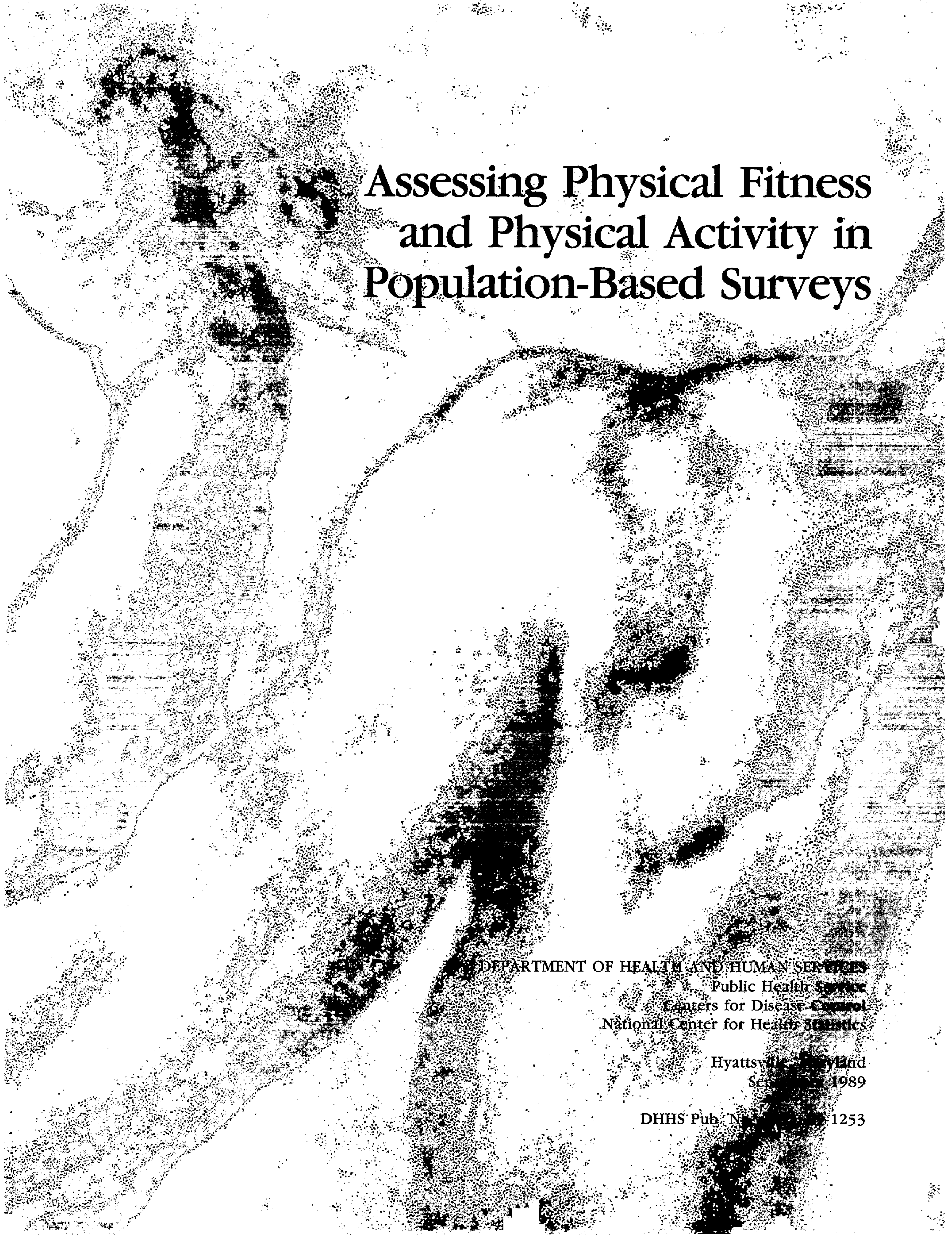
Assessing Physical Fitness and Physical Activity in Population-Based Surveys



DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control
National Center for Health Statistics

Hyattsville, Maryland
Sep[illegible] 1989

DHHS Pub. N[illegible]1253

Copyright Information

Suggested Citation

National Center for Health Statistics. *Assessing Physical Fitness and Physical Activity in Population-Based Surveys.* Thomas F. Drury, ed. DHHS Pub. No. (PHS) 89-1253. Public Health Service. Washington. U.S. Government Printing Office. 1989

Library of Congress Cataloging-in-Publication Data

Assessing physical fitness and physical activity in
population-based surveys.

(DHHS publication; no. (PHS) 89-1253)
Edited by Thomas F. Drury.
"September 1989."
Includes bibliographies and index.
1. Physical fitness—Measurement. 2. Exercise—Measurement. 3. Health surveys. 4. Health surveys—United States. I. Drury, Thomas F. II. National Center for Health Statistics (U.S.) III. Series.
[DNLM: 1. Data Collection—methods. 2. Exertion. 3. Health Surveys. 4. Physical Fitness. WA 950 A846]
RA409.A88 1989 613.7'028'7 88-600045
ISBN 0-8406-0392-4

For sale by the Superintendent of Documents
U.S. Government Printing Office
Washington D.C. 20402

National Center for Health Statistics

Manning Feinleib, M.D., Dr.P.H.
Director

Robert A. Israel
Deputy Director

Jacob J. Feldman, Ph.D.
Associate Director for Analysis and Epidemiology

Gail F. Fisher, Ph.D.
Associate Director for Planning and Extramural Programs

Peter L. Hurley
Associate Director for Vital and Health Statistics Systems

Stephen E. Nieberding
Associate Director for Management

Charles J. Rothwell
Associate Director for Data Processing and Services

Monroe G. Sirken, Ph.D.
Associate Director for Research and Methodology

Division of Epidemiology and Health Promotion

Ronald W. Wilson
Director

Ronald W. Wilson
Acting Chief, Health Status Measurement Branch

Patricia M. Golden
Special Assistant to the Director

Diane K. Wagener, Ph.D.
Chief, Environmental Studies Branch

Foreword

Manning Feinleib, M.D., Dr.P.H., Director
National Center for Health Statistics

During the past 5 years, National Center for Health Statistics (NCHS) staff have made a concerted effort to evaluate the feasibility of developing comprehensive and systematic epidemiologic data on physical fitness and physical activity through the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS). The preparation of a series of expert papers on various aspects of physical fitness and physical activity assessments was central to this evaluation effort.

These papers were initially designed with several purposes in mind. One aim was to contribute to achieving national health promotion objectives in the area of physical fitness and exercise by codifying conceptual, methodological, and measurement approaches to physical fitness and physical activity assessments of the general population. A more immediate purpose was to orient and update NCHS managerial and professional staff on the current state-of-the-art of fitness and activity assessment as a way of stimulating consideration of these topics in NHANES and NHIS planning. Because of their broad import for population-based studies of physical fitness and physical activity, it was always intended to make these papers available to the wider audience of researchers, program planners and evaluators, students, and others with an interest in these topics. The publication of this resource work makes that intention a reality.

The preparation of the papers in this volume coincided with the planning and design of the 1988–94 National Health and Nutrition Examination Survey (NHANES III), which made it possible for their insights and recommendations to be actively and thoughtfully considered. Many of this volume's recommendations in assessing body composition, diet, physical activity, and the physical functioning of older persons have been included in the pilot phases of NHANES III; however, recommendations for the inclusion of tests of cardiovascular endurance, muscular strength and endurance, and flexibility have not been implemented. Whether future cycles of NHANES will include such a fitness test battery remains an open question.

Physical fitness and exercise is only one area of behavioral health for which more adequate national data are needed. In this larger context, the present volume may be viewed as a model of the kind of indepth analysis and consideration that may be necessary to advance other areas of health statistics bearing on national health promotion objectives. Moreover, in the current climate of lively questioning over the types of physical fitness tests and physical activity questionnaires that may be most appropriate for different age groups, this volume also provides an important set of resource materials for future efforts to develop standardized assessments of physical fitness and physical activity in the general population.

Acknowledgments

Overall responsibility for planning and coordinating the content of this publication rested with the Office of Analysis and Epidemiology Program, Division of Epidemiology and Health Promotion, under the general supervision of Jacob J. Feldman and Ronald W. Wilson. *Assessing Physical Fitness and Physical Activity in Population-Based Surveys* was prepared under the direction of Thomas F. Drury, formerly Chief, Health Status Measurement Branch, Division of Epidemiology and Health Promotion, and currently Deputy Director, Epidemiology and Oral Disease Prevention Program, National Institute of Dental Research, National Institutes of Health.

Within the Division of Epidemiology and Health Promotion, day-to-day research assistance and technical review were provided by Laura E. Montgomery and Kathleen M. Danchik. Pennifer Erickson and Bruce Cohen provided special assistance to the project in the areas of health-related quality of life and pulmonary function measurements, respectively. Thomas Stephens provided ongoing consultation and encouragement to the project during a 2-year visitation as NCHS Distinguished Visiting Scientist. Within the Office of Analysis and Epidemiology Program, general administrative assistance and production planning and coordination were provided by Madelyn A. Lane and Dorothea J. Donahue. Thomas Hodgson, Chief Economist, provided ongoing consultation on the economic analysis of fitness and exercise data; Richard Gillum, on the development and uses of measures of cardiovascular endurance.

Within the Publications Branch, publication management was provided by Rolfe W. Larson. Printing and coordination were managed by Linda L. Bean. The cover was designed by Sarah M. Hinkle.

The authors of each chapter, who are to be commended for doing such an exceptionally fine job, are identified in their respective sections. In addition to the authors and NCHS staff identified below, who provided technical review of one or more manuscripts, technical review was also provided by Eleanor M. Pao, Robert S. Gold, Carl Slater, Clifford Clogg, Samuel M. Fox III, and Bernice Cohen.

Publication of *Assessing Physical Fitness and Physical Activity in Population-Based Surveys* would not have been possible without the contributions of numerous staff members throughout the National Center for Health Statistics. The consultation, cooperation, and technical review efforts of staff from the Division of Health Examination Statistics were pivotal, particularly that provided by Robert Murphy, Catherine E. Woteki, Kurt Maurer, Dale Hitchcock, Peter J. Gergen, Stacey Fitzsimmons, Claudia Moy, Christopher Sempos, Suzanne G. Haynes, Trena Ezzati, and Clifford Johnson. Within the Office of Data Processing and Services, Daniel Savage, David L. Larson, Jean S. Findlay, and Charlotte K. Leahy provided technical assistance and orientation to physical and logistical issues arising in mobile examination centers. Staff of the Division of Health Interview Statistics, including Robert Fuchsberg, Owen Thornberry, Stewart C. Rice, Jr., Marcie Cynamon, and Peggy Barker, provided technical assistance in the development and uses of physical activity assessments obtained through the 1985 National Health Interview Survey. In the Office of Program Planning, Evaluation and Coordination, Gail F. Fisher provided health data policy guidance, Marjorie Greenberg provided assistance in the area of program evaluation, Nancy G. Hamilton, with the assistance of Michele Fair and Sheryl D. Ashworth, coordinated conference management services for the Airlie House Workshop, and Richard Havlik provided consultation and technical review in the area of the biochemical correlates of fitness and exercise. Within the Office of Research and Methodology, James T. Massey and Andrew A. White provided invaluable guidance in the areas of survey design and analysis.

A broad spectrum of support and assistance for the project also came from outside NCHS. Within the Office of Disease Prevention and Health Promotion, Henry Montes, Glenn Gilbert, and Larry Green provided important consultation and materials in the early stages of the project, as did Ash Hayes, Glenn Swengros, and Christine G. Spain of the President's Council on Physical Fitness and Sports, and Peter W. F. Wilson, with the Framingham Heart Study of the National Heart, Lung, and Blood Institute. Kenneth E. Powell and Carl J. Caspersen of the Centers for Disease Control provided key assistance in the area of physical activity assessment, particularly at the Airlie House Workshop.

Additional outside consultation was provided by Raymond A. Ciszek on youth fitness testing; by Michael Chubb on fitness resources; by David Jackson on measurement and analysis strategies; by Carl Slater on health education and health promotion uses of fitness and exercise data; by Samuel M. Fox III on ethical and safety aspects of fitness testing; by William G. Reddan on fitness testing of elderly persons; by Ralph E. Berry, Jr., Thomas Deley, and Gerry Oster on economic aspects of fitness- and exercise-related activities; and by Bernice Cohen on genetic aspects of fitness. Between 1985 and 1987, the annual meetings of the American College of Sports Medicine also provided innumerable occasions for project staff to consult informally with members of the College on the broad range of issues dealt with in this publication for which special thanks is gratefully given.

Finally, special acknowledgment needs to be given to Steven N. Blair for his timely and gracious assistance throughout the project, particularly for chairing the physical activity assessment session at the Airlie House Workshop, and to E. R. Buskirk for chairing the entire Airlie House Workshop and for encouraging the project staff to "see it through" to this publication.

Contents

Part I Historical Perspectives

Part II Fundamental Perspectives on Health-Related Physical Fitness and Cardiopulmonary Health

Part III Fundamental Perspectives on Energy Balance, Dietary Patterns, and Physical Activity

Part IV Special Subpopulation Issues

Part V Lessons From Community, National, and International Studies

Part VI Contexts of Evaluation

Part VII Measurement and Analysis Strategies

Introduction

Thomas F. Drury, Ph.D., ***Editor***
National Center for Health Statistics

In 1983, National Center for Health Statistics (NCHS) staff in the Division of Epidemiology and Health Promotion, Office of Analysis and Epidemiology Program, initiated an evaluation study of NCHS fitness and physical activity data. This study, "Assessing Physical Fitness and Physical Activity in NCHS Population-Based Surveys," was concerned with three types of evaluation: (1) an evaluation of effort, (2) an evaluation of effectiveness, and (3) an evaluation of adequacy. Based on these evaluations, an effort was also made to develop recommendations on how inadequacies in existing physical fitness and physical activity measurements might be better addressed through the population-based surveys of NCHS.

In carrying out these evaluations, answers were sought to the following questions: What efforts have so far been made in NCHS data systems to assess physical fitness and physical activity? In the context of what is currently known about the assessment of physical fitness and physical activity in population-based studies, how good is the quality of available NCHS data on specific physical fitness and activity issues? Given what we most need to know about specific aspects of fitness and exercise from scientific, policy, and programmatic perspectives, how adequate are the available data to answer such questions?

The evaluation took place in three phases. In Phase I, a series of state-of-the-art papers were prepared by leading experts related to the assessment of physical fitness and physical activity in NCHS general population surveys. In Phase II of the study, NCHS brought the primary authors of these state-of-the-art papers together at a workshop designed to formulate specific recommendations regarding fitness testing, activity assessment, and related issues in the specific contexts of the National Health and Nutrition Examination Survey, the National Health Interview Survey, and the National Survey of Family Growth. This workshop, held at Airlie House in Warrenton, Virginia, on June 2–5, 1985, was carried out under the NCHS Health Statistics Workshops and Conferences Contract administered by the NCHS Office of Program Evaluation and Planning.

Phase III of the overall evaluative effort has focussed on the preparation of the present volume. It is intended to be a major reference work, containing the revised manuscripts of the background papers prepared during Phase I of the overall evaluation study along with a few additional papers prepared subsequent to the Airlie House Workshop.

In the remainder of this introduction, the structure and content of the volume are outlined. In a brief discussion of scope, suggestions are made for extending the work presented here. Finally, the relationship of this publication to earlier Public Health Service projects is noted.

Structure and content

This volume is organized into seven parts: The first part provides a historical perspective on NCHS efforts in the area of physical fitness and activity assessment. Nancy Pearce (chapter 1) provides an overview of NCHS population-based surveys. Arthur McDowell (chapter 2) documents the efforts made so far in NCHS to assess

---

Note: Thomas F. Drury is currently affiliated with the National Institute of Dental Research.

cardiovascular endurance, muscular strength, and lung function. Dorothy Blair, Jean-Pierre Habicht, and Lee Alekel (chapter 3) trace the efforts made to assess body composition, dietary patterns, and nutritional status.

The second section is devoted to fundamental perspectives on health-related physical fitness and cardiopulmonary health. Jack Wilmore (chapter 4) identifies major components of health-related physical fitness. He provides an indepth discussion of the range of basic design issues and alternatives that need to be considered in assessing these components of health-related physical fitness among apparently healthy adults. Karlman Wasserman (chapter 5) introduces the concept of cardiopulmonary health, outlines measurements of this concept, identifies noninvasive integrative methods for developing these measurements, and discusses the interpretive uses of such assessments.

Part III provides a series of fundamental discussions on energy turnover, energy balance, dietary patterns, and physical activity. E.R. Buskirk (chapter 6) introduces this section with an overview of the issues involved in the assessment of energy intake and energy expenditure. Dorothy Blair (chapter 7) addresses the relevance that the concept of energy balance may have for a large-scale examination survey of the general population. Catherine Woteki (chapter 8) reviews the current state of the art in the assessment of dietary patterns, and Thomas Stephens (chapter 9) performs a similar review of design alternatives in the conceptualization and survey measurement of physical activity.

The fourth part addresses issues that arise in assessing the physical fitness and physical activity of selected subpopulations. Oded Bar-Or (chapter 10) discusses special concerns that arise in the study of children and adolescents. On this same theme, James Ross (chapter 11) codifies the major lessons to be learned from the National Children and Youth Fitness Studies. Barbara Drinkwater (chapter 12) synthesizes and evaluates essential information on the fitness testing of women. This information should be taken into account in the design of population-based studies of women's fitness and activity patterns. Nanette Wenger (chapter 13) reviews and evaluates pertinent findings from the study of cardiac patients in terms of the implications for including such persons in a national survey of physical fitness. Everett Smith and Catherine Gilligan (chapter 14) provide an extensive review of the literature on aging in relationship to cardiovascular endurance, flexibility, and bone; on the basis of these reviews they develop recommendations for health-related fitness testing of the elderly.

The fifth part highlights the lessons to be learned from selected community, national, and international studies. Henry Montoye (chapter 15) develops the lessons to be learned from the Tecumseh studies of physical fitness and physical activity. James Vogel (chapter 16) provides a systematic review of the fitness and activity protocols that he and his colleagues have used in major studies of U.S. Army personnel. Thomas Stephens and Cora Craig provide an inside review of the Canada Fitness Survey (chapter 17). Roy Shephard (chapter 18) rounds out this section with a discussion of the lessons to be learned from recent international studies, particularly those conducted in Canada.

In the sixth part contexts are reviewed within which physical fitness and physical activity measurements need to be evaluated. Robert Malina and Claude Bouchard (chapter 19) provide an up-to-date review of genetic considerations in physical fitness. William Haskell (chapter 20) reviews the current state of knowledge in the area of the biochemical correlates of fitness and exercise. James House and David Smith (chapter 21) review major ways of conceptualizing and measuring physically and psychologically demanding work on and off the job in relationship to health. Arthur Leon (chapter 22) reviews measurement approaches to physical fitness and physical activity in the context of a substantive discussion of the effects of physical activity on health. Steven Blair and Harold Kohl (chapter 23) review what is currently known about the relationship of physical activity to other health behaviors and lifestyle factors. Ronald LaPorte (chapter 24) identifies the need to study the full spectrum of physical activity, reviews what is currently known about the quality of survey measurements of physical activity, and discusses several anomalies that arise in evaluating interrelationships among physical activity, cardiovascular fitness, and other health characteristics. Erika Sivarajan and Victor Froelicher (chapter 25) provide a critical perspective on high-tech assessments of physical fitness in the context of national surveys of the general population and highlight the need for more adequate assessments of physical activity patterns. Robert Hyde and Ralph Paffenbarger (chapter 26) provide an overview of the history of epidemiologic concern with physical activity. They review some of the landmark contributions to this area of study and state the case for national assessments of physical activity.

The seventh section of this volume introduces several major measurement and analysis strategies that, although not yet extensively applied in studies of physical fitness and exercise, hold great promise for the analysis of data bearing on these topics. George Bohrnstedt and Joseph Lucke (chapter 27) introduce the notion of physical fitness and physical activity as latent variables. They suggest that future studies of these contructs might consider using LISREL and other available software packages to evaluate measurement reliability and to adjust for measurement error in substantively oriented analyses of the causes and consequences of fitness and exercise. Bengt Muthen and Lynn Short (chapter 28) extend this discussion by reviewing recent advances in regression and factor analysis, which have import for the analysis and evaluation of fitness and exercise measures when data take the form of dichotomies or other qualitative distributions. Allan McCutcheon (chapter 29) provides an overview of latent class analysis, including the range of uses to which this type of analysis might be put.

Many of the chapters in the seven parts of this volume include extensive appendix material that either is not elsewhere published or is difficult to obtain, particularly by newcomers to this field of study. This is particularly the case with the various schemes that have been used to develop estimates of energy expenditure from reported activity measurements. Accordingly, an appendix to chapter 15 reproduces the energy expenditure coding system used in the Tecumseh studies; and an appendix to chapter 17 shows the coding scheme used in the Canada Fitness Survey for this purpose. In the only stand-alone appendix to this volume, Andrew Dannenberg and Peter Wilson describe the procedures that were used in the Framingham Offspring Study to assess physical activity and to develop estimates of energy expenditure.

Scope

Given limits of time and other resources, any evaluation of the feasibility of assessing physical fitness and physical activity in the context of the population-based surveys of NCHS is bound to give only brief attention to certain topics that others would consider terribly important and that we might well have developed more fully had we known when this project began what we know now. To further clarify the scope of this volume and in hope of inspiring scholarly attention to these issues, some of the more important of these omissions with respect to physical fitness and physical activity assessments in population-based surveys are identified below.

A basic assumption of this volume is that the lack of national assessments of the physical fitness status of the adult population represents a major gap in the national system of health statistics. In designing the papers, however, an effort was made to include discussions (see, for example, chapter 25) that would question the merits of this assumption. The aim here was to stimulate consideration by health data policymakers of the relative merits and uses of physiological measures of fitness in comparison with the merits and uses of self-report information on physical activity, perceived fitness levels, and other healthy lifestyle characteristics.

The protection of human subjects, a key issue in fitness testing, is touched on in a number of the papers. This topic deserves specialized attention. However, at the time the papers in this volume were commissioned, there was great reluctance among potential authors to take on this task. In the interim, the experiences gained from more recent fitness testing in large-scale, community-based studies are finding their way into the literature. It is to be hoped that these will provide the basis for someone to carry out the needed thorough review of issues pertaining to the protection of human subjects in large-scale epidemiologic studies of physical fitness.

Although an effort has been made to consider the special testing needs of persons with various medical conditions, only persons with cardiac conditions have actually received systematic consideration (chapter 13). Other medical conditions, including arthritis, diabetes, chronic respiratory disorders, and obesity, as well as various forms of disability, also need to be considered in the design of physical fitness test protocols for use in population-based surveys.

The economic impact of physical fitness and exercise is a topic that is receiving increasing attention. In view of the fact that national health promotion objectives call for the development of methodologies to assess such an impact and that a number of conceptual and methodological approaches already exist, which might be applied to the case of physical fitness and exercise, an examination of the kinds of national data that these conceptual models and methodological approaches require would be timely.

Relationship to other Public Health Service projects

The project that gave rise to this volume benefited from a number of related projects in the area of physical fitness and exercise, many of which were sponsored or carried out by various agencies of the United States Public Health Service. In the initial stages of the project, NCHS staff profited immensely from the Office of Disease Prevention and Health Promotion (ODPHP) reviews of the literature on physical fitness and physical activity participation, as well as from the various planning activities surrounding ODPHP's sponsorship of the National Children and Youth Fitness Studies. In 1984, Public Health Service agencies carried out three workshops on various aspects of physical activity that made it possible for NCHS staff to focus more intensely on physical fitness issues than might have otherwise been the case (1–3).

Concluding remarks

In the preparation of their manuscripts, authors were asked to provide recommendations for NCHS surveys. These recommendations are included here, not because they represent in all particulars a blueprint for local community, State, or regional surveys of physical fitness or physical activity, but rather because they highlight a number of important issues that smaller surveys need to resolve in order to demonstrate more fully the feasibility of developing national assessments.

References

1. Morgan, W. T., and S. E. Goldston, eds. 1987. *Exercise and Mental Health.* New York: Hemisphere Publication.
2. Wilson, P. W. F., R. S. Paffenbarger, J. N. Morris, and R. J. Havlik. 1986. Assessment methods for physical activity and physical fitness in population studies: report of a NHLBI workshop. *Am. Heart J.* 111 (6): 1177–92.
3. Powell, K. D., and R. S. Paffenbarger. 1985. Workshop on epidemiologic and public health aspects of physical activity and exercise: a summary. *Public Health Rep.* 100 (2): 118–26.

Part I

Historical Perspectives

Chapter 1

General Population Surveys of the National Center for Health Statistics: An Overview

Nancy D. Pearce, M.A.
National Center for Health Statistics

Introduction

The National Center for Health Statistics (NCHS) is one of the major Federal statistical organizations. It operates a diverse survey and inventory program with legislative authorization (42 USC 242k) to collect statistics on the following:

- The extent and nature of illness and disability of the population of the United States, including life expectancy, maternal morbidity, and mortality.
- The impact of illness and disability of the population on the economy of the United States and on other aspects of the well-being of its population.
- Environmental, social, and other health hazards.
- Determinants of health.
- Health resources, including health professionals by specialty and type of practice, and the supply of services by hospitals, extended-care facilities, home health agencies, and other health institutions.
- Utilization of health care, including utilization of ambulatory health services, services of hospitals, extended care facilities, home health agencies, and other institutions.
- Health care costs and financing.
- Family formation, growth, and dissolution.

The Department of Health and Human Services 1990 Objectives for the Nation (1) in the areas of fitness and exercise call for the development of methodologies to monitor physical fitness and exercise patterns in the general population. In response to these objectives, staff of the National Center for Health Statistics have been attempting to assess the extent to which NCHS general population surveys might provide a context for the development of national statistics on fitness and exercise.

This paper provides background information about NCHS and describes its population-based survey program, including recent and expected changes in the program. The first part of the paper describes general aspects of survey planning, data collection and processing, and data release. Next, each of the population-based surveys is described. These surveys are the National Health Interview Survey, the National Medical Care Utilization and Expenditure Survey, the National Health and Nutrition Examination Survey (NHANES), the first NHANES (NHANES I) Epidemiologic Followup Survey, and the National Survey of Family Growth. Finally, there is a discussion of recent and potential changes in the NCHS survey program. These changes include an integrated survey design through which the samples for other NCHS population-based surveys are linked to that of the National Health Interview Survey, the capability to conduct random-digit-dialed telephone surveys, and the ability to use the National Death Index to determine the fact and cause of death for individuals in NCHS surveys.

Although this paper is nontechnical in nature, it does include some technical terms. For the purposes of NCHS surveys, many terms have highly specialized definitions. No effort has been made here to provide a comprehensive series of such definitions.

Background

It is important to be aware of the general environment in which NCHS data collection programs operate. This section briefly describes the processes through which surveys are developed and implemented and some of the constraints on data collection, analysis, and dissemination.

Survey planning process

A somewhat lengthy survey planning and development process is followed in NCHS. This is a result of the budget process (which includes request and approval for funds to conduct each year's data collection activities), the actual survey-planning process (which requires time for development and testing of survey questionnaires, sample design, and procedures), and administrative processes, such as approval from the Office of Management and Budget (OMB).

Typically 2–3 years elapse from the time a decision is made to conduct a survey and funds are requested through the budget process to the time that data collection begins. After the data are collected, additional time is required to complete data reduction, editing, and analysis for dissemination.

As the statistical agency with responsibility for health data, NCHS makes every effort to meet data needs of other Public Health Service agencies that can best be met through NCHS survey programs. In addition to consulting with health program planners, managers, and researchers in the Federal Government, NCHS staff consult widely with experts in the private sector. Decisions are made about survey content and design based on priorities identified in the planning process.

Following broad decisions about content areas, staff responsible for survey planning and development review the availability of prototype questions and procedures that have previously been developed and tested for reliability and validity. To the extent that reliable, valid questions and procedures suitable for use in a particular survey do not exist, the development and testing process is lengthened.

Reviews and approvals

Several reviews and approvals must be obtained before data collection may begin. These reviews are directed at technical aspects of survey design and human subject concerns. Within NCHS both types of review occur, and it is only after any issues are resolved that the plans for the survey are forwarded to the Office of Management and Budget, which, under the requirements of the Paperwork Reduction Act of 1980 (44 USC 3501–3520), must review and approve any collection of information from more than 9 respondents.

Technical aspects of the survey design and methodology are reviewed by the NCHS Office of Research and Methodology. The purposes of this review (2) are as follows:

- To verify that the proposed statistical methodological techniques employed in NCHS projects are consistent with accepted statistical theory and practice.
- To verify that each project can conceptually meet its stated objectives.

In addition to technical aspects of the survey, consideration must be given to the potential sensitivity of individual questions or procedures. The plans and materials for each survey are also reviewed by the Committee on Protection of Rights of Human Subjects (3) to assure that the project has intrinsic value or scientific merit and to provide the following safeguards:

- Respect of the individual respondent's privacy, dignity, comfort, and legal rights.
- Freedom of choice for the subject as to participation in the project.
- Protection against undue hazards to the health and safety of the subject and others who might be endangered by the conduct of the research.

Any Department of Health and Human Services (DHHS) project that includes biomedical or behavioral research involving humans is subject to the provisions of 45 CFR 46 (revised January 26, 1981) for the protection of human research subjects. The oversight program is administered by the Office for Protection From Research Risks, Office of the Director, National Institutes of Health. NCHS projects are reviewed by an Institutional Review Board instead of the Committee on Protection of Rights of Human Subjects to assure that appropriate informed consent procedures are followed when conducting physical tests and examinations or when eliciting particularly sensitive information.

When both technical and human subject reviews have been completed, approval for the survey is sought from the Office of Management and Budget. (OMB review is conducted according to provisions of 5 CFR 1320, the final rule implementing the Paperwork Reduction Act of 1980.) In its review of data collection plans, OMB gives particular attention to the need for and expected uses of the data, the total cost of the project in relation to the need for the data, how essential the project is to the proper fulfillment of the sponsoring agency's mission, and whether there is duplication of existing information. Pilot projects and pretests often require separate review and approval by OMB, a process that requires approximately 3 months from the time the request leaves NCHS.

Statistical standards and guidelines

Data collection undertaken by NCHS is subject to statistical standards established by OMB (4) and to standards and guidelines of DHHS (5) and of the Public Health Service (PHS). These standards and guidelines relate to survey methodology, to how certain data items are to be obtained and coded, and to how data are presented in printed reports.

Statistical standards established by OMB relate to the adequacy of sample design and the achievement of an adequate response rate, generally considered to be at least 75 percent for statistical surveys. In order to maximize comparability of data collected by Federal agencies and to minimize the recordkeeping burden on businesses and the public, OMB has also established standards for certain data items. For example, when a project obtains information on the race of individuals, it is also necessary to determine whether the individual is of Hispanic origin; minimum categories have been established for the collection and reporting of such information. These standards must be adhered to, and justification for noncompliance must be provided by the sponsor and accepted by OMB.

The programs of data standards and guidelines of DHHS and PHS expand on OMB standards. They include, for example, specifications that exact date of birth rather than age at last or nearest birthday be asked to determine age of individuals and that certain minimum age groupings be used when only aggregate information about persons is obtained.

NCHS has established certain additional data standards. These relate to data collection, data processing, and data analysis. Although it is not feasible to discuss these various standards and guidelines in detail, it is important that they exist and may affect how particular information is obtained and used.

Confidentiality of data

Data collection activities of the Center are subject to the specific provisions of both the Privacy Act of 1974 (Public Law 93–579) and the Health Services Research, Health Statistics, and Health Care Technology Act of 1978 (Public Law 95–623). The Privacy Act covers all federally sponsored and operated data collection that involves creation of a system of records containing unique personal identifiers; the latter is limited to a portion of the Public Health Service.

In keeping with the requirements of the Privacy Act of 1974 and NCHS policy, each individual, household, or establishment asked to provide data for NCHS is informed of the following:

- The authorization for soliciting the information.
- The fact that disclosure of such information is voluntary.
- The principal purpose or purposes (usually statistical or research) for which the information is intended.
- The routine uses (as published in the *Federal Register*) that may be made of the information.
- The effects on the respondent, if any, of not providing all or any part of the requested information.

NCHS is also bound by the provision of its authorizing legislation stipulating that information obtained in the course of its statistical activities may not be used for any purpose other than that for which it was supplied. Such information may not be published or released in another form if the particular establishment or person supplying the information or described in it is identifiable, unless the establishment or person has consented to its publication or release in another form.

Almost all data collected by NCHS are requested under a pledge to the respondent, either an individual or establishment, that the information will be used only for statistical purposes and will not be released in such a way that individual persons, households, or facilities will be identifiable. NCHS never releases information that would identify an individual or household. The *NCHS Policy Statement on Release of Data for Individual Elementary Units and Special Tabulations* (6) and the *NCHS Staff Manual on Confidentiality* (7) present and explain in detail the confidentiality policy.

Data collection mechanisms

NCHS does not have a large data collection staff of its own. It collects most of its data through interagency agreements with the U.S. Bureau of the Census or through contracts with non-Federal organizations. When a competitive contract is awarded for conduct of a survey, additional lead time is required for NCHS staff to prepare the specifications for the request for proposal, for the procurement process to solicit technical proposals, for review of proposals received, and for negotiation and award of the final contract. Staff of the U.S. Bureau of the Census and of contractors are bound to maintain confidentiality of survey respondents in the same way as NCHS staff are.

Morbidity and mortality data coding

NCHS collects a wide range of information about illnesses and health problems of the population of the United States. To enhance the comparability of international statistics and of morbidity and mortality estimates from such diverse sources as household respondents and health care provider records, these data are coded to appropriate adaptations of the International Classification of Diseases (ICD). ICD is a system for classification of diseases and injuries for statistical purposes and is revised periodically under the auspices of the World Health Organization.

All NCHS mortality and morbidity information is classified according to the most recent revision of ICD. Data for the approximately 10-year period prior to January 1, 1979, are coded according to the *Eighth Revision International Classification of Diseases, Adapted for Use in the United States,* ICDA–8 (8). Mortality data for the period beginning January 1979 are classified according to the 9th Revision of the International Classification of Diseases (9), and morbidity data are classified using the clinical modification of the 9th Revision, ICD–9–CM (10).

Data release mechanisms

NCHS releases its data in several ways. When they are issued, the *Vital and Health Statistics* series (detailed reports on the design of the various data collection systems, findings of methodological studies, and detailed cross-tabulations of data), *Advance Data* reports (which summarize newly available data on topics of special interest), and *Monthly Vital Statistics Reports* are distributed without charge to libraries, medical colleges, schools of public health, and other selected institutions that have requested all issues in the series. Other organizations and individuals receive publication notices and information to purchase copies from the U.S. Government Printing Office. The annual *Catalog of Publications of the National Center for Health Statistics* (11) contains a list of reports issued during the previous 5 years, and the quarterly *Publication Note* is issued to provide the latest information.

NCHS also issues a catalog that describes public use data tapes available for purchase. The data tape program, begun in 1969, permits researchers to perform their own analyses of data. Tape availability and contents, along with information on ordering the tapes, are detailed in a periodic *Catalog of Public Use Data Tapes From the National Center for Health Statistics* (12) and *Data Tape Update.* The majority of these data tapes are now sold by the National Technical Information Service (NTIS) and are not available from NCHS. Eventually, all NCHS public use data tapes will be available from NTIS. The order form in the back of the catalog gives the address and instructions for ordering from NTIS or, in a few instances, from the National Center for Health Statistics.

Many requests for unpublished data are filled regularly using tabulations that have been previously compiled. In addition, many special tabulations are prepared each year to meet data requests that cannot be fulfilled in any other way. When special tabulations are necessary, the requester is given a cost and time estimate. Tabulations and public use data tapes are carefully reviewed to ensure that confidentiality is maintained.

Population-based surveys

Table 1 provides an overview of NCHS population-based surveys. The following brief descriptions provide additional detail of each survey's objectives, sample design, questionnaire content, and data collection procedures. More detail is provided in technical appendixes of analytic reports from the surveys.

National Health Interview Survey

Background

The National Health Interview Survey (NHIS) is a principal source of information on the health of the civilian noninstitutionalized population of the United States. The National Health Survey Act of 1956 provided for a continuing survey and special studies to secure on a voluntary basis accurate and current statistical information on the amount, distribution, and effects of illness and disability in the United States and the services rendered for or because of such conditions. The survey referred to in the act, now called the National Health Interview Survey, was initiated in July 1957. In its early years the survey was known to many as the National Health Survey, the name now given to a broader program of surveys in the National Center for Health Statistics.

Purpose and scope

The purpose of NHIS is to provide national data on the incidence of acute illness and accidental injuries, the prevalence of chronic conditions and impairments, the extent of disability, the utilization of health care services, and other health-related topics. A major strength of this survey lies in the ability to display these health characteristics by selected demographic and socioeconomic characteristics of the U.S. civilian noninstitutionalized population.

Because NHIS data are obtained during household interviews from the people themselves, the statistics measure health status and experiences and reflect the social and economic dimensions of health as reported by individuals—the extent and impact of illness and disability and the resulting uses of health care services by the people experiencing them.

Interviews are conducted each week throughout the year in a probability sample of households. Data collected over the period of a year form the basis for the development of annual estimates of the health characteristics of the population and for the analysis of trends in those characteristics.

The survey covers the civilian noninstitutionalized population of the United States. Because of technical and logistical problems, several segments of the population are not included in the sample or in the estimates from the survey. Persons excluded are patients in long-term care facilities (data are obtained on patients in some of these facilities through the NCHS National Nursing Home Survey); persons on active duty with the Armed Forces (although their dependents are included); U.S.

nationals living in foreign countries; and persons who died during the calendar year preceding the interview.

Sample design

The National Health Interview Survey is a cross-sectional household interview survey. The sampling plan follows a multistage probability design that permits the continuous sampling of households. Every 10 years, the NHIS sample design is revised following the decennial census of the population. Beginning in 1985, the NHIS sample design has reflected changes based on the 1980 census and an integrated survey design capability has been implemented. The integrated survey design, discussed in greater detail in a subsequent section, will permit use of NHIS as the sampling frame for other NCHS population-based surveys, which in the past have been designed independently.

Conceptually, the sampling plan for the 1985 NHIS remains the same and follows a multistage probability design with continuous sampling of the civilian noninstitutionalized population of the United States. The first stage consists of a sample of primary sampling units drawn from approximately 1,900 geographically defined primary sampling units that cover the 50 States and the District of Columbia. A primary sampling unit consists of a county, a small group of contiguous counties, or a metropolitan statistical area. Within primary sampling units, smaller units, called segments, are defined in such a manner that each segment contains an expected number of households. The sampling plan is designed to yield national estimates, although estimates can be obtained separately for the four geographic regions.

The households selected for interview each week are a probability sample representative of the target population. Each calendar year through 1984, data were collected from approximately 40,000 households, including about 110,000 persons. The sample design implemented in 1985 includes approximately 50,000 households with about 135,000 individuals; because of budgetary restrictions, however, it was possible to implement only a three-fourths sample for 1985. The annual response rate of NHIS is usually at least 95 percent of the eligible households in the sample. The 5-percent nonresponse is divided equally between refusals and households where no eligible respondent can be found at home after repeated calls.

Data collection procedures

Data are collected through personal household interviews conducted by a permanent staff of interviewers employed and trained by the U.S. Bureau of the Census according to procedures specified by the National Center for Health Statistics.

All adult members of the household 17 years of age and over who are at home at the time of the interview are invited to participate and to respond for themselves. The mother is usually the respondent for children. For individuals not at home during the interview, information is provided by a responsible adult family member (19 years of age or over) residing in the household. Approximately 65 percent of the adults 17 years and over are self-respondents. On occasion, a random subsample of adult household members is selected to respond to questions on selected topics. There are also instances in which followup supplements are completed for either the entire household or for individuals identified as having particular health problems. As required, either these supplements are left for the appropriate person to complete and return by mail or the interviewer calls again in person or by telephone to secure the information directly. There are, however, additional costs associated with leaving a supplement for self-completion (these costs include followup to obtain a completed questionnaire and processing of separate documents that must be linked to the original household questionnaire) or for additional contacts to complete a personal interview with the sample person.

Nationally there are approximately 110 interviewers, trained and directed by health survey supervisors in each of the 12 Census Bureau regional offices. The supervisors are career Civil Service employees whose primary responsibility is the National Health Interview Survey. The interviewers are part-time employees, selected through an examination and testing process. Interviewers receive thorough training in basic interviewing procedures and in the concepts and procedures unique to NHIS.

On the average, the interviews require about 45 minutes in the household. Depending on the family size and the nature and extent of health conditions of family members, the length of interview typically ranges from 15 to 90 minutes.

Content of questionnaire

The questionnaire consists of two basic parts: a core set of health, socioeconomic, and demographic items and one or more sets of supplementary health items. The core items constitute approximately 70 percent of the questionnaire and are repeated each year. The arrangement of core items, complemented by rotating and single-time supplements (13, 14), allows the survey to respond to changing needs for data and to cover a wide variety of topics while providing continuous information on fundamental topics. Consideration is currently being given to shortening the core questionnaire in order to permit additional interview resources to be devoted to supplement topics.

The questionnaire now includes the following types of core questions (15):

- Basic demographic characteristics of household members, including age, sex, race, education, and family income.
- Disability days, including restricted-activity days, bed days, work-loss days, and school-loss days, occurring during the 2-week period prior to the week of interview.

- Physician visits occurring during the same 2-week period.
- Acute and chronic conditions responsible for these disability days and physician visits.
- Long-term limitation of activity resulting from chronic disease or impairment and the chronic conditions associated with the disability.
- Short-stay hospitalizations, including the number of persons with hospital episodes during the past year and the number of discharges from short-stay hospitals.
- The interval since the last physician visit.

The questionnaire also includes six lists of chronic conditions. Each condition list concentrates on a group of chronic conditions involving a specific system of the body (for example, digestive, skin and musculoskeletal, circulatory, and respiratory). The body systems approach to chronic conditions was adopted because it was found that organizing questions around a particular body system resulted in more thorough reporting and increased the number of conditions for which estimates of prevalence could be made. Beginning in 1978, each of six representative subsamples of households has been asked questions based on one of the six lists of chronic conditions. In this way, national estimates on each of the six body systems are obtained during the same interview year. Prior to 1978, only one condition list was asked each year.

The supplements to the questionnaire change in response to current interest in special health topics. Throughout 1981, a child health supplement was included. In 1982, supplement topics were health insurance coverage and preventive care. Health habits (including alcohol consumption), dental care, and access to care were the supplement topics in 1983. For 1984, aging and health insurance coverage were the supplement topics. Throughout 1985, the supplement is on health promotion and disease prevention, and 1986 supplements are planned on health insurance, dental care, and use of vitamin and mineral supplements.

Suggestions and requests for special supplements are solicited and received from many sources. These include university-based researchers, administrators of national organizations and programs in the private and public health sectors, and specialists from other parts of the Department of Health and Human Services (for example, the National Institutes of Health and the Centers for Disease Control). Although it is not possible to include all of the suggested topics, every effort is made to be responsive to the data needs of such groups. A lead time of at least 1 year is required to develop and pretest questions for new topics to be included as special supplements. For example, supplements selected in 1985 will be developed and pretested in 1986 and implemented in 1987 or 1988. Relevant portions of the questionnaire are included in an appendix to each Series 10 report, and each year's questionnaire is reproduced in its entirety in the annual Current Estimates report.

Data processing and quality control methods

Throughout the data collection and processing phases, there are extensive quality control activities. Each interviewer edits all completed work before returning it to the regional supervisor, and in the regional office there is a preliminary edit of all questionnaires. As part of the quality control program, interviewers are provided feedback on their errors. In addition, a small sample of households is reinterviewed, usually by telephone, by a supervisor or senior interviewer. The reinterview questionnaire includes questions from both the core and supplement(s) and is intended to verify that the interview was conducted correctly and completely.

The Census Bureau regional offices forward the questionnaires to the National Center for Health Statistics for coding and data processing. At this time, each questionnaire is checked again at NCHS for completeness of field coverage, omissions, and other errors. Illnesses, diseases, and injuries reported by the households are coded to a modified version of the International Classification of Diseases. All coding is subject to recoding on a sample basis to ensure a high level of accuracy.

Thus, potential nonsampling errors such as those in reporting, processing, and nonresponse, which are inherent in any sample survey, are kept to a minimum by methods built into the survey procedures. In addition, with regard to sampling error, standard error charts are created and provided so that a user can calculate the variations in the estimates that might occur because only a sample of the population is surveyed.

National Medical Care Utilization and Expenditure Survey

Background

The National Medical Care Utilization and Expenditure Survey (NMCUES) is a unique source of detailed national estimates on the utilization and expenditures for various types of medical care. This survey builds on the experience of the National Health Interview Survey, the National Medical Care Expenditure Survey, and the former Current Medicare Survey. The first cycle of the survey covered calendar year 1980.

Purpose and scope

NMCUES is designed to be directly responsive to the continuing need for statistical information on the health care expenditures associated with health services utilization for the entire U.S. population. Cycle I was designed and conducted in collaboration with the Health Care Financing Administration to provide detailed utilization and expenditure data for persons in the Medicare and Medicaid populations. NMCUES will produce estimates over time for evaluation of the impact of legislation and programs on health status, costs, utilization, and illness-related behavior in the medical care delivery system.

Cycle I was composed of several related surveys. The household portion of the survey consisted of a national survey of the civilian noninstitutionalized population and a separate survey of the Medicaid-eligible populations of the States of New York, California, Texas, and Michigan. These two surveys each consisted of five interviews over a period of about 15 months to obtain information on medical care utilization, expenditures, and other health-related information. A third survey, an administrative records survey, was designed to verify the eligibility status of the household survey respondents for the Medicare and Medicaid programs. Insurance claims filed with the national Medicare program and Medicaid programs in each of the four States were checked for persons in the sample of Medicaid eligibles.

Sample design

The national Cycle I household survey comprised persons residing in about 6,000 households. The sample for this survey was a multistage area probability sample drawn from 106 primary sampling units representing the 50 States and the District of Columbia. The State Medicaid household survey sample consisted of about 1,000 families in each of the four States; these families were selected with a known probability of selection from the State Medicaid enrollment lists. Thus, the total sample for the survey was about 10,000 households.

An overall response rate for the household interviews of 89.4 percent was achieved in the first interview in both household surveys in Cycle I: For the national household survey, the response rate was 91.4 percent, and for the State Medicaid survey, the rate was 86.7 percent. Attrition over the course of interviewing resulted in final response rates of 84.9 percent for the national household survey and 76.1 percent for the State Medicaid household survey.

Data collection procedures

NMCUES is a panel study conducted by private contractors according to NCHS specifications. Interviews in Cycle I were conducted with each household at approximately 3-month intervals, with interviewing beginning in February 1980 and ending in March 1981. The first two interviews were conducted by personal visit of the interviewer to the household, the next two were conducted by telephone (if a telephone was available and acceptable to the household), and the final interview was conducted in person. In each round of interviewing, questions were asked about the period since the last interview (except that in the first round, questions were asked about the period of time following January 1, 1980).

Collection of data from the households was facilitated by the use of a calendar and a summary. At the time of the first interview, the household respondent was given a calendar on which to record information about health problems and health services utilization and to use in assembling physician and other provider bills between interviews. Following each household interview, information about health provider contacts and the payment of charges associated with them was used to generate a computer summary of information provided. This summary was then printed out in a simple format and mailed to the household for review for accuracy and completeness prior to the next interview. At the subsequent interview, the interviewers reviewed this information with the household respondent to ensure accuracy and to obtain information not available during a previous interview.

In recognition of the commitment made by the household respondent to participate in the series of interviews and to review the computer-generated summary of utilization, charges, and sources of payment, a modest remuneration was provided. At the end of the first and second interviews, the household respondent was given $5, and at the end of the fifth interview, the respondent was given $10.

The administrative records survey of Cycle I was a check of the eligibility and claims records of persons reported as covered by Medicare or Medicaid. The supplementary and confirmatory data on Medicaid and Medicare enrollees provided information from the administrative records of the programs for comparison with that reported by the household respondents. The survey was designed specifically to meet the needs of the Health Care Financing Administration for programmatic data that are not otherwise available. For those individuals identified in the surveys as being enrolled in either the Medicare or Medicaid programs, sufficient identifying information was collected to allow the Health Care Financing Administration to flag incoming claims for services paid by Medicare, to allow the State Medicaid agencies in the four States in the State Medicaid household survey to abstract data from claims, and to permit all State Medicaid agencies to confirm eligibility and type of eligibility under their State Medicaid program.

Content of questionnaires

Questionnaires for the household surveys were designed to obtain some information on a repeated basis throughout the survey and some information only one time. The repetitive core of questions for Cycle I included health insurance coverage, episodes of illness, and the number of bed days, restricted-activity days, hospital admissions, physician and dental visits, other medical care encounters, and purchases of prescribed medicines. For each contact with the medical care system, data were obtained on the nature of the health conditions, characteristics of the provider, services provided, charges, and sources and amounts of payment. Questions asked only once included data on access to medical care services, limitation of activities, occupation, income, and other sociodemographic characteristics.

NMCUES is a unique source of extremely detailed health care utilization and financing information that will enable detailed measurement of utilization and expenditures for different subgroups of the population. Specific items of service are identified that can be related to chronic and acute illness, disability, work-loss days, and limitation of activity. Such data are essential to the measurement of the impact of legislation and health insurance programs on the charges for and financing of health services and insurance. This survey will be a principal source of ongoing data needed to monitor changes over time in the impact on individuals and families of private and public programs. Data from Cycle I of the survey are being published in a special series of joint reports from the Health Care Financing Administration and NCHS. The questionnaires and survey procedures have been published (16).

Data processing and quality control methods

Data collection and processing phases incorporated a number of quality control activities, including a detailed edit by each field interviewer and, immediately upon receipt at the central office, a premachine edit of all key linkage variables. Errors discovered in this premachine edit and in further edit of a sample of all items provided feedback to interviewers and supervisors.

Illness conditions, operations performed, specialties of medical providers not precoded on the questionnaire, and the industries in which persons were employed were coded prior to data entry. Quality control samples provided continuous monitoring of each coder, with retraining of coders and recoding of data, as appropriate, to ensure high levels of accuracy. Data entry through key-to-disk minicomputers provided a level of editing for valid ranges and codes. Check digits embedded in key identifying numbers kept miskeyed linkage numbers to an absolute minimum.

National Health and Nutrition Examination Survey

Background

The first National Health and Nutrition Examination Survey, referred to as NHANES I, was initiated in 1970, with data collection beginning in April 1971. NHANES I was a modification and expansion of the Health Examination Survey (HES), which had been initiated a decade earlier and carried out as three separate programs. The restructuring and modification of HES reflected the assignment to NCHS of an additional specific responsibility—the measurement of the nutritional status of the population and subsequent monitoring of changes in that status over time. On completion of NHANES I, data were collected for the second National Health and Nutrition Examination Survey (NHANES II), which began in February 1976 and ended in February 1980. A special health and nutrition examination survey directed to persons in families having one or more members of Hispanic origin or descent and living in areas with a high concentration of Hispanics was conducted for the period between the completion of NHANES II and the beginning of NHANES III. Pretesting for the Hispanic Health and Nutrition Examination Survey occurred in 1981 and 1982; the main survey was conducted from July 1982 through December 1984.

Planning has begun for NHANES III, which is scheduled to begin in 1988. NHANES III is expected to be conducted over a 6-year period, with a sample of approximately 20,000 persons every 2 years. This would permit national estimates every 2 years, estimates for black persons within 4 years, and estimates for Hispanics in 6 years.

Purpose and scope

NHANES and its predecessor program, HES, share a common purpose—the collection and utilization of data that can be obtained best or only by direct physical examination, clinical and laboratory tests, and related measurement procedures. This information, which cannot be furnished by the people themselves or by the health professionals who provide their medical care, is of two kinds. Prevalence data are collected for specifically defined diseases or conditions of ill health, and normative health-related measurement data are collected that show distributions of the total population with respect to particular parameters, such as blood pressure, visual acuity, or serum cholesterol level.

In the surveys, probability samples of the U.S. population are used to provide representative national data that are analyzed and made available in a series of reports. Successive surveys in the HES and NHANES programs have been directed to different segments of the population and have had different sets of target conditions. Thus, the first Health Examination Survey (first cycle) involved examining a sample of adults, with the focus primarily on selected chronic diseases. The second and third cycles of HES were directed to children ages 6–11 years and to youths ages 12–17 years, respectively. In both surveys, growth and development data and sensory defects were emphasized. The nutrition component of the first NHANES was directed to a probability sample of people in the broad age range 1–74 years; the detailed health examination component focused on the population ages 25–74 years. NHANES II was again directed to a broad population, 6 months–74 years, and the data on nutrition that were collected will be used in conjunction with the earlier NHANES I data to monitor changes in nutritional status over time.

Sample design

The samples for all of the HES and NHANES programs are multistage, highly clustered probability samples. All of the samples are stratified by broad geographic region and by population-density grouping. Within the strata, the sampling stages employed are the primary sampling unit, the census enumeration district, the segment, the household, and lastly the individual person. Until the household stage is reached, all sampling is carried out centrally in conjunction with the U.S. Bureau of the Census.

The next stage of the sampling is conducted in the field in the particular chosen area. It involves interviewer visits and questionnaire completion at each selected household, with the final selection of individuals included in the sample being dependent on information elicited by the household interview questionnaire. The size of the sample in the survey program has varied. In each of the three HES programs, the sample size was approximately 7,500 persons. In NHANES I, the sample selected for the major nutrition components of the examination contained approximately 28,000 people and yielded about 21,000 examined persons. A comparably sized sample for NHANES II again yielded approximately 21,000 examined persons.

Data collection procedures

Household interviewing during NHANES II was conducted by U.S. Bureau of the Census personnel, a departure from previous surveys, in which NCHS employees served this function. However, NCHS employees did the rest of the interviewing, as well as the history taking, examining, testing, and measuring in the mobile examination centers. Data collection teams consisted of specially trained interviewers and examiners, including physicians, nurses, dentists, dietitians, and medical, laboratory, and x-ray technicians.

NHANES examinations take place in the survey's specially constructed mobile examination centers, each consisting of three truck-drawn trailers. These trailers are interconnected and provide a standardized environment with equipment for the performance of specific parts of the examination. This standardized environment is necessary for such components of the examination as audiometry, which requires hearing chambers within which the ambient noise level conforms to the American Speech Association standards for acoustical measurements.

The general pattern of data collection has meant that each survey has been conducted over a period of 3 or 4 years. This results from the limitation on the number of persons examined in a given timespan (for example, the number of field teams and the number of sample areas). The kinds of data to be collected are also limited because conditions that might show marked year-to-year variation or seasonal patterns cannot be included. However, many important chronic diseases and health-related measurements are not subject to such changes in prevalence within shortrun periods. The distribution of the population according to unassisted visual acuity levels and the prevalence of such conditions as diabetes or hypertensive heart disease may vary over long periods of time, but not so rapidly as to prevent data collection over a 3- or 4-year period from giving a correct picture of the population levels during or at the midpoint of that period. It is hoped that, beginning with NHANES III, the data collection period can be shortened to 2 years for some topics and population groups, although longer periods of time may be necessary for some topics and special population groups.

Voluntary sample surveys present a problem if no data are collected on a large fraction of the selected sample because individuals selected to participate in the program are not willing to do so. In the HES and NHANES program, much attention has always been and continues to be devoted to the question of the response rate, the proportion of sample persons who are actually examined. In NHANES there have been, as anticipated, more problems in the area of response than had been encountered in the earlier HES programs. The difficulties faced have led to a variety of innovative measures, including a policy of remunerating examined persons. A token remuneration of $10 for persons completing the detailed examination was begun during NHANES I. During NHANES II, the remuneration was raised to $20, and that level was maintained for the Hispanic Health and Nutrition Examination Survey. Both NHANES programs succeeded in obtaining household interview data on about 99 percent of the sample population. More detailed health data are obtained by use of the medical history questionnaires; these were completed for 88 percent of the selected sample persons in NHANES I and for 91 percent of the selected sample persons in NHANES II. Finally, in NHANES I, 74 percent of the sample persons selected for the nutrition component and 70 percent of the persons selected for the detailed health component were given the standard examinations and tests; in NHANES II, the overall response rate for the examination component was 73 percent. There is considerable ancillary information on most of the persons in the sample population who were not examined, and it is possible to make use of that data in the process of imputation and analysis of nonresponse bias. There is, moreover, some evidence that data obtained through examinations, tests, and measurements such as those used in these surveys are less susceptible to potential bias from a given rate of nonresponse than are data provided by the individuals themselves.

Content of questionnaires, tests, and examinations

The kinds of information collected in NHANES and other examination survey programs are so varied and extensive that they are only illustrated here. With respect to nutrition, the following four types of data are included:

- Information concerning dietary intake—obtained from 24-hour recall interviews and food frequency questionnaires, both administered by an interviewer who is a trained dietitian.
- Hematological and biochemical tests—a sizable battery of such tests, with processing at the mobile examination centers where necessary but most processing at a central nutrition laboratory established at the Centers for Disease Control.
- Body measurements—an especially important battery in connection with infants, children, and youths, whose growth may be affected by nutritional deficiencies.

- Various signs of high risk of nutritional deficiency—based on clinical examinations.

The health (as distinguished from nutrition) component of the NHANES program includes detailed examinations, tests, and questionnaires that have been developed to obtain a measure of prevalence levels of specific diseases and conditions. These vary with the particular program and have included such conditions as chronic rheumatoid arthritis and hypertensive heart disease. Important normative health-related measurements, such as height, weight, and blood pressure, are also obtained.

A major element in the health component of NHANES I was an assessment, using index conditions, of unmet health needs. For the index conditions, the examination established whether emphysema or another chronic respiratory disease was present. At the same time, information was obtained from the examined person with respect to self-perceived health needs and actions taken in seeking medical care. Analyses included the interrelating of these two kinds of information to produce measures of unmet health care needs.

In NHANES II, the nutrition component remained nearly identical to that fielded previously. From an early analysis of the NHANES I data, it was decided to focus the nutritional examination elements around an anemia-related assessment approach. This involved the addition of certain medical history items and a more tailored set of laboratory determinations. Less was done in the area of health care needs. However, the emphasis placed on the effects of the environment on health will probably be continued in future programs. Data were gathered to measure the levels of pesticide exposure, the presence of certain trace elements in the blood, and the amounts of carbon monoxide present in the blood. In the medical area, primary emphasis in NHANES II was placed on diabetes, kidney and liver functions, allergy, and speech pathology.

The time required for the examination varies with the content of the examination and the age of the examinee. The time constraint included among planning factors has been that the total examination time not exceed 2–2½ hours. In the planning process and in the pretests of a survey, much attention is given to the actual flow of examinees through the examination center, and every effort is made to streamline this process in order to reduce the time burden on the sample person. Additional respondent burden arises from the interview, from the completion of forms and questionnaires in the household, and from the varying time required by the sample person to travel to the examination site. Data collection forms used in HES I, II, and III; NHANES I; the NHANES I Augmentation Survey; and NHANES II are reproduced in Series 1 reports numbered 4, 5, 8, 10b, 14, and 15, respectively (17–22).

Data processing and quality control methods

The data collected in NHANES require a variety of data processing methods. X rays must be interpreted; blood and other laboratory specimens must be processed through the appropriate laboratory operations. Certain data, such as electrocardiographic tracings, are recorded directly onto magnetic tape and must undergo appropriate processing to be translated into digital tape form and, subsequently, to be interpreted. Examination record forms and interview and questionnaire data must be coded and put onto magnetic tape.

Methods for handling specific examination elements are chosen on the basis of appropriateness. The program makes use of precoded forms, marked-sense record forms, self-administered forms, interviewer-administered forms, automatic recording devices, photographs, and so forth. Throughout whatever process is chosen, constant emphasis is placed on quality control measures, such as editing, verifying, and replicating. Finally, outside consultants frequently collaborate in the analysis and reporting of data.

NHANES I Epidemiologic Followup Survey

Background

The NHANES I Epidemiologic Followup Survey was established to make maximum use of detailed, baseline data obtained in the first National Health and Nutrition Examination Survey, conducted from 1971 through 1975, as described previously. The objectives of the Followup Survey were to identify chronic disease risk factors associated with morbidity and mortality; to ascertain changes in risk factors, morbidity, functional limitation, and institutionalization between the time of NHANES I and the followup recontacts; and to map the natural history of chronic diseases and functional impairments in an aging population. This survey was initially a joint project of the National Center for Health Statistics and the National Institute on Aging. Since its origin, several other institutes of the National Institutes of Health and the Alcohol, Drug Abuse, and Mental Health Administration have become involved in the development of hypotheses important to their specialty areas and in the design of the procedures for the collection of data to test these hypotheses. These cosponsors have also provided financial support for the survey.

A feasibility study conducted in 1980 indicated that sample persons examined in the NHANES I survey could be located after a 10-year interval and that they would authorize access to information from their medical records. Initial contact with the sample persons was made in 1983 and 1984. Pretesting has begun for further recontact with these individuals beginning in 1985.

Purpose and scope

Through the NHANES I Epidemiologic Followup Survey it is possible to investigate the relationships between physiological, environmental, nutritional, social, psychological, and demographic factors and the mortality and morbidity from specific diseases, as well as hospitalization and nursing home utilization. In the Initial Followup in 1983–84, information was obtained from in-person interviews with the subjects or a proxy. In addition to information about their health since the previous survey and about their current nutrition habits, information obtained included weight and blood pressure measurements, hospital and nursing home utilization, and copies of death certificates for those who had died. The primary purpose of the Continued Followup is to extend the period for which data about the subjects are available. This will provide a more complete profile of hospital and nursing home use for all respondents and will increase the age span of respondents covered by the study. Because NHANES I was restricted to people under 75 years of age, the Initial Followup extended only to age 85. The Continued Followup will provide important information about the surviving cohort as they enter extreme old age. It is hoped that it will be possible to follow these individuals indefinitely.

Sample design

The Followup Survey universe is the 14,407 persons who were aged 25–74 at the time they were interviewed and examined in the original NHANES I survey. These persons comprise a national sample of the civilian noninstitutionalized population (after adjustment to account for nonresponse). The followup includes all these persons, and proxy information is sought for those who are deceased or unable to respond. It was possible to trace over 92 percent of the original NHANES I participants. Of those traced, it was possible to obtain information for the Initial Followup for over 94 percent, so that the final response rate was approximately 85 percent.

Data collection procedures

This survey is conducted by a private contractor, following specifications set by NCHS. An advance introductory letter from NCHS was sent to all sample persons, informing them about the intent of the study. A brochure explaining the study in more detail was included with the letter. Following the mailing of the letter, an interviewer telephoned each individual to set an appointment for the household interview, which was conducted face to face in the respondent's home.

Data collection was carried out by region, with interviewing beginning in the Northeast in late 1982, in the South beginning in January 1983, in the Midwest beginning in April 1983, and in the West beginning in October 1983. Hospital and nursing home data collection began about 6 months after the initiation of household interviews in each region and continued over about a 6-month period. Tracing continued throughout the survey period, which ended in July 1984.

A brief mental status questionnaire was included in order to screen for dementia; for persons judged not suited to participate in the interview, response was sought for many items from a proxy. When a proxy respondent was needed for someone who was deceased or incompetent, an effort was made to identify the person who had been living with the sample person most recently. The order of priority for a proxy was spouse, children, other relatives, companion, nurse, or friends. Only factual and historical questions were asked of proxy respondents.

For the Followup Survey initial contact, it was possible to locate or determine the vital status of all but approximately 1,000 of the NHANES I examinees. Only 25 individuals firmly refused to participate in the initial contact interview. Survey participants were given a $10 payment for participating in the interview and for signing a consent form to permit access to their medical records. This payment was made in recognition of the significant amount of time required for the interview (about 2 hours) and of the fact that examinees in NHANES I were paid $10 as remuneration for their time and inconvenience.

In the Continued Followup, study participants will also be sent an advance letter to inform them that an interviewer will be contacting them. This interview, however, will be conducted by telephone for persons having a telephone; persons without a telephone will be sent a self-administered questionnaire to complete and return. As in the Initial Followup, information from proxy respondents will be sought in the Continued Followup for those who have died in the interim or who are incapacitated such that they are unable to participate in a telephone interview. The interview is expected to be 15–20 minutes in length.

In the Continued Followup, efforts will continue to locate individuals who were not located during the Initial Followup period. Respondents will again be asked to sign a release to authorize access to information in their medical records in hospitals and other health care facilities, and these providers will be contacted to obtain information about health care episodes. A copy of the death certificate for each known decedent will be obtained from the appropriate State vital statistics office.

Data collection in the Continued Followup will again be carried out by region, in the same sequence as for the Initial Followup. In each region the bulk of the telephone interviews will take approximately 2 months to complete; an additional 6 months will be required for data collection from medical providers.

It is planned that the elderly, those 75 years and over, will be contacted each year and that those under 75 years will be contacted every other year. These time intervals were selected in order to minimize the costs of tracing and relocating respondents and to maximize the accuracy of recall information.

Content of questionnaires

The Initial Followup interview questionnaire drew from questions and scales that had been used extensively in other surveys. About 2 hours were required for the household interview. Questionnaire topics included family history of cancer; a health history pertaining to heart disease; questions on stroke and other neurological symptoms, cancers, and respiratory problems; detailed questions on arthritis; questions on functional impairment; a medical conditions checklist; smoking history; questions on lifetime pattern of usual drinking; questions to identify sleep disorders; questions on current use of vitamins and prescription medications and on physical activity; and questions about current occupation, income, and the occupation of longest duration. Nutrition data were gathered using an expanded food frequency questionnaire developed to best meet the needs of this survey. To facilitate subsequent followup through the National Death Index (discussed later), each respondent was asked to provide his or her social security number. Physical measurements taken by the interviewers were limited to pulse, weight, and blood pressure. As in NHANES, quality control measures to monitor the performance of the measurement equipment and of the interviewers were put in place. Respondents also signed an authorization form to permit obtaining information from their medical records in hospitals or other health care facilities.

The main purpose of the hospital record-check procedure was to validate the respondent's reporting of aspects of the medical history. For each episode of overnight hospitalization since 1970, a respondent was asked to provide the name and address of the hospital and approximate date of hospital discharge. Hospitals were asked to provide information on all admissions of the patient since January 1971. Photocopies of hospital admission and discharge notes were obtained. For patients reporting myocardial infarction, a copy of the electrocardiogram taken the third day after myocardial infarction was requested to validate self-reports of cardiovascular disease. Information was obtained on principal diagnosis and other diagnoses to estimate rates of disease, and information on surgical procedures was obtained to validate certain diagnostic information. A similar medical record form was used to request a subset of this information from other inpatient medical facilities such as nursing homes. Dates of admission and discharge and diagnosis (or reason for admission) were requested of all facilities.

When it was determined that a sample person had died, the vital registration office in the State of death was contacted in order to obtain a copy of the death certificate. The cause of death was coded by NCHS using the same procedures as are used in the national mortality data program.

The questionnaire for the Continued Followup will include questions on functional impairment, change in personal health practices, preventive health practices, hospitalizations, and changes in sociodemographic status. Information to be requested from health care institutions will be the same as that in the Initial Followup.

Analysis of data from the survey is a collaborative activity of staff of NCHS, the National Institute on Aging, and other sponsoring institutes and agencies. A background paper describing the 1982–84 survey activities has been published (23), and a report containing the questionnaires and detailing the survey operation is being prepared (24).

Data processing and quality control methods

Several measures were included in the Initial Followup to maintain quality in the data collection and processing activities. Interviewers were observed and evaluated by a traveling supervisor in each region. Completed questionnaires were edited by the interviewer and again in the field office. Feedback from the editing process was provided by the supervisor to the interviewer. A 15-percent sample of each interviewer's work was verified in a brief telephone reinterview to assure that the interview was actually conducted and to check the reliability of a few items in the questionnaire. When a respondent did not have a telephone, the reinterview was carried out by mail.

Interviewers received an intensive 3-day training session on physical measurement preceding the general training on the conduct of the interview. Experience of sources of error from prior surveys and screening programs was used to develop training approaches. Only interviewers who successfully completed training in taking physical measurements were employed in the survey. The blood pressure equipment and the scale used to weigh sample persons were selected for reliability and durability under field conditions. Each piece of equipment was calibrated and prepared for field use by a technician. Interviewers were trained to check their equipment daily, and equipment was thoroughly reconditioned by a technician when interviewing was completed in each region. Blood pressure data were monitored for potential problems such as digit preference and extreme values.

After receipt of the survey interviews by the contractor, careful editing and coding were completed. Specialized coders coded the industry and occupation codes and medical diagnoses. Range and logic checks were incorporated in the computer data processing.

In the Continued Followup, computer-assisted telephone interviewing (CATI) will be used. The CATI system will bring the proper question to the screen for the interviewer to read to the respondent. The programming will be designed so that the interviewer will be provided with automatic prompts depending on the circumstances; for example, a proxy will have different wording than a participant. Edit checks will be built into the CATI programming. In addition, interviewing will be monitored and evaluated by a supervisor.

National Survey of Family Growth

Background

The National Survey of Family Growth (NSFG) is a multipurpose survey that provides a wide range of information serving needs of persons and organizations concerned with the dynamics of population change, family planning, and maternal and child health. Developmental funds and necessary positions were provided for establishing NSFG in NCHS in fiscal year 1971. Fieldwork for the first cycle of the survey was begun in July 1973 and completed in January 1974. The second cycle began in January 1976 and was completed in September 1976. A third cycle was conducted from September 1982 through February 1983. Planning has begun for Cycle IV, to be conducted in 1987.

Purpose and scope

NSFG is designed to produce data on factors influencing trends and differentials in fertility, family planning practices of the population, sources from which family planning advice and services are obtained, the effectiveness and acceptability of the various methods of family planning, and those aspects of maternal and child health that are most directly related to fertility and family planning. The survey is based on personal interviews with women of childbearing age selected from a nationwide area probability sample of households and group quarters, exclusive of the institutionalized population and women living on military bases.

Sample design

Cycles I and II of the National Survey of Family Growth were based on a cross-sectional sample of women in the conterminous United States, 15–44 years of age, who were married, had been married, or had never been married but had offspring of their own living with them in the household. Excluded from the sample were women living in group quarters (that is, five or more unrelated persons 18 years or older who are unrelated to the head of the household and who live and eat together, or six or more unrelated adults who live and eat together—as, for example, in college dormitories, barracks, or long-term care institutions), and never-married women without children.

NSFG is designed as a multistage area probability sample. In the first cycle of this survey, the first-stage primary sampling units included 101 standard metropolitan statistical areas, counties, parts of counties, and independent cities. Secondary sampling units consisted of enumeration districts or block groups within selected primary sampling units. Where feasible, secondary units were subdivided into third-stage listing units with a probability of selection proportional to estimated housing. In the fourth stage, housing units within the third-stage listing units were chosen by systematic sampling. A fifth level of sampling among eligible women within a household was required to obtain one and only one interview in each household with at least one eligible respondent. To assure sufficient precision for presenting intragroup comparisons on various fertility variables, the black population was oversampled.

The overall sample design for Cycle I called for 10,000 completed interviews, including approximately 4,000 black women and 6,000 women of white and other racial groups. The overall response rate was 81 percent. The final number of interviews, 9,797, consisted of interviews with 3,856 black respondents and 5,941 respondents of white and other races.

For Cycle II of NSFG, a multistage area probability design based on 79 primary sampling units was developed. The remaining stages were broadly similar to those of the first cycle, except that a stratum for new housing (housing built since 1970), sampled from building permit listings, was added. Overall expected sample size remained at 10,000, with comparable racial composition. After a certain level of effort was completed in the fieldwork, remaining nonrespondents were subsampled at a rate of 1 in 2. The response rate for Cycle II was an overall 83 percent. The actual number of completed extended interviews was 8,611—2,946 interviews with black respondents and 5,665 interviews with respondents of white and other races.

For Cycle III, the sample design was broadly the same as for Cycle II but with adjustments to accommodate the more inclusive universe of women ages 15–44 years of any marital status. The sample of 7,600 women included women in group quarters such as college dormitories and boarding houses with more than five unrelated persons. As in the previous cycles of the survey, military bases and the institutionalized population were not sampled.

Data collection and quality control procedures

Data collection for NSFG is conducted by private contractors according to NCHS specifications. Personal indepth interviews are conducted with women identified as eligible for extended interview by means of a household screening interview.

The questionnaires used for data collection include a household screener designed to obtain household composition data and to identify eligible extended-interview respondents in the sample households. For Cycles I and II, two different versions of the questionnaire were used for the extended interview—a currently married questionnaire for women who were married at the time of the interview and a postmarried questionnaire for women who were widowed, divorced, separated, or who had never been married but had their own children living in the household. The differences between the two versions were a result primarily of rewording the questions and of deleting questions related to husbands in order to make the postmarried questionnaire appropriate for respondents not married at the time of the interview. Throughout the questionnaires, instructions were used to skip interviewers over questions or entire questionnaire sections that were not applicable to the individual situations of respondents. Cycle III of the survey involved a screener questionnaire similar to that used in the previous cycles and two separate questionnaires for the extended interviews, one for women under age 25 and one for women 25 years of age and over. The construction and format of the questionnaires were modeled on those of the previous cycles.

Standardized interviewer training programs, which provide indepth training on the questionnaires and on NSFG concepts and procedures, are conducted in several different locations in the country. In addition to successfully completing training, interviewers are required to conduct, for review and approval, several practice interviews prior to beginning their field assignments. Only women are employed as interviewers and observers in NSFG. Field supervisors receive the same indepth training as the interviewers, in addition to intensive training in field procedures.

Several quality control procedures were developed in Cycle I and refined in Cycle II to ensure the quality of the collected and processed data; a systematic field edit of selected interview items and a validation of a sample of each interviewer's work (a sample recheck) were performed throughout the fieldwork. Data preparation was validated through a 5-percent sample recode of all questionnaires, in addition to systematic verification of each coder's or keyer's work. A comprehensive legal-code and consistency cleaning program was developed and used for the data tapes. Similar quality control procedures were specified for Cycle III.

Content of questionnaires

In Cycles I and II, the NSFG questionnaires covered marital history, a detailed pregnancy history, fecundity and expected or intended future births, pregnancy planning practices and utilization of specific contraceptive methods, the source and financing of family planning services, and a broad range of socioeconomic and demographic characteristics. In Cycle III, information was collected on sex education and the sexually active population as well as on the topics covered in Cycles I and II.

In the first cycle of the survey, the average length of an interview was 73 minutes; in the second cycle, the average length of an interview was reduced to about 56 minutes; in the third cycle, interviews were again kept within an hour on the average.

A copy of the Household Screener questionnaire for Cycle I is contained in appendix III of Series 2, Number 76 (25); the screener for Cycle II is in appendix II of the comparable report for Cycle II, Series 2, No. 87 (26). Relevant portions of the extended interview questionnaires are reproduced in appendixes of Series 23 reports. The currently married questionnaire for Cycles I and II is reproduced in appendix III of Series 4, No. 18 (27).

Recent and possible future activities

A number of recent or potential future activities related to the NCHS population-based survey program are of note because they affect how these surveys are or may be conducted. Each of these activities is discussed briefly in this section.

National Death Index

The National Death Index (NDI) is a central, computerized index of death record information obtained by NCHS from the States on magnetic tape under contractual arrangements with the State vital statistics offices. The tapes contain a standard set of identifying information about each decedent, beginning with deaths occurring in 1979. This file is designed to assist in the mortality ascertainment activities of investigators conducting prospective studies in health and medical research.

Investigators use the NDI to determine whether persons in their studies may have died. If so, they are provided with the names of the States in which those deaths occurred and the corresponding death certificate numbers. An NDI user then arranges to procure copies of death certificates from the State vital statistics offices in order to obtain such statistical information as cause of death.

Prior to the establishment of the NDI, investigators often had to contact all or most State vital statistics offices, asking each to search its files to see if a death record may have been filed for any individual in the entire study group. The large size of many study groups and the time required for the States to complete their file searches meant that considerable time and effort were required to determine the vital status of study subjects. The NDI facilitates prospective medical and health research studies by reducing the time, expense, and effort involved in State file searches.

NDI files are updated annually. All State data for a given calendar year are usually received, processed, and added to the national file within 12–14 months after the end of the calendar year.

Investigators wishing to use the NDI service must submit an application to NCHS for review by staff and consultants, who include representatives from State vital statistics offices, the National Institutes of Health, the National Institute for Occupational Safety and Health, and schools of medicine and public health. The review is intended to ensure that the proposed use of information obtained from the NDI (as well as any information subsequently acquired from State vital statistics offices) conforms with NCHS confidentiality requirements and the contractual agreements between NCHS and the State vital statistics offices and that the potential user's data are technically acceptable.

Users are encouraged to provide as many as possible of the following NDI data set items (28): *last name, *first name, middle initial, *social security number, *month and year of birth, day of birth, *father's surname (primarily for females), age at death (actual or estimated, if known), sex, race, marital status, State of residence (last known State of residence), State of birth. A user must be able, at a minimum, to provide first and last names *and* social security number *or* month and year of birth.

Items marked with asterisks are key items used in the searching procedure to determine whether a particular death record in the NDI file constitutes a possible match with a given user record. The remaining items are matched against corresponding items in the NDI file when a possible record match occurs on the basis of some combination of the key data set items. Matches or nonmatches on items not used in the searching procedure enhance a user's ability to assess the quality of each possible record match, especially when more than one possible record match is generated by a particular user record.

Recently, NCHS has begun to include all of the NDI items in its population-based survey questionnaires. In most instances this means adding only a few items (usually the social security number, State of birth, and father's surname) to the questionnaire. For example, the NDI items were included in the Hispanic Health and Nutrition Examination Survey and in the NHANES I Epidemiologic Followup Survey, and as of 1985, the NDI items are part of the core questionnaire used in the National Health Interview Survey for each person 18 years of age and over. Separate data tapes are being created for later NDI searches as part of the data processing for these surveys.

Integrated survey design

Traditionally, the samples for various NCHS population surveys have been independently designed and selected. The sample design for the National Health Interview Survey has been reevaluated and updated following each decennial census of the population to reflect changes in characteristics of the population and new statistical methodologies. The fact that NHIS is a continuous survey with a large sample makes it a possible sampling frame for the periodic population surveys. However, confidentiality restrictions associated with the fact that the sample was derived from lists of addresses compiled during the decennial census prohibit this type of linkage of NCHS surveys.

As part of the redesign following the 1980 census, it was decided to use an area sample (rather than the list frame used by the Census Bureau in surveys it conducts for other Federal agencies) and to explore whether using NHIS as a frame from which to select samples for the other NCHS population surveys was advantageous with respect to costs, sampling error, nonresponse, respondent burden, and data utility (29). The NHIS sample design implemented in 1985 is comprised of four panels, one or more of which can be used as a sampling frame for other surveys.

Many of the conceptual and statistical issues related to integrating the designs of the National Survey of Family Growth and the National Medical Care Utilization and Expenditure Survey with NHIS have been investigated, and it is clear that significant cost savings can be realized for NSFG and also for NMCUES if certain subgroups of the population are oversampled in NMCUES. The practical feasibility of linking the surveys must be investigated also, however. Questions such as the following have recently been addressed for NSFG and will soon be addressed for NMCUES: Should the sampling unit be a dwelling unit (regardless of whether the persons living there at the time of the NHIS interview are still there) or residents of the unit (regardless of where they live at the time of recontact)? Over what period of time should the sample be accumulated?

It is expected that the next cycle of NSFG and a NMCUES-like survey will be conducted with samples selected from the NHIS sample. Because the investigation of issues associated with linking the National Health and Nutrition Examination Survey to NHIS has not begun, NHANES III will not be linked to the NHIS sample.

Telephone surveys linked to NHIS

An additional possibility under the new NHIS sample design is that of using NHIS as a frame from which to select samples for telephone surveys designed to provide data on particular topics in a relatively short period of time. The fact that the names, addresses, and telephone numbers as well as health, demographic, and socioeconomic data collected in NHIS would be available makes this a potentially efficient way to target surveys toward specific health issues or specific population subgroups.

A contract was recently awarded to the Survey Research Center at the University of Michigan to investigate the feasibility and potential problems associated with fielding a linked telephone survey and producing needed estimates within 3–6 months after identification of a survey topic. A particular focus of the study is the effect of sampling frame deterioration, choice of sample unit, and method of initial contact on the cost, turnaround time, and statistical integrity of the linked telephone survey. The project will also provide information on the time needed to design a computer-assisted telephone interview questionnaire, the time needed to develop and test computer edit specifications, the time needed to produce estimates, and possible modifications to current NHIS operating procedures that would facilitate more rapid telephone survey sampling and data collection. Findings will be compared with other possible methods of collecting data on a rapid turnaround basis, such as random-digit-dialed telephone surveys.

Random-digit-dialed telephone surveys

Since 1977, the National Center for Health Statistics has engaged in an extensive program of research and development in random-digit-dialed sampling and telephone interview procedures (30–32). The lower expected costs of telephone surveys, along with the increased ability to monitor interviewers' work when a centralized telephone interviewing facility is used, led NCHS staff to embark on a program to study the potential benefits and problems associated with use of the telephone as a data collection mechanism, either as an enhancement to the National Health Interview Survey or for special separate surveys.

Table 1. Profile of NCHS population-based surveys

Survey	Periodicity	Population covered	Data collection methodology	Sample design
National Health Interview Survey	Continuous since 1957; supplements to core questionnaire change each year or more frequently	Civilian noninstitutionalized population of the 50 States and District of Columbia	Initial household interview conducted face to face with an adult household respondent; supplements may require self-response (face to face or by telephone) from a sample person	Multistage probability sample; each week's sample is broadly representative of the target population and weekly samples are additive over time 1973–84: 376 primary sampling units; based on 1970 census registers plus area segments and permit segments (using building permits issued since 1970); 42,000 households containing 110,000 persons 1985–present: 201 primary sampling units; based completely on independent listing of areas in the survey sample; 50,000 households containing approximately 135,000 persons; because of budget constraints, only a 3/4 sample was implemented for 1985
National Medical Care Utilization and Expenditure Survey	Approximately quinquennial First cycle conducted in 1980; next cycle planned for 1987	Civilian noninstitutionalized population of the 50 States and District of Columbia in national sample; Medicaid-eligible households in New York, California, Texas, and Michigan	3 face-to-face and 2 telephone interviews with an adult household respondent; review of administrative eligibility and claims records for persons reported to be covered by Medicare and Medicaid	Multistage area probability sample drawn from 106 primary sampling units for national sample of 10,000 households; probability samples of about 1,000 households from State Medicaid eligibility lists in each of the 4 States
National Health and Nutrition Examination Survey	New cycle begins approximately every 4 years; Cycle I, 1971–75; Cycle II, 1976–80; Cycle III planned to begin in 1988	Civilian noninstitutionalized population of the conterminous United States in Cycle I; expanded to cover all 50 States in 1976 Cycle I—persons ages 1–74 years; Cycle II—persons ages 6 months–74 years	Face-to-face interview with an adult household respondent to determine household membership, obtain background information; and select sample person(s); face-to-face household interview with sample person; detailed tests, examinations, and dietary interview for sample person in survey's mobile examination center	Multistage probability sample; Cycle I used 100 primary sampling units and examined approximately 21,000 persons in the detailed nutrition components; Cycle II used 64 primary sampling units and examined approximately 21,000 persons
NHANES I Epidemiologic Followup Survey	Initial followup in 1983–84; recontact planned on annual basis for persons 75 and over and biennially for those under 75	All 14,407 persons who were ages 25–74 at time of examination in NHANES I	Initial followup was face-to-face interview plus blood pressure and weight measurements; proxy interview for those who had died or were incapacitated; continued followup for 1985–86 by telephone; signed authorizations obtained at each contact to permit obtaining information from hospital and nursing home records	Efforts made to trace all 14,407 persons in the universe; located more than 92 percent; effort continues to locate persons lost to followup
National Survey of Family Growth	Approximately triennial Cycle I, 1973–74; Cycle II, 1976; Cycle III, 1982–83; Cycle IV planned for 1987	Cycles I and II: Ever-married women and never-married women with own child living in the household, excluding women in group quarters such as college dormitories; Cycle III: All women ages 15–44 regardless of marital or childbearing status, including women in group quarters	Female interviewers visit sample addresses to enumerate household members and select sample women for extended interview; face-to-face interview conducted in strict privacy with sample women; only 1 extended interview conducted in a household; parental permission obtained before minors are asked to participate in the survey	Multistage area probability sample; black women are oversampled; Cycle I used 101 primary sampling units and obtained 9,797 extended interviews; Cycle II used 79 primary sampling units and obtained 8,611 extended interviews; Cycle III used 79 primary sampling units and obtained 7,969 extended interviews

NCHS has used a random-digit-dialing (RDD) procedure developed by Waksberg to select the samples for its studies. This RDD procedure permits the interviewer to dial a randomly selected telephone number and conduct an interview without prior knowledge of who the respondent is or where the home is located. It also permits the probabilities of selection to be estimated by determining the number of telephone numbers for each household and the number of persons who use each of the separate telephone numbers.

Methodological studies have been completed that tested and demonstrated the ability to administer NHIS supplements by telephone, the comparability of estimates derived from a RDD telephone survey with those from the NHIS face-to-face interviews, alternate approaches to administration of the NHIS core questionnaire, and use of computer-assisted telephone interviewing to conduct the interview and enter the data directly into the computer. These studies have shown the following:

- Telephone interviewing and RDD sampling can be substantially less costly than the face-to-face interview (although the amount of savings varies by the type of application and by the organization collecting the data).
- Some problems remain because not all households have a telephone, but problems can be reduced or adjustments made through statistical techniques.
- For RDD surveys on health issues with a Federal agency as the sponsor and a "knowledgeable adult" as the respondent, overall response rates of up to 85 percent can be achieved.
- For a well-designed interview, there are no major differences between data collected via telephone and data from face-to-face interviews.
- It is feasible to conduct the entire NHIS interview by telephone.
- Use of CATI enhances both the quality of data collection and the timeliness of data dissemination.

In recognition of the findings of these studies, serious consideration was given to using a dual-frame approach for the redesigned NHIS sample that was implemented in 1985. The advantage of a dual-frame dual-methodology approach is that coverage problems are minimized while design efficiency is maintained. However, the decision to link the other NCHS population-based surveys to NHIS led to the conclusion that it would be preferable to conduct all NHIS interviews face to face.

Summary

This paper has provided an overview of the current population-based survey program of the National Center for Health Statistics. The description of each survey is necessarily brief. Table 1 provides a profile of each of the four surveys discussed. Additional technical information about each survey is available in published reports containing data from the survey.

Additional analytic potential for data from the surveys should be realized as a result of the recent addition to the survey questionnaires of the data items needed for matching with the National Death Index. Linkage to the National Health Interview Survey of the samples for the other population-based surveys and continued exploration of the possibility of conducting rapid-turnaround telephone surveys of special population groups using the NHIS sample or independent random-digit-dialing telephone surveys are also probable future enhancements to the NCHS program.

In considering ways in which data on physical fitness and activity patterns might be collected through any of these surveys, it will be important to keep in mind the environment in which NCHS data collection activities are planned and carried out. Numerous reviews and approvals are required, and often they place constraints on what data are collected and how the information is obtained. There may, however, be more flexibility in what can be done and how it can be done than appears at first look. An optimistic, open-minded approach should always be taken.

References

1. Office of the Assistant Secretary for Health and the Surgeon General: Promoting Health/Preventing Disease—Objectives for the Nation. Washington. U.S. Government Printing Office, Fall 1980.
2. National Center for Health Statistics: NCHS Staff Manual Guide, General Program Policy, No. 2—The NCHS Technical Review Program. Public Health Service. Hyattsville, Md. May 18, 1982.
3. National Center for Health Statistics: NCHS Staff Manual Guide, General Program Policy, No. 1—Protecting Human Subjects of Research Study and Data Collection. Public Health Service. Hyattsville, Md. Aug. 31, 1971.
4. Office of Federal Statistical Policy and Standards: Statistical Policy Handbook. Department of Commerce. Washington. U.S. Government Printing Office, May 1978.
5. Health Data Advisory Committee: Standards and Guidelines for Administrative and Statistical Data Collection in Health Programs of the Department of Health and Human Services, Issuance No. 1. Department of Health and Human Services. Washington. July 1981.
6. National Center for Health Statistics: *NCHS Policy Statement on Release of Data for Individual Elementary Units and Special Tabulations.* Public Health Service. DHEW Pub. No. (PHS) 78–1212, Hyattsville, Md., May 1978.
7. National Center for Health Statistics: *NCHS Staff Manual on Confidentiality.* DHHS Pub. No. (PHS) 84–1244. Public Health Service. Hyattsville, Md. Sept. 1984.
8. National Center for Health Statistics: *Eighth Revision International Classification of Diseases, Adapted for Use in the United States.* PHS Pub. No. 1693. Public Health Service. Washington. U.S. Government Printing Office, 1967.
9. World Health Organization: *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death.* Based on the Recommendations of the Ninth Revision Conference, 1975. Geneva. World Health Organization, 1977.
10. Public Health Service and Health Care Financing Administration: *International Classification of Diseases, 9th Revision, Clinical*

Modification. DHHS Pub. No. (PHS) 80–1260. Public Health Service. Washington. U.S. Government Printing Office, Sept. 1980.

11. National Center for Health Statistics: *Catalog of Publications of the National Center for Health Statistics 1979–83.* DHHS Pub. No. (PHS) 84–1301. Public Health Service. Hyattsville, Md. June 1984.
12. National Center for Health Statistics: *Catalog of Public Use Data Tapes From the National Center for Health Statistics.* DHHS Pub. No. (PHS) 81–1213. Public Health Service. Hyattsville, Md. Nov. 1980.
13. National Center for Health Statistics: Health Interview Survey procedure, 1957–74. *Vital and Health Statistics.* Series 1, No. 11. DHEW Pub. No. (HRA) 75–1311. Health Resources Administration. Washington. U.S. Government Printing Office, Apr. 1975.
14. National Center for Health Statistics, M. G. Kovar and G. Poe: The National Health Interview Survey Design 1973–1984 and Procedures 1975–1983. *Vital and Health Statistics.* Series 1, No. 18. DHHS Pub. No. (PHS) 85–1320. Public Health Service. Washington. U.S. Government Printing Office. In press.
15. National Center for Health Statistics: Current Estimates from the National Health Interview Survey. *Vital and Health Statistics.* Series 10. Public Health Service, DHHS, Hyattsville, Md. To be published.
16. National Center for Health Statistics, G. S. Bonham: Procedures and questionnaires of the National Medical Care Utilization and Expenditure Survey. *National Medical Care Utilization and Expenditure Survey.* Series A, Methodological Report No. 1. DHHS Pub. No. 83–20001. Public Health Service. Washington. U.S. Government Printing Office, Mar. 1983.
17. National Center for Health Statistics: Plan and initial program of the Health Examination Survey. *Vital and Health Statistics.* Series 1, No. 4. PHS Pub. No. 1000. Public Health Service. Washington. U.S. Government Printing Office, July 1965.
18. National Center for Health Statistics: Plan, operation, and response results of a program of children's examinations. *Vital and Health Statistics.* Series 1, No. 5. PHS Pub. No. 1000. Public Health Service. Washington. U.S. Government Printing Office, Oct. 1967.
19. National Center for Health Statistics: Plan and operation of a health examination survey of U.S. youths 12–17 years of age. *Vital and Health Statistics.* Series 1, No. 8. PHS Pub. No. 1000. Public Health Service. Washington. U.S. Government Printing Office, Sept. 1969.
20. National Center for Health Statistics: Plan and operation of the Health and Nutrition Examination Survey, United States, 1971–73. *Vital and Health Statistics.* Series 1, No. 10b. DHEW Pub. No. (HSM) 73–1310. Health Services and Mental Health Administration. Washington. U.S. Government Printing Office, Feb. 1973.
21. National Center for Health Statistics, A. Engel, R. S. Murphy, K. Maurer and E. Collins: Plan and operation of the HANES I Augmentation Survey of Adults 25–74 Years, United States, 1974–75. *Vital and Health Statistics.* Series 1, No. 14. DHEW Pub. No. (PHS) 78–1314. Public Health Service. Washington. U.S. Government Printing Office, June 1978.
22. National Center for Health Statistics, A. McDowell, A. Engel, J. T. Massey and K. Maurer: Plan and operation of the Second National Health and Nutrition Examination Survey, 1976–80. *Vital and Health Statistics.* Series 1, No. 15. DHHS Pub. No. (PHS) 81–1317. Public Health Service. Washington. U.S. Government Printing Office, July 1981.
23. J. Cornoni-Huntley, H. E. Barbano, J. A. Brody, et al.: National Health and Nutrition Examination I—Epidemiologic Followup Survey. *Public Health Reports* 98(3): 245–251, May–June 1983.
24. National Center for Health Statistics, B. Cohen: Plan and operation of the NHANES I Epidemiologic Followup Survey. *Vital and Health Statistics.* Series 1, Public Health Service, DHHS, Hyattsville, Md. To be published.
25. National Center for Health Statistics, D. K. French: National Survey of Family Growth, Cycle I, sample design, estimation procedures, and variance estimation. *Vital and Health Statistics.* Series 2, No. 76. DHEW Pub. No. (PHS) 78–1350. Public Health Service. Washington. U.S. Government Printing Office, Jan. 1979.
26. National Center for Health Statistics, W. R. Grady: National Survey of Family Growth, Cycle II: sample design, estimation procedures, and variance estimation. *Vital and Health Statistics.* Series 2, No. 87. DHHS Pub. No. (PHS) 81–1361. Public Health Service. Washington. U.S. Government Printing Office, Feb. 1981.
27. National Center for Health Statistics: Statistics needed for national policies related to fertility, a report of the United States National Committee on Vital and Health Statistics. *Vital and Health Statistics.* Series 4, No. 18. DHEW Pub. No. (HRA) 78–1455. Health Resources Administration. Washington. U.S. Government Printing Office, Jan. 1978.
28. National Center for Health Statistics: User's manual, the National Death Index. DHHS Pub. No. (PHS) 81–1148. Public Health Service. Hyattsville, Md. Sept. 1981.
29. M. G. Sirken and M. S. Greenberg: Redesign and integration of a population-based health survey program. Paper presented at the 44th session of the International Statistical Institute. Madrid. Sept. 12–22, 1983.
30. J. Fitti: Some results from the telephone health interview system. *Proceedings of the American Statistical Association Section on Survey Research Methods, 1979,* 244–249.
31. National Center for Health Statistics: An experimental comparison of telephone and personal health surveys. *Vital and Health Statistics.* Series 2. Public Health Service, Hyattsville, Md. To be published.
32. U.S. Bureau of the Census and National Center for Health Statistics: Results of the 1984 NHIS/RDD feasibility study, final report. Unpublished document. Feb. 28, 1985.

Chapter **2**

Cardiovascular Endurance, Strength, and Lung Function Tests in the National Health and Nutrition Examination Surveys

Arthur J. McDowell

Introduction

In the course of planning a major health survey, the purpose of each step in the process must be understood. The present section looks at the history of five national health examination surveys, focusing particularly on several measurements and tests related to certain aspects of physical functioning. Why? If we believe, and we do, in the "prime importance" (1, 2) of being specific about the purpose of the survey we are planning, then we must put this principle into practice here and always.

It is a truism to say that we must look at our history in order to learn from our past experiences. It is essential, however, that we truly learn, which means that we will wisely adapt—not merely adopt—the previous patterns. James Russell Lowell said: "New occasions teach new duties . . . time makes ancient good uncouth" (3). Remembering this can help us to avoid Epaminondas's foolish repetition of behavior that would perhaps have been appropriate earlier but is no longer so (4). Lest the point be missed, examples could be cited of planners of health examination surveys in a developing country who were insistent on utilizing certain sophisticated and delicate measuring instruments because "that is the way it was done in the United States," ignoring pronounced and uncontrollable variations in environmental situations that interfered with the use of these instruments. There are, on the other hand, instances in which U.S. advisers were confounded when the national planners were proven right in their confidence that *in their cultural situation* it was possible to obtain compliance of very high percentages of selected sample persons under operational procedures that would not be acceptable in the United States.

The purpose, then, of setting forth here some history of the operations of examination surveys of the National Center for Health Statistics (NCHS) is to examine *critically* the way in which:

- The survey planning was done.
- The operations of the survey were carried out.
- The resultant data were made available.

Purposes of the surveys and their relation to assessing physical fitness

Before detailing what was done, how it was done, and how well it was done in each of the surveys, it is appropriate to examine why that which was done was done at all. What were the goals of these programs?

The first Health Examination Survey

One of the goals of the very first Health Examination Survey (HES I) was to establish that it was possible to successfully conduct a national health survey in which many of the data were collected by individual physical examinations and tests of selected sample persons. In 1950 the National Morbidity Subcommittee of the U.S. National Committee on Vital and Health Statistics was established to review the needs for morbidity statistics and to prepare a program for meeting them (5). That subcommittee's report in 1953 recommended establishment of two continuing national sample surveys (6, p.4). One survey was needed to collect data by health interviews; another was to utilize "mobile examination units to obtain data on undiagnosed and non-manifest disease, by means of laboratory screening, detection, and physical examination" (7).

However, at the time planning was under way for the first U.S. national health examination survey, there was no precedent for such a program. Health examination surveys had been conducted in limited geographical areas in the United States, but they had very different constraints (e.g., in the time available to obtain participation from sample persons). The plan developed took account, therefore, of the possibility that this first survey program might, for various reasons, have to be terminated before completion (8). Happily, this concern proved unfounded.

Obviously the first HES program had purposes beyond establishing the feasibility of such surveys. This concern, however, had an impact on the setting of the survey's goals. Some procedures that were believed likely to discourage participation by the selected sample persons were avoided for that reason. Thus, for example, although one of the goals was to determine the prevalence of diabetes, the survey's serum glucose test was not based on a blood specimen taken in a fasting state and the level was determined only at 1 hour after challenge. By the time the second Health and Nutrition Examination Survey was carried out, it was deemed feasible to obtain fasting specimens and also specimens at 1 hour and 2 hours after glucose challenge.

In addition to the specific target of diabetes prevalence, the HES I had the goals of obtaining data indicative of the prevalence of the specific forms of arthritis and of cardiovascular disease. Baseline data on certain physical and physiological measurements (blood pressures, serum cholesterol levels, skinfolds, electrocardiographic tracings, height, weight, auditory and visual acuity levels, etc.) were also obtained in the program. The purposes of that first survey, then, were specific and somewhat limited and did not aim at an assessment of physical functioning per se.

The second and third Health Examination Surveys

When the first HES program was in place, proving that such surveys could be successful, attention was directed to the question of what should be the next program. Operations of HES I (also called Cycle I), begun in November 1959, had reached the planned full-scale level in early 1961. The experience already gained demonstrated the great importance of the skills and dedication of the trained field staff and the complexity of assembling the necessary equipment and recruiting the required personnel. This awareness was a reason for the decision to adopt a pattern of simultaneous operation on three levels. The concept was that over a particular time period of, say, 2 or 3 years, the program would be engaged in analyzing the data from the past survey (HES I), collecting data in the succeeding survey (HES II), and carrying out general and then detailed planning for the next survey (HES III). Under this concept, it was hoped to avoid a complete dismantling of the carefully constructed field apparatus (human and mechanical) at the end of a survey and to move directly into pilot testing, pretesting and operation of a new survey program (9, p. 2).

The target population for a second health examination survey was recognized almost from the beginning as the children and youth who had been excluded by the limitation on the HES I target population of 18–79 years of age. As the adult examination survey had come to the attention of health workers in government and in the private sector, those individuals, groups, and agencies concerned with child health expressed interest in a survey in which those under 18 years would be examined. In the earliest phases of planning HES II, the tentative designation of the group to be studied had been children and youth ages 6 through 17 years. As the planning proceeded, it became apparent that differences between the children in the early years of this age segment and those in the later years were so great as to require separate programs. Such matters as feasibility of self-administered tests, types of motivational approaches to be used, and size of some of the equipment led to the decision to redefine the target age group to the population of children ages 6–11 years (9, p. 4).

The determination of the substantive goals of the second Health Examination Survey necessarily reflected the interests and concerns that had been expressed about the health of these children. It was obviously a different set of concerns from those of the HES I program. Chronic diseases such as arthritis and diabetes were not appropriate topics of concern because prevalence rates were known to be too low for study through this sort of survey. The acute conditions that do present health problems in the age group 6–11, conditions such as communicable diseases, cannot be studied in a survey that has the kind of geographic and time constraints characteristic of HES.

There was broad agreement that data related to growth and development of U.S. children were needed. Normative standards of physical growth, measured not only by height and weight but also by a variety of segmental lengths, of girths, and of skinfold thicknesses, were needed. It became apparent that some data relevant to the intellectual growth and development of children should be included. Measures of visual acuity levels and of auditory acuity levels for a representative sample of the child population were seen to be important, both as indicators of health problems requiring attention and as a basis for establishing norms.

As the various interests were examined, it became clear that a study focused on growth and development would require some measurements of physical functioning, assessments of specific aspects of fitness. Thus the targeted conditions included several that are related to the goals indicated for the third National Health and Nutrition Examination Survey (NHANES III), although the difference in age groups for the two surveys limits the applicability of that experience.

Selecting the target population of the third Health Examination Survey (HES III) presented no problem. At the time the age range for HES II had been set at 6–11 years, it was recognized that a later survey would be needed to address the older children and youth. Thus, the planning of HES III aimed at continuing many of the measurements made of children 6–11 years (heights, weights, skinfolds, visual and auditory acuities, etc.), so the selected parameters of growth and development could be followed up to what is, for most individuals, the end of the growth period. These measurements were continued. In addition, the plan and the purposes of this survey were expanded and modified from those of the predecessor survey. In part this reflected the growing experience in the examination survey operation methodology. In part it reflected the population differences, with the targeted age group covering the period of adolescence, except for early adolescence. In part the changes grew out of the longer planning period and the broader range of expert consultants who participated in the development of this program (10, pp. 3–7).

One innovation carried out in the Cycle III program had been planned earlier. This provided that, for the first time, an element of longitudinality would be introduced into the series of HES programs. It had been decided to conduct the Cycle III survey, in which those 12–17 years would be examined, in the same 40 sample areas and the same sample segments as were utilized in the Cycle II program, involving children 6–11 years. This meant that some of the children examined in Cycle II would be examined again in Cycle III. A child aged 10 years in the first of these programs, for example, might be aged 13 and still living in the same place when the later survey was carried out (10, p. 8).

To an even greater extent than in the HES II program, attention in HES III was addressed to physical functioning. Contracts were used in the planning process to develop instrumentation appropriate for the mobile examination center operating method. This was not a new concept. In planning Cycle I (HES I), a single-visit cardiovascular examination had been developed by the contract process (11). The extent of this activity was expanded for Cycle III. In the area of psychological testing and measurement of behavior, the data pertaining to the child had been obtained from the parent or guardian and from the school. With the older children able to respond for themselves, considerably more data were collected bearing on these important health-related variables of health habits, history, and behavior (10, pp. 49–61). This, at least in part, reflects the interest in the total person that the pediatrician and adolescent medicine specialist had with regard to their patient populations, an interest frequently greater than that of the medical practitioners dealing with adults. This is less true today as the importance of the individual's own input into his or her health maintenance has come to be widely recognized.

The first National Health and Nutrition Examination Survey (HANES I), including HES IV

The present paper is not intended to provide a history of the development of the series of HANES programs now carried out by NCHS. Elsewhere in these chapters a general picture is presented of the evolution of HANES from the HES programs and of their relation to the total set of NCHS general population surveys. However, it may be instructive to examine briefly certain stages in that evolution. The exercise will illustrate, with respect to planning national surveys in the context of a Federal Government health agency, how necessary it is to adapt plans to changing demands and exigencies, and how flexibility helps offset the "gang aft a-gley" factor (12) in plans that are never the "best laid" ones in the first place.

Under the three-level operation concept described earlier, by the year 1969 the program should have finished the planning phase of the next (fourth) Health Examination Survey (HES IV), completed the data collection (field operation) phase of HES III, and been ready to shift the analytic and publication efforts from HES II data to HES III data. This was not the situation at that point in time. In the latter part of 1964, about halfway through the data-collecting phase of Cycle II, budgetary problems had arisen, necessitating some adjustments. Severe restrictions on personnel and dollar resources had resulted in a decision to slow down the data-collecting work and thus prolong the time period (9, p. 14). This added constraint on the survey program had not been removed for the Cycle III operation and it remained in 1969 (as did the war in Vietnam, which, of course, was related).

Work on a plan for a Cycle IV Health Examination Survey had been carried out during the 1967–68 period, and many tentative decisions had been reached. These called for a target population of adults and a set of specific diseases that would be among the target conditions for the program. Some of them (e.g., cardiovascular diseases) had been target conditions in Cycle I, and the aims of the planned Cycle IV program included some indications of changes over time. Heart disease prevalence may have been lowered because much effort had been expended by public health workers to promote preventive measures such as dietary changes and exercise. Other target conditions (e.g., chronic pulmonary disease) were to be measured for the first time (13).

The general thrust of the early planning took into account the past HES experience, the gaps in the health data already collected, and the expressed needs of many different disciplines in the health field. In hindsight, however, it is clear that the planning was still focused on the long-range goals of providing better research data. Such data could ultimately serve in improving knowledge about disease prevalence and health-related normative data and so, it was hoped, in improving health and medical care. These kinds of gradualist goals had been acceptable in the late 1950's and early 1960's, when the first examination surveys were planned and supported.

By the time of the late 1960's enormous changes had taken place. National attention had been brought to problems of hunger, malnutrition, and disease found to exist in many parts of the United States, particularly among impoverished persons, with initial evidence based on examination of children in some Southern States. A report by a "Citizens' Board of Inquiry into Hunger and Malnutrition," hearings and a field inspection trip by a subcommittee of employment, manpower, and poverty (under the Senate Committee on Labor and Public Welfare)—these things and more led to pressures requiring increased Federal health resources to address this present problem.

In December 1967 the U.S. Congress passed legislation (14) requiring a "comprehensive survey of the incidence and location of serious hunger and malnutrition and health problems incident thereto." Serious problems were inherent in the nature of the charge of this legislation and the requirement for a report to the Congress within 6 months giving "findings and recommendations." These, along with other budgetary, administrative, and jurisdictional considerations, placed impossible limitations on obtaining a statistically valid sampling national survey. The survey that was carried out, the "National Nutrition Survey," which came to be known as the "Ten State Nutrition Survey" (15) did provide, at the least, evidence of some nutrient deficiencies. Before any final data from that survey were available, however, it was apparent that, because of the design and methodology, valid measures for the U.S. national population and designated subgroups thereof (age, regional, income, etc.) could not be provided. The Secretary of Health, Education, and Welfare therefore decided that the authority of the 1956 National Health Survey Act would be used to institute a continuing program of national nutritional surveillance. Within the National Center for Health Statistics it was decided, after detailed study, to assign to the Health Examination Statistics Division, which had conducted the HES programs, the task of planning and carrying out the required survey (6, pp. 21–24).

The detailed history of the beginnings of HANES is set forth elsewhere (6, pp. 14–21; 13, pp. 1–5). The point here is to note that after about 2 years of planning the content of the HES IV program, political and administrative considerations required a new and major kind of survey program. Other considerations also required the tentative plan to be reworked. At about the same time, and growing out of the same new concern for the practical usefulness of survey data, the National Center for Health Statistics was directed to focus more of its resources on obtaining data concerning health care needs, particularly unmet health needs. The plan decided on for the HANES I program did not abandon any of the three different purposes:

- Obtaining some prevalence data on specific diseases in adults.
- Measuring the nutritional status of the total population.
- Obtaining data to serve as indexes of unmet health needs.

It is the first of these three purposes that yielded data most nearly related to the present effort to include in HANES III measures of physical functioning and fitness. This first purpose could not be satisfactorily accomplished with examinations of only the subsample of the adult population in HANES I, which had been projected as about 30,000 persons for the entire sample ages 1–74 years. It was decided to continue examination of more of the adults on completion of the HANES I cycle, an activity labeled as the "HANES I Augmentation" (or sometimes as the "HES IV Continuation").

The second National Health and Nutrition Examination Survey

In mid-1974, as the data collection phase of the first HANES program was completed, formal planning for the next such program was begun. About that time, the acronyms used and the names of the programs were modified so as to include the designation "national," which all of these related HES and HANES programs had always been. The principal purposes of NHANES II were to be the dual ones of again making an assessment of the nutritional status of the U.S. population and obtaining detailed data regarding a number of specific diseases. The nutritional assessment was required to fulfill the charge to measure and monitor nutritional status. The disease-related data included prevalence of certain new target conditions (e.g., diabetes, allergy, and lead levels in blood). It was also planned to obtain normative data on pulmonary function for persons 6–24 years of age similar to those collected in NHANES I for persons 25–74. The original plan called for an investigation of cardiac arrhythmia as part of the continuing collection of data related to cardiovascular disease. A more detailed description of the plan and the planning process has been published (16), as is the case for each of the HES and NHANES surveys.

Measurements of or related to cardiovascular endurance, lung function, and hand grip strength

The description of what was done in these five national surveys and how it was done is well documented in several hundred pages of separate program reports in Series 1 of the NCHS *Vital and Health Statistics* publications. Beyond that level, there are available, on a limited basis, printed or mimeographed (but not formally published) manuals of staff instruction for particular surveys. These relate to the entire survey and give evidence of the concern for quality control, standardization of operational procedures, and similar matters that are essential to the successful conduct of a statistically valid sampling survey involving health examinations. An example of one portion of one such manual, covering only the steps related to the specific tests of lung function and the electrocardiographic tracings, is presented as appendix A.

In the following paragraphs, the detail will be limited to brief descriptions of what kinds of tests were done and what equipment was used for each of the specific sets of measurements of particular interest here. Along with this information concerning how it was done, there will be some evaluations of how successfully it was done and what kinds of problems were encountered.

Tests of cardiovascular endurance

In the first HES program (HES I), there was no test of cardiovascular endurance (table 1). In HES II, children 6–11 years were given an exercise tolerance test. The equipment used was the bicycle ergometer—Model AM 368—manufactured by Elema-Schonader of Sweden. The end point of the test was the examinee's pulse rate, which was monitored and recorded using special equipment (Kenelco) fastened to the examinee by an electrode attachment. The equipment was set for the required workload using a reference table specific for age, sex, and weight. If the pulse rate had advanced appropriately after a 1-minute test run, the test was continued for 2 minutes. If the test run indicated underloading or overloading, the load was adjusted and the test continued. Values recorded were the four pulse rates: Before exercise, at 2 minutes after ending exercise, at the maximum point reached during the 5-minute rest after ending the exercise, and at the end of that 5-minute rest. In addition to these values, there was a timed graph of the pulse rate throughout the test, which the equipment automatically traced.

Throughout the HES II data collection, the experience with the bicycle ergometer could have been regarded as highly satisfactory. It became the favorite part of the entire examination for the young children and was featured in the brochures used in the household visit by the HES interviewer, when parental consents and additional data were obtained. A picture of the child taken on the bicycle ergometer was given to him or her after the total examination in the mobile examination center was completed. The HES experience with this instrument was not truly successful, however, because of problems in the resultant data. One of the bicycle ergometers used turned out to have been out of calibration, at least by the end of the survey, with respect to the other one and the true workload that the ergometer should have been presenting. This problem was recognized in a study (17) made after the survey was completed, but because there seemed to be no obvious way of identifying the errors so as to permit adjustment, the data were set aside for later study. The pressures of other data sets and other surveys were such that this was never done.

Even before detection of the calibration error in the Cycle II exercise tolerance test, the planning work with respect to measuring cardiovascular endurance in the Cycle III survey had moved in another direction. After considerable consultation, it had been decided to use a treadmill exercise tolerance test. This involved a walk on a specially constructed treadmill at a speed of 3.5 miles per hour. The grade of the incline was zero at the outset. The treadmill incline was elevated by an electric motor to a 10-percent grade for the last 3 minutes of the 5-minute work period. The end point of the test was the subject's pulse rate before and at various intervals during the walk on the treadmill. The pulse rate was monitored and recorded by means of precordial leads going to a cardiotachometer, which gave both the electrocardiogram and instantaneous pulse rate (appendix B).

The planning work for the exercise tolerance test had involved much consultation with the Laboratory of Physiological Hygiene at the School of Public Health and the Exercise Laboratory at the School of Education, both of the University of Minnesota. In June 1964 a contract relationship was entered into between those two groups and NCHS. Under this contract and its later amendments, the work to be carried out involved developing and determining the feasibility of a test of aerobic capacity. This contract also involved carrying out replicate measurements in the laboratories at Minnesota made on sample youth who had recently been examined in the Health Examination Survey in sample segments located in that area. This aspect of the contract would incidentally provide a check on the measurements made under field operating conditions by permitting comparison with several repeated tests in the laboratory. Primarily, however, it was part of a continuing effort to develop equations that could be used to predict the maximal oxygen consumption under the conditions of the HES treadmill test.

The report of this research effort was not received in NCHS until mid-1975. This was 5 years after completing the data collection in the HES III survey and a year after the succeeding HANES I field operations had ended. This delay, pressures of new surveys, changes in key personnel, and other factors all are among the reasons that there have been no reports of the results of the treadmill exercise tolerance test.

Table 1. Selected tests related to physical fitness performed in the Health Examination Surveys (HES) and National Health and Nutrition Examination Surveys (NHANES) with indication of tests used and of available information

Survey, age of sample persons tested, and dates of data collection	Cardiovascular endurance test	Lung function test	Hand grip strength test	Other relevant or correlated measures	
				ECG	Body measurements
HES–I, adults 18–79 years, 1959–62	None done.	None done.	None done.	12-lead EKG Details and findings in Series 11, No. 10.	Height, weight, 3 girths, 2 skinfolds, 11 anthropometric measurements. Details: Series 1, No. 5. Findings: Series 11, Nos. 8, 14, 35.
HES–II, children 6–11 years, 1963–65	Bicycle ergometer exercise tolerance test; pulse rate. Details: Series 1, No. 5; Series 2, No. 21. Findings: No published reports. No public use data tapes.	Spirometry—Collins water-sealed 6-liter vitalometer. Details: Series 1, No. 5; also Series 2, Nos. 72, 86. Findings: Series 11, No. 164. No public use data tapes.	Hand dynamometer, 3 tests each hand, record of handedness Details: Series 1, No. 5. Findings: No NCHS published reports. See App. D, this paper.	10-lead EKG Findings: No NCHS published reports.	Height, weight, 5 girths, 3 skinfolds, same 11 anthropometric measurements as in HES–I plus 11 more. Details Series 1, No. 5. Findings Series 11, Nos. 104, 119, 120, 123.
HES–III, youth 12–17, 1966–70	Exercise tolerance test using treadmill walk at 3.5 mph, for 2 minutes at zero grade and 3 minutes at 10% grade. Details: Series 1, No. 8. (See also table 2 this paper.) Findings: No NCHS published reports. Public use data tapes do include.	Spirometry recorded in analog data form on magnetic tapes to be converted to digital tape form using PHS computer program. Details: Series 1, No. 8. Also Series 2, No. 72, 86. Findings: No NCHS published reports. No public use data tapes.	Hand dynamometer, 3 tests each hand; record of handedness. Details: Series 1, No. 8. Findings: No NCHS pub. Published reports (see table 2.)	12-lead ECG done in conjunction with spirogram using Data Acquisition Unit. Analog tapes converted to digital. Details: Series 1, No. 8. Findings: No NCHS published reports.	Height, weight, 6 girths, 4 skinfolds, plus 26 additional measurements, 17 of which were made in HES–II. Details: Series 1, No. 8. Findings: Series 11, Nos. 124, 126, 132, 143.
HANES–I, Nutrition Exam 1–74 years 1971–74 Detailed Exam, 25–74 1971–75	None done.	For detailed exam subset—spirogram, recorded by Beckman Digicorder onto magnetic tape in digital form. Details: Series 1, No. 10a, of this paper, and Series 2, No. 86. Findings: Series 11, No. 222. Public data tapes do include best trials data only.	None done.	For detailed exam subset—12 lead electrocardiogram with results recorded on magnetic tape in digital form. Details: Series 1, No. 14a. (Also see appendix A of this paper.)	For both nutrition and detailed exam, height, weight, 2 skinfolds, 2 breadths (bitrochanteric and elbow), plus others for ages 1–7 years. Details: Series 1, No. 10a. Findings: Series 11, Nos. 208, 211, 224.
NHANES–II, Nutrition exam 6 months–74 years, Health exam variable ages 1976–80	None done.	For persons aged 6–24–spirogram, recorded by Beckman Digicorder onto magnetic tape in digital form. Details: Series 1, No. 15, Series 2, No. 86. Findings: No NCHS reports. Data on this, 6–24 ages spirometry released 1986.	None done.	For persons 25–74, supine ECG performed. 3 channels of data (12 std, lead and 3 Frank lead) recorded in digital magnetic tape. Details: Series 1, No. 15. Findings: No NCHS published reports.	Height, weight, 2 skinfolds, 2 breadths, 1 girth, plus certain others for very young. Details: Series 1, No. 15. Findings: No NCHS published reports.

The draft report (18), accepted as a final one on this study, presented a somewhat tentative recommendation that the following equation be applied to the HES data in order to estimate maximal oxygen intake from pulse rate after exercise (P), taking account of height, weight, age, and sex:

$$\text{Max VO}_2\ (\text{ml/kg}) = A + B(P) + C(\text{wt}/\text{HT}^3).$$

Values derived in the contract study for the coefficients A, B, and C in this equation were presented separately for each of six age-sex groups in table 2.

Despite the great amount of time and resources

Table 2. Least squares regression coefficient and standard error of estimate for estimation of MVO_2(cc/kg/min) from submaximal pulse rate, weight, and height: Health Examination Survey sample data

Sex and age	N	A	B	$C \times 10^{-5}$	S.E.E.
Lab pulse rates					
Boys:					
12–13 years	28	84.40	−0.1479	−10.5851	4.34
14–15 years	28	87.47	−0.2389	−1.2199	5.05
16–17 years	15	71.60	−0.1396	−0.3870	4.15
Girls:					
12–13 years	28	87.40	−0.1965	−11.1803	3.48
14–15 years	17	86.61	−0.2697	−3.2963	3.38
16–17 years	15	50.76	−0.0642	−3.4440	2.55
Trailer pulse rates					
Boys:					
12–13 years	28	88.74	−0.1559	−11.9510	4.30
14–15 years	28	90.34	−0.2545	−0.3184	4.16
16–17 years	17	79.28	−0.1953	+1.2134	4.63
Girls:					
12–13 years	28	97.32	−0.2310	−12.0705	3.19
14–15 years	18	95.22	−0.2747	−6.2547	3.45
16–17 years	15	55.92	−0.0844	−4.0004	2.30

Notes: PR = five minute N.H.S.T. pulse rate. "Lab. Pulse Rates" refer to fourth Laboratory visit. $MVO_2 = A + B(PR) + C(Wt/Ht.^3)$.
Source: Table 16 from report listed as reference 18: Taylor, H.L., and others: Preliminary Report on the Development and Feasibility of a Test of Aerobic Capacity for the Health Examination Survey.

devoted to obtaining data on cardiovascular endurance in the HES II and HES III programs, no findings reports have been produced.

No attempt was made to obtain measures of cardiovascular endurance in either the NHANES I or NHANES II programs. In the planning work for the latter program, however, there was an attempt to obtain data on cardiac arrhythmia, including heart monitoring by use of portable Holter electrocardiogram recorders. These recordings were to be made over a 2-hour period while the other parts of the examination were carried out. This was not done in NHANES II because of interactions between this and certain other procedures (e.g., glucose tolerance testing). Pressures of survey operating schedules prevented resolution of that problem in time to include the test. It is mentioned here because there might be relevance to cardiovascular endurance testing if such a test were to carried out in a future NHANES program.

The experience of the five national surveys discussed here is of limited value in planning for measurement of cardiovascular endurance measurement per se in a NHANES program to assess the physical fitness status of U.S. adults. In none of these surveys were tests to measure this variable administered to an adult population. The surveys in which children and youth were examined (HES II and III) did establish that in these age groups it was possible to obtain high response rates for participation in the examinations (96% and 90%, respectively). Nearly all of those examined were willing and able to perform the tests of cardiovascular endurance (bicycle ergometer and treadmill, respectively). However, this experience is not what could be expected from an adult population. In the Tecumseh Community Health Study (19), for example, there is a decrease in the percentage of the sample willing and able to take the modified Harvard Step Test as age increases. More than 90 percent of those aged 10–19 years completed this test; just under 50 percent of those aged 50–69 completed the test (20).

The significance of the concern with the expected participation rate for a test of cardiovascular endurance must be understood in the light of the purpose of the program. Where that purpose is largely methodological or is the study of relationships among the test results and other data collected in the same program, the composition of the population for which data are obtained is relevant but not of preponderant importance. Where that purpose is the description of the entire population for which the group tested is a sample, the representativeness of that sample comes to be of overwhelming importance. If NHANES III, now being planned, has as a principal purpose an assessment of the physical fitness status of the American adult population, then any test included for that purpose must be one that can be expected to yield participation rates that will permit generalization of the findings to the total U.S. population. There is no precise answer to the question of how high the participation rates must be to avoid nonresponse rates that will invalidate the generalizability of the findings. A number of NCHS reports have addressed the problems related to nonresponse (21–24). When account is taken of the fact that only with the expenditure of enormous effort has it been possible to achieve overall response rates in the range of 70–75 percent for the all-ages populations included in the NHANES samples, it becomes clear that it is unwise to consider including tests that will add appreciably to nonresponse.

Tests of lung function

In the first Health Examination Survey no attempt was made to measure lung function. The second survey, directed to a sample of children 6–11 years, included a vital capacity test. The spirometry test made use of a Collins 6-liter vitalometer. This test produces a tracing made on a timed rotating cylinder as the examinee blows back a deep breath into the tube on the equipment. These tracings can be measured to provide parameters such as maximal forced expiratory volume and peak flow rate. Three separate recordings were supposed to be made for each examinee, but for most subjects there were no more than two. Because the test is dependent on maximal effort by the examinee, the technician administering the test encouraged and exhorted the examinee throughout. During the data collection period, it was not recognized that this test presented unusual problems in getting the sample persons to take the test. Test tracings were available on 6,932 children—all but 187 of the 7,114 children 6–11 years who came for the exam. The retrieval of the data from the tracings produced did require the additional step of carefully measuring each of the charts to obtain the values needed. This was done, and some of the results are set forth in a published report (25). As shown in this report, however, the data have very serious shortcomings. These problems were encountered during analysis of the data. Additional measuring of the data by various persons skilled in the area of spirometry revealed that the measurements made by clerical personnel were seriously flawed, and both the measurements and the decisions as to such matters as selecting best trials were carefully reconsidered. This led to decisions to reject some of the data and to make imputations based on other data available. What was done and the implications thereof are set forth in the previously referenced findings report and also in a report in the methodology series (26).

In the third Health Examination Survey, a changed method of spirometry was introduced. A module that used computerized data collection techniques was used. The data were recorded in analog form on magnetic tape, and the tapes were translated into digital form using the spirometry computer program developed by a unit within the Public Health Service. Most unfortunately, the tapes containing these data were largely destroyed. The loss of these data is described in a methodology report (26).

In the work of planning the first National Health and Nutrition Examination Survey (NHANES I), further refinements were made in the spirometry data collection module. The analog data system used in HES III was replaced by digital tape equipment. The field operating procedures manuals were revised to reflect the then-current methodology (e.g., use of a standard set of five trials to ensure maximal values). A report in the NCHS methodology series gives a detailed description of the spirometry data acquisition hardware system (27).

A report presenting some findings from the spirometry testing in the NHANES I program has been published (28). This report presents the findings data for nine parameters for the some 5,500 persons tested. Approximately 84 percent of those tested had totally satisfactory data, as defined by presence of reproducible best trials. The data are shown separately for this group and for the some 800 persons who did not have reproducible best trials. No data were obtained for about 1,300 sample persons who, for various reasons, did not start or did not complete the test, although these were sample persons who had come to the mobile examination center to be examined. A total of about 6,900 persons did come in for the NHANES I detailed examination. That total was out of a sample size of 9,881 persons in the target population. Thus the overall response rate for the detailed examination was about 70% (70.7% on an age-adjusted basis). The total nonresponse (survey overall nonresponse plus item nonresponse) on this test, then, was essentially 50%. This would seem to preclude making statistically valid projections to the total U.S. population sampled. Some uses are being made, however, of lung function data from NHANES I. An example is in the analysis of the NHANES I Epidemiologic Followup Study data (29). Reduced pulmonary function has been shown to be significantly associated with an increase in mortality, even after adjusting for smoking behavior (30).

Tests of hand grip strength

No test of hand grip strength was included in the first Health Examination Survey (HES I). In the second and third programs (HES II and III), a test was included. The procedure was similar for both programs. Using a hand dynamometer, the technician tested the examinee, adjusting the handle to fit the hand properly and urging maximum effort in squeezing. Three tests were made with each hand, alternating hands after each test. The technician also recorded the response to a question about handedness and noted any relevant impression as to the validity of the test.

No NCHS publications have included the findings of the hand grip strength tests for either Cycle II of III. The test data are included in the public use data tapes that have been made available by NCHS. These standardized microdata tape transcripts are described in a published NCHS catalog (31). Some findings from the survey of children aged 6–11 years are contained in table 3.

No hand grip strength tests were included in either the NHANES I or NHANES II programs.

Table 3. Hand grip strength findings in Health Examination Survey of U.S. children and youth, 1963–65: Means ($\bar{x}$) and standard deviations (SD), separately for each hand, by sex, age and race

	Male				Female			
	Right hand grip strength		Left hand grip strength		Right hand grip strength		Left hand grip strength	
Race and age	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
Total								
6–11 years	13.7	4.7	12.9	4.6	11.8	4.5	10.9	4.3
6 years	8.8	2.7	8.3	2.6	7.6	2.4	7.0	2.4
7 years	10.7	2.9	10.1	2.8	9.0	2.6	8.4	2.5
8 years	12.7	3.1	11.9	3.2	10.6	2.7	9.8	2.8
9 years	14.5	3.4	13.6	3.3	12.3	3.2	11.4	3.0
10 years	16.4	3.5	15.4	3.5	14.2	3.6	13.3	3.6
11 years	18.8	4.1	17.8	4.1	17.0	4.5	15.7	4.3
White								
6–11 years	13.6	4.7	12.9	4.5	11.6	4.4	10.8	4.1
6 years	8.8	2.7	8.3	2.6	7.5	2.4	6.9	2.3
7 years	10.7	2.8	10.1	2.8	8.9	2.5	8.3	2.4
8 years	12.6	3.1	11.8	3.2	10.5	2.7	9.7	2.7
9 years	14.4	3.3	13.5	3.3	12.2	3.1	11.2	2.9
10 years	16.4	3.5	15.3	3.5	14.0	3.4	13.0	3.5
11 years	18.7	4.1	17.7	3.9	16.7	4.2	15.4	4.0
Black								
6–11 years	14.1	5.0	13.1	5.0	12.6	5.2	11.7	5.0
6 years	9.2	2.7	8.5	2.8	7.9	2.4	7.1	2.5
7 years	10.8	3.1	9.7	2.9	9.5	2.8	8.8	2.7
8 years	13.6	3.4	12.3	3.4	10.9	3.1	10.0	3.0
9 years	15.0	3.8	14.1	3.3	13.5	3.7	12.4	3.5
10 years	16.7	3.8	15.9	3.8	15.8	4.1	15.1	4.1
11 years	19.6	4.3	18.6	4.8	18.7	5.8	17.4	5.6

Notes: The sample numbers on which the table is based are 3,574 males (3,101 white and 458 black). Totals include persons of other races not shown as separate categories.

Source: Adapted from tables in personal communication from Dr. Robert Malina, Department of Anthropology, University of Texas at Austin.

Other measurements relevant to or correlated with these physical fitness tests

Electrocardiograms

The experience of the HES and HANES programs with respect to use of electrocardiograms has possible relevance to the tasks of designing an NHANES III fitness protocol. Measures of individual physical fitness frequently use heart rate as the end point measured along with specific work performance tasks involving known energy expenditure. Although this has not been done for any of the adults included in these past programs, 12-lead, resting electrocardiograms have been recorded. The only ECG data included in published NCHS reports are those for the HES I program (32), but the NCHS public use data tapes, referred to earlier, include these test results for the HANES I and HANES I Augmentation (or Cycle IV detailed examination) as well as for the HES I program. Other ECG data, including those from NHANES II and Hispanic HANES, are expected to be released in 1987.

The methodology employed in these programs has varied from the use of paper tracings in HES I, through analog magnetic tape form in HES III, to use of Beckman Digicorder equipment, which produces digitally recorded data on magnetic tapes. The difficulty with the first of these methods is in the measurement and interpreting process required. How this was done in the HES I program is described in the Series 11 report of findings (32). The second of these methods, which records the ECG data in analog tape form, requires that these tapes be translated into digital magnetic tape form. The third method, the recording of ECG data directly onto digital form magnetic tape, as was done in the NHANES I and NHANES II programs, does not have either of the problems associated with the other two. The detailed instructions followed by the technician in using the Beckman Digicorder to do this in the two NHANES programs were set forth in the staff procedures manual, the relevant portion of which is included as appendix A.

Body measurements

In each of the five HES and HANES programs, the recorded data have included a broad range of body measurements. Many of these are relevant to the process of measurement of physical fitness. In predicting maximal oxygen intake in tests of aerobic capacity, the regression equations may make use of both body weight and some index of obesity. This may also involve use of height and of some skinfolds and girths. A study based on HANES I data assessing dietary intake and cardiovascular risk factors, specifically blood pressure correlates, found that body mass index (weight/height squared) and subcutaneous adiposity (measured as the sum of subscapular and triceps skinfolds) were importantly related to systolic and diastolic blood pressures (33). This association was found in every sex, age, and race group examined. Because of the correlations and relationships, it is appropriate to note here what related data have been obtained in the HES and HANES programs.

In every one of the five national survey programs, height and weight have been recorded with meticulous concern for accuracy. The details of how these measurements were done are set forth in the already referenced program description reports and in the reports of findings (34–41).

There have been skinfold measures recorded in each of the HES and HANES programs. These readings have been made in a standardized manner in so far as specifying just how the procedure is to be done and training and retraining the technicians can ensure standardization. Two skinfolds, triceps skinfold measured as the right upper arm and subscapular skinfold measured on the right side, have been done in each one of the five surveys. In the children and youth surveys (HES II and III), there was also a skinfold taken on the right lateral chest wall. The Cycle III program also included two additional skinfold measurements, right-side suprailiac and a medial calf skinfold made on the left side. Some girths were included in each of the surveys. Details on these, various measurements of breadth, and other body measurements such as sitting height and segmental lengths are set forth in the published reports of findings (42–47).

Lessons from the HES-NHANES experience

Issues relevant to NHANES III planning

It would be hard to overstate the critical importance of specific recognition of the purpose for each element considered in the design of NHANES III. It is not enough to speak of wanting to obtain data related to physical fitness of American adults. Because what we are considering is a national survey involving scientific sampling and medical examinations, every aim of the survey must be identified and subjected to the following inquiry: Is the rather elaborate mechanism needed to fulfill this aim? If the aim is, for example, to establish the prevalence rate of diabetes in U.S. adults so that planning for health programs and medical care can be done, then the answer is probably affirmative. If the aim is only to examine the interrelationship between a particular laboratory finding and a specific disease, the answer may be negative. In the past there have been occasions when a survey has ended up with inadequate valid data to generalize about the entire population sampled for the test in question. In such cases it may be possible only to salvage some investigation of interrelationships among internal parts of the data, but we should not plan to do that at the outset and we should make every effort to avoid ending up with only that.

A design issue deserving some attention is whether the plan should aim at collecting only cross-sectional data or whether there should be some planning to have a longitudinal element in the survey. Each of the past surveys has been basically an attempt to get a one-time picture of the state of the studied population. Although it has taken several years to collect the national data for a single survey, the conceptual model has been obtaining data that could be assumed to characterize the total population at the midpoint of the survey collection period—point prevalence. There have been, however, some steps in the direction of adding an element of followup. Thus, a separate NHANES I Epidemiologic Followup Survey, described elsewhere, was carried out by an NCHS contractor using the findings for the sample population examined in the NHANES I program as a starting point and comparing those findings with newly obtained data on these same individuals about 10 years later. Again, as pointed out earlier, the Cycle III sample aged 12–17 years was designed to ensure some followup examinations of persons who had been examined in the HES II program. Where it is thought fairly likely that this kind of thing might be done later, some steps should be taken regarding the identifying data recorded and the consents obtained.

A sample design issue that must be given attention is the best relationship, if any, between the sample selected for the NHANES III program and the samples selected for the successive National Health Interview Surveys (NHIS) carried out by NCHS. This issue is not a new one. In the very first HES program it was considered, and it continues to recur. Most frequently, it is suggested that the examination in the HANES program might be given to a sample that is a subset of the NHIS sample. In this simplest form, the suggestion is not feasible for a variety of reasons. One major reason is that the NHIS sample is—for good design considerations—located in 1,900 different primary sampling units throughout the United States. The smallness of the numbers of persons interviewed for NHIS in any one primary sampling unit, even in as long a period as 1 year, is inconsistent with the numbers of persons required for NHANES in one geographic location. Several hundred sample persons are the minimal numbers in order to move in, set up, and relocate a team of medical and paramedical examiners with anything approaching cost effectiveness. There are other reasons, too, why this subset concept is not feasible. They have been developed in reports made to the NCHS Director by a special committee set up within NCHS to study this question (White-McDowell committee) and by the responsible division director within NCHS to both the General Accounting Office and the Office of Management and Budget in response to their questions.

There can be a relationship between the two samples and the two sets of data (NHANES and NHIS) without the one sample being a subset of the other. In the first HES program, the same set of interview questions as used in NHIS at the time HES began was administered to the persons in the HES sample. These interview questions remained unchanged for the HES sample persons throughout the 3 years of the survey, although some later changes occurred in the instrument used in NHIS. Some sort of linkage between the two survey programs provides great potential advantages. This is especially true as NHIS moves more in the direction of obtaining data describing the lifestyle patterns of the population. To study the health of the U.S. population in the 1980's without collecting information on such matters as the amount and kind of physical exercise and like activity would be to lag decades behind the state of knowledge. To be at the cutting edge, it would seem desirable to include not only these things but areas that are the province today of such pioneers as Norman Cousins (48, 49) and Richard Symington (50). That would involve finding out the extent to which the U.S. population was recognizing the responsibility of the individual for his health maintenance and disease control. Is this issue too controversial to be considered?

The three-level operation concept has been described in these pages and in referenced publications. When one considers the experience of these five surveys, it is clear that the survey operations have strayed widely and to an increasing extent from this model. The difficulties of ensuring continued Federal funding (including authorizations and salaries for needed personnel) to carry out health examination surveys extending over several years has been a factor in this changed pattern. The mechanism of contracting the conduct of survey operations to private agencies outside the Federal Government permits multiyear contracts in years when dollar availability exists. This methodology has other problems, which will be discussed shortly. In the process of designing NHANES III, it will be important to consider by whom each of the three operations of planning, data collection, and data analysis and reporting should be performed and how each of them should be phased.

The experience of the five national health examination surveys should make planners conscious of the importance of asking the question, How much is too much program content? It is not easy to answer the corollary question, How do we tell? The answers will, of course, be dependent on many factors, such as the available resources, the various time limitations (e.g., time allowed for obtaining the complete survey data and time burden that can be borne by each examinee), and the value of the resultant data from the specific item considered for inclusion in the program. Some help in answering the questions about "too much" will come from applying the guidance already given (1) about being conscious of the purpose and being specific about its details. Experience shows, however, that there is a barely resistable urge to add on still another item because "as long as the effort has been made to get the sample person there for the examination, it will be a simple matter to do the test for item X as well." That urge should be resisted. Consideration of the data that have been collected and not analyzed or put to possible use should help in resisting it.

An issue, separate from but not unrelated to the "too much" issue, is the matter of whether certain tests are inappropriate for use in a national health examination survey because they require such elaborate and complex instrumentation. The methodology for use in testing cardiovascular endurance should aim for use of some simpler exercise challenge than the elevating treadmill, which may perhaps have been defensible for teenagers. The use of the simpler and well-tested Harvard Step Test or modifications thereof would seem better advised. The methodology used for spirometry in the adult detailed examination of the NHANES I and Augmentation programs probably represents the best state of the art. The fact is that, in the population samples for those programs, the number of usable test results constitute data for only a little more than half of the people who should have been persuaded to come in for examination and should have produced a satisfactory test. This kind of response rate makes it impossible to generalize about the total national population. Although there are simpler tests that relate to lung function, the data they yield are, of course, less comprehensive and may raise questions of validity. In any event, it is important to be clear as to the purpose of the test and then to decide on the required complexity of instrumentation.

Issues related to data collection

The issue of contract mechanism versus in-house effort should not be misunderstood as a matter of either one or the other. The issue is rather the extent to which the NCHS (in-house) staff will actually carry out the field operation, as compared with setting forth standards and guidelines in a contract to some agency outside NCHS. All of the HES and NHANES programs have had a mix of the two operating methods, but in markedly differing proportions. In most of these surveys, the U.S. Bureau of the Census has acted as a contractor for NCHS in the initial phase of interviewing at potential sample households. In some of the surveys, they have also carried out a large part of the process of explaining the survey to and obtaining consents from selected sample persons. In two related surveys (the Hispanic HANES and the NHANES I Epidemiologic Followup Survey), the actual data collection has largely been done by an outside contractor, with guidelines from and some participation by staff of NCHS.

There are advantages and disadvantages to both methodologies. The Bureau of the Census is especially expert in the conduct of the early phases of such surveys, and their access to official Census Bureau address lists provides an important advantage. The HES–NHANES experience gives some evidence that the use of NCHS personnel has advantages in the "persuasion process," the obtaining of consents for examination. There is also some limitation on the flexibility of scheduling and operating the survey when the Census Bureau has responsibility for the first wave of interviewing. There is an important disadvantage to having most of the data-collecting work performed on an outside contract basis because the personnel in NCHS who should be responsible for many aspects of the interpretation of the final data cannot properly perform this function if they have been dependent on some other agency to carry out the work of collecting those data.

The second issue noted in this section concerns the use of mobile examination centers. This issue has some relationship to the immediately preceding one because, when use of outside contractors is considered, there may be suggestions to do the data collecting—the physical examination part of it, that is—in fixed sites, medical clinics, hospitals, or the like. At times in the past HES–NHANES experience such suggestions were made, and at one point the conduct of a pilot study to investigate the feasibility of that operation was even required. It is well to keep in mind the several kinds of reasons that have led to the decisions to use mobile examination centers. They include such concerns as the requirements for very special housing for some measuring instruments, such as spirometry and lung function tests or treadmill exercise cardiovascular endurance tests. For these, there not only must be adequate electrical power and the like, but there needs to be a controlled environment with respect to temperature and humidity. Even if no such complicated and sensitive measuring instruments were included in the examination, there is need for a standardized setting for making almost all of the measurements. The measurement of blood pressure, for example, will be affected by very many situational factors. One might not get valid sample data for projection to a national population if the subject's blood pressure is markedly affected by household factors, which are not standardized and are variable in their effects. There is a temptation to think that one can carry out physical examinations in hospitals, clinics, or doctors' offices, which will be found in all parts of the country, but experience with the vagaries of national scientific sampling surveys makes one aware that sometimes the location will be in an area where such optimism is belied. A mountainous rural county in Appalachia or an Indian reservation location in Arizona come to mind. Then, even in our cities, there may be overcrowding of existing medical facilities, which precludes access.

Another issue related to the data collection process is the matter of the kinds of persons who should supervise the steps in the examination process. The interdisciplinary nature of the operation is a part of this. Some operations in some examination surveys may be conducted by health technicians, medical electronics engineers, nurses, physicians, dentists, psychologists, nutritionists, dietitians, interviewers, and even by specialists in certain medical fields (e.g., dermatologists and cardiologists). It is essential that definite responsibilities for aspects of supervision be clearly established. There must be some single overall responsibility because interests may sometimes conflict among separate pieces of the examination. The possibilities include a survey statistician experienced in health surveys and skilled in diplomacy, a physician with good appreciation of statistical problems and issues and with research requirements, an epidemiologist, and perhaps others. There needs to be definite understanding of the roles of the various staff members and of the authority to resolve problems, both at the field location and at the overall headquarters. In the HES and HANES experience, it has sometimes proved useful to make use of an adapted academic research model in which a particular expert in a specialty field plays a significant role in planning how a piece of the data will be collected, training the field personnel who will be responsible for administering the test, making numerous field visits to observe the performances, reviewing the data as they are collected, and participating in the analysis and presentation of the findings. In terms of the supervision question, this has the potential of adding an additional problem, but if lines of authority are clearly specified, the advantages of this model can come without undue disruption.

An issue that can never be disregarded is the importance of staff motivation. The planner can expend great cost and effort to design a survey but the end point of it all is the report of findings. Essential to success in the effort is a high degree of motivation on the part of every member of the survey team. The cooperation of the sample persons must be obtained, first in agreeing to the examination, then in coming, in staying throughout, in exerting maximal effort where required, in agreeing to the less pleasant aspects such as having blood drawn, and in other ways. It is the individual staff members, one after another, who must be able to obtain this cooperation and who need to be highly motivated in belief in the importance and concern for the success of the survey. The process starts with staff selection, but training and retraining are required. There is no magic formula, but where experience has shown success, the retention of demonstrated abilities seems indicated.

In this rapidly changing world of information recording and transmitting, an obvious issue concerns the ways in which the survey should make use of the latest developments in electronic communication. This area has been the subject of some study within NCHS. The topic included study of the extent to which the data could be recorded directly in the field in a form that would be transmitted electronically to the headquarters computer center. Related research was carried out by the Bureau of Census.

Finally, with respect to issues in the data collection, one must emphasize the overriding importance of a response rate high enough to permit statistically valid conclusions pertaining to the total population of which the survey is a sample. When a program director defending a health examination survey is asked how much each one of these examinations costs, it is essential to be able to point out that the real measure of cost is not the single examination, because that one particular sample person actually represents thousands of like persons in the population surveyed. This point can be made, but only so long as the survey's response rates remain acceptable. The solution to this problem is not an easy one. It is not a problem that can be solved by money alone. It has been pointed out that, in the NHANES I program, NCHS did, for the first time, begin remunerating the sample persons who participated. This seemed to have some favorable effect on the response rate and it has been continued. A later study examining the effect of an increased amount of payment did not show that clear gains would result from such change. Today's climate may possibly be one in which we should not take for granted the favorable effect of even the minimal payment used in the past. The concern with budget deficits and Government costs, coupled with the many statements denigrating the work of Federal Government employees, may present new problems in response rates for surveys.

Data release issues

There will be some among the users of NHANES III data who will be notably less concerned with how the data were obtained than when the data will be available. The need for speed issue is frequently opposed to the need for quality research, but ways must be found to satisfy it while maintaining high quality standards. One way in which this issue can be partly met is in planning the survey to be completed in 3 years but to consist of two parts, each a representative sample of the population. One half-sample would be scheduled to be completed in about the first 18 months of the survey's data-collecting period, and some broad preliminary data could be made available from that smaller sample.

A second issue related to the release of the data from NHANES III concerns the medium to be used for release of data. In the early history of the National Center for Health Statistics, the major means for release of data from its programs was in various series of reports published by NCHS (or by the U.S. Government Printing Office for NCHS) after preparation, most frequently by NCHS staff. Sometimes reports of particular findings were published in the appropriate scientific literature, either by NCHS staff or by persons obtaining special tabulations of NCHS data. Beginning as a formal program in 1978, NCHS made available public use data tapes containing the raw data collected in the survey. Any data that would permit the identification of individual sample persons were omitted from the tapes. The appropriate sampling weights were included along with other administrative and demographic data. This has proved to be a very useful means of data release and has greatly increased the usefulness of those data. It would be unfortunate, however, if this kind of data availability were to be thought of as lessening the need for continuation of the publication of separate reports on particular findings of the survey. Many health research and public health workers are not able to take advantage of the availability of computer tapes. Equally important, the importance of having NCHS staff persons present the data with their knowledge of how the total data collection process may have affected the findings is quite great. Equally great is the availability of a vehicle in which the NCHS staff persons can follow through on the planning and directing work of the survey and gain the satisfaction of a final product.

A final issue related to release of data from the HES and NHANES programs concerns the need for collaboration with the related professional communities. Data from these surveys gain great usefulness by being incorporated in textbooks, reference works, and the like, which reach a far wider audience of medical and health-related workers than can be reached otherwise. Many examples could be given to illustrate this, but a single one will make the point. A very large body of survey data with respect to body measurements characterizing the children and youth of the United States has been incorporated in the reference work produced by Malina and Roche in two volumes, entitled *Physical Status* and *Physical Performance*, respectively (51). This kind of collaboration has produced similar examples in a broad range of medical areas that have been subjects of past surveys in this series of programs. This fine record will, it is hoped and believed, be continued with data from NHANES III.

Recommendations

1. In planning NHANES III, a deliberate effort should be made at every step of the process to specifically identify the purpose of each aspect of the plan and to assess the purpose's importance, its realistic chance of achievement, and its relevance to other and to overall purposes.
2. NHANES III should be planned as a single survey in which cross-sectional data on the sample population are collected, but data should include sufficient identifying information to make future followback study feasible if this proves desirable and possible. This information should include social security number if the obstacles to this have been or can be overcome.
3. It is recommended that the NHANES III sample be selected independently of the NHIS sample, but a possibility of some linkage should be secured through administering a set of identical interview questions to each of the two population samples. This set of questions should include not only ones that would characterize lifestyle with respect to exercise, activity (work and leisure), and like matters, but also questions concerning attitudes towards health and self-responsibility in disease prevention and therapy.

4. It is recommended that a deliberate effort be sustained throughout the planning process to avoid overloading the content of the examination beyond what is necessary to accomplish the essential purposes of the survey. It may help to remember the crudely expressed "KISS" rule, attributed to the Army. The acronym stands for "keep it simple, stupid!"
5. For the adult population of NHANES III, it is recommended that less complicated tests than used for the HES youth program be used for measuring cardiovascular endurance and lung function. In the survey, the kinds of data that are correlated with the targets of those tests should be collected, including blood pressures and measures of adiposity and body mass. In addition, some simpler measures of diminished vital capacity and of increased heart rate should be considered. Pilot testing of such measures, as well as steps to ensure standardization, validity, and reliability, would be required.
6. Every effort should be made to ensure that adequate staffing (both numerically and skillwise) be provided to NCHS so that direction of every phase of the program—planning, data collecting, and analyses—can be carried out on an in-house basis. To the extent that outside contracting must be used, effective control of what is done and how it is done must be retained.
7. It is recommended that the mobile examination center type of data collecting operational mode be used in NHANES III.
8. All members of the staff involved in the NHANES III program, most certainly including all field staff members, should participate in a training program to ensure awareness of the importance of the survey, the benefits that should result from its successful completion, and the role they can play in this.
9. In order to provide information from NHANES III on as rapid a time schedule as possible, it is recommended that the sample be planned and scheduled for a 3-year data collection period, with a representative half-sample of locations included in the first about 1½ years of data collection. Thus preliminary data from this half-sample can be made available.
10. It is recommended that release of the data from NHANES III be accomplished through reports published in the established NCHS series as well as through making available detailed public use data tapes and through professional journal articles.
11. Finally, it is recommended that the NHANES III program continue the high level of cooperation and collaboration with the relevant professional and scientific communities both in the Federal Government and in the private sector. This has been a hallmark of the survey programs and is of prime importance in their success.

References

1. National Center for Health Statistics: Plan and Operation of the Second National Health and Nutrition Examination Survey. *Vital and Health Statistics.* PHS Pub. No. 1000–Series 1, No. 15 (p.2).
2. McDowell, A.J.: Health examination surveys—in theory and in application, in P. Armitage, ed., *National Health Survey Systems in the European Economic Community.* Proceedings of a conference held in Brussels on Oct. 6–8, 1975. Commission of the European Communities, EUR 5747e, 1977.
3. Lowell, James Russell: "The Present Crisis." Quoted in *Bartlett's Famous Quotations*, 14th Edition. Little, Brown and Company, Boston, 1968 (p.62).
4. Bryant, S.C.: *Epaminondas*, in V.S. Hutchison, *Chimney Corner Stories* (pp.102–108).
5. National Center for Health Statistics: Origin, Program, and Operation of the U.S. National Health Survey. *Vital and Health Statistics.* PHS Pub. No. 1000–Series 1, No. 1 (p.4).
6. National Academy of Public Administration: Improving the Health and Nutrition Examination Survey. An evaluation by a panel of NAPA (p.4).
7. U.S. Public Health Service: *History of the United States National Committee on Vital and Health Statistics, 1949–1964.* U.S. Public Health Service, June 1966 (pp.7–8).
8. National Center for Health Statistics: Plan and Initial Program of the Health Examination Survey. *Vital and Health Statistics.* PHS Pub. 1000–Series 1, No. 4 (p.16).
9. National Center for Health Statistics: Plan, Operation, and Response Results of a Program of Children's Examinations. *Vital and Health Statistics.* PHS Pub. No. 1000–Series 1, No. 5.
10. National Center for Health Statistics: Plan and Operation of a Health Examination Survey of U.S. Youths 12–17 Years of Age. *Vital and Health Statistics.* PHS Pub. No. 1000–Series 1, No. 8.
11. National Center for Health Statistics: Evaluation of a Single-Visit Cardiovascular Examination. *Health Statistics.* PHS Pub. No. 584–D–7, Public Health Service, Wash., D.C. U.S. Government Printing Office, Dec. 1961.
12. Robert Burns: "To a Mouse," quoted in Bartlett, John: *Bartlett's Famous Quotations*, Fourteenth Edition, Little, Brown and Company, Boston, 1968, (p.492).
13. National Center for Health Statistics: Plan and Operation of the Health and Nutrition Examination Survey. *Vital and Health Statistics*, Series 1, No. 10a. DHEW Publication No. (HSM) 73–1310.
14. Section 14, Public Law 90–174, Dec. 5, 1967.
15. Ten-State Nutrition Survey, 1968–1970. Health Services and Mental Health Administration, U.S. Department of Health, Education, and Welfare. DHEW Pub. No. (HSM) 72–8130.
16. National Center for Health Statistics: Plan and Operation of the Second Health and Nutrition Examination Survey, 1976–80. *Vital and Health Statistics*, Series 1, No. 15. DHHS Pub. No. (PHS) 81–1317.
17. National Center for Health Statistics: Calibration of Two Bicycle Ergometers Used by the Health Examination Survey. *Vital and Health Statistics.* PHS Pub. No. 1000–Series 2, No. 21.
18. Taylor, H.L., and others: Preliminary Report on the Development and Feasibility of a Test of Aerobic Capacity for the Health Examination Survey. Unpublished manuscript. National Center for Health Statistics from the Laboratory of Physiological Hygiene (School of Public Health) and the Exercise Laboratory (School of Education), University of Minnesota.
19. Francis, T., Jr.: Aspects of the Tecumseh Study. *Public Health Reports,* 1961, 76:963–966.
20. Derived from data in Table I: Montoye, H.J., and others: Heart Rate Response to Modified Harvard Step Test: Males and Females, 10–69. *The Research Quarterly*, 40:1 March 1969. American Association for Health, Physical Education, and Recreation. NEA.
21. National Center for Health Statistics: Cooperation in Health

Examination Surveys. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 2, No. 9.

22. National Center for Health Statistics: Factors Related to Response in a Health Examination Survey, United States, 1960–62. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 2, No. 36.
23. National Center for Health Statistics: Quality Control in a National Health Examination Survey. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 2, No. 44.
24. National Center for Health Statistics: A Study of the Effect of Remuneration upon Response in the U.S. Health and Nutrition Examination Survey. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 2, No. 67.
25. National Center for Health Statistics: Forced Vital Capacity of Children 6–11 years, United States. *Vital and Health Statistics*. Series 11, No. 164. DHEW Pub. No. (PHS) 78–1651.
26. National Center for Health Statistics: Methodologic Problems in Children's Spirometry. *Vital and Health Statistics*. Series 2, No. 72. DHEW Pub. No. (PHS) 78–1346.
27. National Center for Health Statistics: Computer Assisted Spirometry Data Analysis for the National Health and Nutrition Examination Survey. *Vital and Health Statistics*. Series 2, No. 86. DHHS Pub. No. (PHS) 81–1360.
28. National Center for Health Statistics: Basic Data on Spirometry in Adults 25–74 Years of Age, United States, 1971–75. *Vital and Health Statistics*. Series 11, No. 222. DHHS Pub. No. (PHS) 81–1672.
29. Cohen, B.B., Cohen, B.H., and Drizd, T. The Effect of Reduced Pulmonary Function on Mortality. *Am. J. of Epid.*, Vol. 124, No. 3, p.538. September 1986.
30. National Center for Health Statistics, B.B. Cohen, H.E. Barbano, C.S. Cox, et al.: Plan and operation of the NHANES I Epidemiologic Followup Study, 1982–84. *Vital and Health Statistics*. Series 1, No. 22. DHHS Pub. No. (PHS) 87–1324. Public Health Service. Washington. U.S. Government Printing Office, June 1987.
31. Catalog of Public Use Data Tapes from the National Center for Health Statistics: NCHS. DHHS Publication No. (PHS) 81–1213.
32. National Center for Health Statistics: Coronary Heart Disease in Adults, United States, 1960–62. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 10.
33. National Center for Health Statistics: Dietary Intake and Cardiovascular Risk Factors, Part I. Blood Pressure Correlates: United States, 1971–75. *Vital and Health Statistics*. Series 11, No. 226. DHHS Pub. No. (PHS) 83–1676.
34. National Center for Health Statistics: Weight, Height, and Selected Body Dimensions of Adults, United States, 1960–62. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 8.
35. National Center for Health Statistics: Weight by Height and Age of Adults, United States, 1960–62. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 14.
36. National Center for Health Statistics: Height and Weight of Children 6–11 Years, by Age, Sex, Race, and Geographic Region, United States. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 104.
37. National Center for Health Statistics: Height and Weight of Children: Socioeconomic Status, United States. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 119.
38. National Center for Health Statistics: Height and Weight of Youths 12–17 Years, United States. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 124.
39. National Center for Health Statistics: Weight by Height and Age for Adults 18–74 Years, United States, 1971–74. *Vital and Health Statistics*. Series 11, No. 208. DHEW Pub. No. (PHS) 79–1656.
40. National Center for Health Statistics: Weight and Height of Adults 18–74 Years of Age, by Age, Sex and Race: United States, 1971–74. *Vital and Health Statistics*. Series 11, No. 211. DHEW Pub. No. (PHS) 79–1659.
41. National Center for Health Statistics: Height and Weight of Adults 18–74 Years by Socioeconomic and Geographic Variables, United States. *Vital and Health Statistics*. Series 11, No. 224. DHHS Pub. No. (PHS) 81–1674.
42. National Center for Health Statistics: Skinfolds, Body Girths, Biacromial Diameter, and Selected Anthropometric Indices of Adults, United States, 1960–62. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 35.
43. National Center for Health Statistics: Skinfold Thickness of Children 6–11 Years, United States. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 120.
44. National Center for Health Statistics: Selected Body Measurements of Children 6–11 Years, United States. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 123.
45. National Center for Health Statistics: Body Weight, Stature, and Sitting Height: White and Negro Youths 12–17 Years, United States. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 126.
46. National Center for Health Statistics: Skinfold Thickness of Youths 12–17 Years, United States. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 132.
47. National Center for Health Statistics: Body Dimensions and Proportions, White and Negro Children 6–11 Years, United States. *Vital and Health Statistics*. PHS Pub. No. 1000–Series 11, No. 143.
48. Cousins, N.: *Anatomy of An Illness As Perceived by the Patient: Reflections on Healing and Regeneration*. Norton, 1979.
49. Cousins, N.: *The Healing Heart: Antidote to Panic and Helplessness*. Norton, 1983.
50. Simonton, C., et al: *Getting Well Again—A Step by Step Self Help to Overcoming Cancer for Patients and Their Families*. J.P. Tarcher, Publisher. 1978.
51. Roche, A., and Malina, R.: *Manual of Physical Status and Physical Performance*. Vols. 1 and 2. Plenum Press, 1983.

Appendix A. Excerpt from NCHS Instruction Manual. Data Collection, Part 15c (Chapters 10 and 11) Examination Staff Procedures Manual for the Health Examination Survey, 1974–1975.

CHAPTER 10

ELECTROCARDIOGRAPHY AND SPIROMETRY

Equipment

The Beckman digicorder is used in HES to record spirograms and electrocardiograms. All data gathered by this unit are recorded on magnetic tapes which are to be forwarded to the Collection and Analysis Linkage Section at the end of each stand.

Aside from recording data, the little unit also provides a special coding circuit. The code is a series of pulses which the computer program identifies as the lead code; the recording location; the examinee's identification number; age, sex, race, and height; technician number; and a calibration constant.

This calibration constant is the product of the temperature correction factor and the water vapor pressure factor. The constant corrects volumes from ambient temperature and pressure saturated with water vapor (ATPS) to body temperature pressure saturated with the water vapor (BTPS). The code initiates each recording.

Code Interpretation

1. Coded information to be recorded with lead selector switch turned to Electrocardiogram STD:

1	2	3	4	5	6	7	8	9	10	11	12
Stand Number		Tech. Number		temp.	B.P.	Sample Number					Not used

2. Coded information to be recorded with lead selector switch turned to EKG leads 1 through 12:

1	2	3	4	5	6	7	8	9	10	11	12
Height		Weight			Age		S/R	9	9	I.D	Not used

3. Coded information to be recorded with lead selector switch turned to Aux. Lo (spirometer electronic calibration) and Aux. Hi I (spirogram):

1	2	3	4	5	6	7	8	9	10	11	12
Height		Weight			Age		S/R	9	9	I.D	Not used

Codes

Stand Number:

00-99

Technician Number:

71-99, 01-09

Temperature:

8 = 28°C
9 = 29°C or greater
0 = 20°C or less
1 = 21°C
2 = 22°C

3 = 23°C
4 = 24°C
5 = 25°C
6 = 26°C
7 = 27°C

Sample Number:

First two digits: 01-99,00 = stand number
Last three digits: 600-799 = examinee number

Height:

01-98 = height to nearest inch
99 = height unknown

Weight:

001-998 = weight to nearest pound
999 = weight unknown

Age:

25-74 = age in years
99 = age unknown

Barometric Pressure:

0 = 695 mm. to 704.9
1 = 705.0 to 714.9
2 = 715.0 to 724.9
3 = 725.0 to 734.9
4 = 735.0 to 744.9
5 = 745.0 to 754.9
6 = 755.0 or greater
7 = less than 675.0
8 = 675.0 to 684.9
9 = 685.0 to 694.9

Sex/Race:

1 = white male
2 = white female
3 = Negro male
4 = Negro female
5 = other male
6 = other female
9 = race unknown

I.D.:
0-9 = last digit of sample number

Digicorder Computer Tape

1. Label all tapes with the stand number, location, date, and the number of the tape for that stand, e.g., Stand 01, Philadelphia, Pa., Tape 1, 4-27-71.

2. Don't record more than 30 examinations on one computer tape.
3. If the power is shut off while the digicorder is on, remove the data tape and mount a new tape on the digicorder.
4. Send the computer tapes, all tracings, and two copies of the log book to Headquarters at the end of each stand.

EKG-Spiro Data Book

At the beginning of each stand record the barometric pressure at the top of the page.

On the top of each right hand page write the stand number and tape number. Below this make six columns headed respectively: Examinee ID (sticker), Date, Technician, EKG, Spiro, Temp. 29. Enter the appropriate information in these columns as the examinations are done.

On top of each left hand page write "Comments." Record on this page, directly opposite such comments as:

1. Miscoded info on 01999; should be...
2. Info not coded on 01999; should be...
3. Canc/PHY (Exam cancelled per physician)
4. Exam not done (or not completed) due to maximum time
5. Exam done on backup unit
6. New tape mounted on digicorder

The EKG-Spiro data book is to be used as the unusual occurrence form for these two sections. Any information which the technician regards as pertinent to the exams should be noted.

Electrocardiogram

1. General

For the electrocardiogram analysis, information from twelve leads is recorded on magnetic tape. A computer program has been developed to analyze this information and print diagnostic messages along with the data.

2. Procedure

Record the time on the control record. Put up the table and ask the examinee to lie on his back on the table and make himself comfortable.

A. Place the electrodes on the examinee.

(1) Place the electrodes over areas with the least muscle movement (approximately 2 to 3 inches above wrist and ankle joints over medial radius and medial tibia).
(2) Sensitize the areas where limb lead electrodes are to be placed by rubbing them lightly with electropads and leaving the pads in place.

B. Turn the selector switch to the STANDARD (STD) position.
C. Enter the examinee's identifying information in the following positions:

1 2	Location (stand number)
3 4	Technician number
5	Temperature in °C
6	Barometric pressure
7 8 9 10 11	Sample number
12	Not used

D. Record a standard by depressing the RECORD DATA push button. Criteria for an adequate STD are listed below.

10 mm. in height ± 1/2 mm.
baseline on 25 mm. line ± 5 mm.
four standard complexes present not varying vertically
by ± 3 mm. between complexes

E. Change the examinee's identifying information code to the following:

1	Height to the nearest inch
2	
3	Weight to the nearest pound
4	
5	
6	Age
7	
8	Sex/Race
9	9
10	9
11	ID (same as STD)
12	Not used

F. Turn the selector switch to Leads I, II, III, aVR, aVL, aVF in turn and record these leads. Wait a few seconds after turning selector switch before recording lead to allow machine to stabilize.

G. Place the chest lead on the examinee:

(1) Sensitize the areas where chest leads are to be placed by rubbing them lightly with the electropads and leaving the pads in place.

(2) Place the lead on the examinee before turning the selector switch to the position for that lead. Automatic centering is activated by positioning the chest lead before turning the selector switch.

H. Record chest leads, V1, V2, V3, V4, V5, V6.

3. Recording

A. The quality of the tracings should be checked in the monitoring oscilloscope before recording.

B. If any of the below described artifacts are present, check the possible causes and repeat the lead. The computer will process the last recording of each lead EXCEPT FOR LEAD V_6 which should not be repeated even if it is unacceptable.

The following criteria should be used to determine if a recorded lead is acceptable.

(1) Noisy signals in a lead are interference causing the baseline to be thickened; obscuring or obliterating the P wave. This can

be corrected by having the examinee relax completely and not touch any metal objects on or off the bed. Also check that the electrode clamps are not too tight.

(2) Wandering baseline occurs when the vertical difference between consecutive complexes is 3 mm. or more. If the baseline is wandering the electrodes should be reapplied using new electropads and frictioning the area well. Loose electrode clamps will cause high contact resistances that will result in a wandering baseline. By allowing a few seconds for the baseline to stabilize after changing the lead selector switch you may avoid many wandering leads.

(3) Check that the entire complex is on the tracing. If not, correct by adjusting centering control on the digicorder.

(4) Check the tracing for adequate stylus performance. Complex peaks should be clearly defined, the baseline should not be fuzzy. Adjustments can be made with the heat control on the digicorder.

(5) Four acceptable complexes must be present on each lead. The length of each lead should be 5 sec. (25 boxes).

(6) Check that the leads are marked correctly.

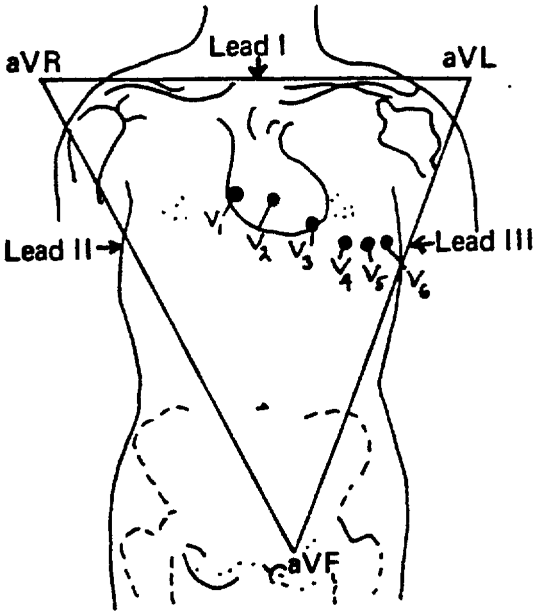
4. Lead positions

The basis of electrocardiography is the neuromuscular mechanism of the heart. The heart muscle contracts and the heart "beats" in response to a stimulus. The action of this stimulus sets up a tiny electric current which can be received and recorded by electrodes. Since electrical current in the heart flows from negative to positive, EKG electrodes are paired negative with positive to record the flow. Below is a listing of the negative and positive electrodes for each lead.

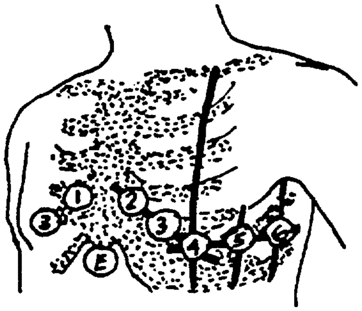
LEAD	*POSITIVE*	*NEGATIVE*
I	LA	RA
II	LL	RA
III	LL	LA
aVR	RA	LL & LA
aVL	LA	LA & RA
aVF	LL	LL & RA
V	Chest	LL & LA & RA

Graphic locations of standard leads (I, II, III), augmented limb leads (aVR, aVL, aVF), and precordial leads (V_1 - V_6) with reference to the heart are shown in the diagram. The electrical potential of any extremity is the same anywhere from its point of attachment to the torso to its most distal end. Therefore, electrodes attached to wrists and ankles will give the same EKG pattern as those placed at the point of union of torso and extremity. The extremity electrodes should be placed over the area where muscle is at a minimum. This prevents picking up muscle potential artifacts. The electrodes should be placed on the inside area of arms and legs, approximately 2-3 inches above ankles and wrists. The electrodes may be moved around to obtain a better recording at the discretion of the technician doing the test. The electrodes should be cleaned weekly with an abrasive cleaner.

STANDARD ELECTRODE POSITIONS



Lead positions



The accompanying chart illustrates the most commonly used electrode positions for chest or precordial leads. The positions shown are located or determined as follows:

V_1	Fourth intercostal space, at right border of sternum.
V_2	Same interspace, at left sternal border.
V_3	Midway between positions 2 and 4.
V_4	Fifth interspace, at left midclavicular line.
V_5	Same level as 4, in anterior axillary line.
V_6	Same level as 4 and 5, in midaxillary line.

Spirometer

1. General

The volume output of an electronic spirometer is recorded on magnetic tape. A computer program has been developed that is capable of measuring the forced vital capacity, the forced expiratory volumes at 1, 2, and 3 seconds, the maximum expiratory flow rate, the maximum mid-expiratory flow rate, and various peak flow rates for the tape recorded signal.

The examinee, who must stand during the test, is instructed to inhale maximally from room air, then exhale with maximal force and speed into the spirometer. A nose clip, if needed, and a disposable mouthpiece are provided for the examinee.

The important variables that affect the performance of the Forced Expiratory Spirogram (FES) fall into three categories:

Equipment function
Technician skill
Examinee comprehension and motivation

The section labeled "Calibration Procedure" is designed to help you control the first variable. The second variable refers to the technician's ability to establish initial rapport with the examinee, clearly administer the test instructions, provide the necessary stimulus and motivation for each examinee to perform maximally, and to judge the quality and reproducibility of the spirometry signals. Examinee comprehension and motivation are the end results of technician skill, i.e., the interaction between the technician and the examinee.

Should your best efforts not provide an acceptable test, indicate on the chart and in the EKG Spiro data book that the exam is VOID; and record on the chart your comments concerning the reason for such poor recordings; i.e., language barrier, submaximal effort (SME), inhalation artifact (IH), premature termination (PT).

2. Procedure

A. Depress the ON button on the storage display unit.
B. Set the BTPS factor on the calibrator in the spirometer to 000.

C. Set the BTPS factor on the F/V converter to 0.

D. Record the spirometer calibration as follows:

(1) Turn the selector switch to AUX LO. Use the same coding as in EKG lead 1-12 (page 10-2).
(2) Depress the RECORD DATA push button.
(3) Turn the spirometer VOLUME CALIBRATION selector switch from 0 to 5 three times, recording each selection for a minimum of 1 second.
(4) Set the spirometer volume calibration selector switch on "Operate."
(5) Set the spirometer flow calibration selector switch on "Operate."

E. Turn the selector switch to AUX HI.

F. Instruct the examinee according to the following instructions:

(1) "Take in a great big deep breath of air." (Have the examinee inhale maximally from room air.)
(2) "Hold all of the air in." (Have him hold his breath long enough to insert the cardboard mouthpiece into his mouth while you start the recording.)
(3) "Put the mouthpiece into your mouth with your teeth resting lightly on it. Seal your lips tightly around it." (Demonstrate the right way.)
(4) "Blast your air into the tube as fast as you can, like a cough." (The exhalation should be made with the lips tight around the mouthpiece with maximal force and speed. Do not have the examinee use the nose clip unless he exhales through his nose.)
(5) "Keep on blowing out the same breath of air—don't stop, don't take another breath—until you just can't blow anymore." (Have him blow out as hard and fast as he can and to keep blowing until he "empties" his lungs.)

G. Demonstrate to the examinee a deep inspiration, proper placement of the mouthpiece, and the bursting of air into the tube. Continue to blow for at least 4 seconds.

H. Have the examinee do the first trial; observe closely to insure that all instructions are being followed.

I. Using the oscilloscope, calculate the examinee's approximate volume and flow rate. Refer to the regression chart and find Peak Flow Rate (PFR) and Vital Capacity (VC) cutoff points for the examinee.

J. If the examinee's values fall below the cutoff points, give additional instructions.

K. Make sure that the following requirements are met:

  (1) The onset of the spirogram should occur within 3 seconds after you depress the Record Data button.
  (2) An optimal baseline should not be less than 1 second or more than 3 seconds in length.
  (3) The spirogram should not fall at the completion of the examinee's effort. It is a good technique to try to maintain a straight line.
  (3) The computer program demands that the peak of the spirogram be maintained for at least 0.2 second (5 mm.).

L. Proceed with trials 2, 3, 4, and 5. Five trials must be done but no more may be done without allowing the examinee to rest for 20 min.

M. Try your best to get the full cooperation of the examinee. Since there is a great deal of voluntary control over respiration, the success of lung-function tests depends on your getting the examinee to actively and fully participate.

N. If two of the trials are not within the allowable range for both volume and flow rate after five trials, stop testing and let the computer decide which two trials are the most comparable.

O. Label and keep all tracings.

3. Determining data

A. Monitoring the volume signal with a strip chart recorder

A clinically acceptable curve must be smooth and free of inhalation artifacts. The extent of the examinee's cooperation should be questioned if successive total volumes vary more than ± 5 percent for volumes greater than 3 liters or ± 10 percent for volumes less than

3 liters. This variability can be easily estimated on the oscilloscope by eye and on Sanborn paper in the following manner:

5 percent of 6 liters (30 mm.) is 1.50 mm.
5 liters (25 mm.) is 1.25 mm.
4 liters (20 mm.) is 1.00 mm.
10 percent of 3 liters(15 mm.) is 1. 5 mm.
2 liters (10 mm.) is 1. 0 mm.
1 liter (5 mm.) is 0.5 mm.

B. Monitoring flow and volume with an oscilloscope

A memory oscilloscope with an X-Y axis is the most precise method of monitoring examinee effort. Flow is registered on the Y axis (vertically one square equals 2 liters per second). Volume is registered on the X axis (horizontally one square equals 1 liter). Both parameters can be calibrated so that flow and volume readings are read directly from the face of the CRT. Each respiratory effort results in a flow-volume loop. This is displayed on the oscilloscope and compared with subsequent efforts. Discreet changes in examinee effort and cooperation can be monitored by observing the shape of the loop and peak flow deflection. Small variations in respiratory effort cannot successfully be monitored with a volume curve alone. Acceptable spirograms result in reproducible loops. Reproducibility is determined by superimposing one loop over the other or comparing them side by side.

C. Oscilloscope evaluation

After the examinee has blown into the spirometer and you have watched the FV loop on the screen, evaluate the examinee's effort. If the vertical and horizontal movement of the scribing point of the scope is low or reduced, look for the following conditions in your examinee:

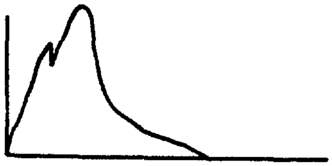
(1) The mouthpiece not being inserted into the mouth far enough or putting the lips in front of instead of around the mouthpiece.
(2) Collapsing of the mouthpiece by excessive mouth pressure.
(3) Tongue occluding the mouthpiece opening.

(4) Submaximal effort due to a lack of understanding of the procedure, reluctance to give a full effort, or improper instructions.
(5) Inability to comprehend instructions.

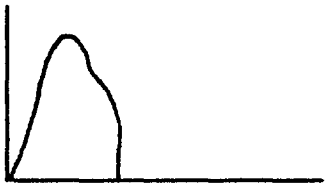
OSCILLOSCOPE EVALUATION USING THE FLOW-VOLUME LOOP

Typical flow loop

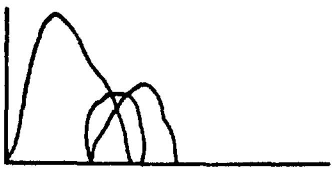
Sub-maximal effort (SME)



Hesitation in initial expiration effort - inhalation artifact (IH)

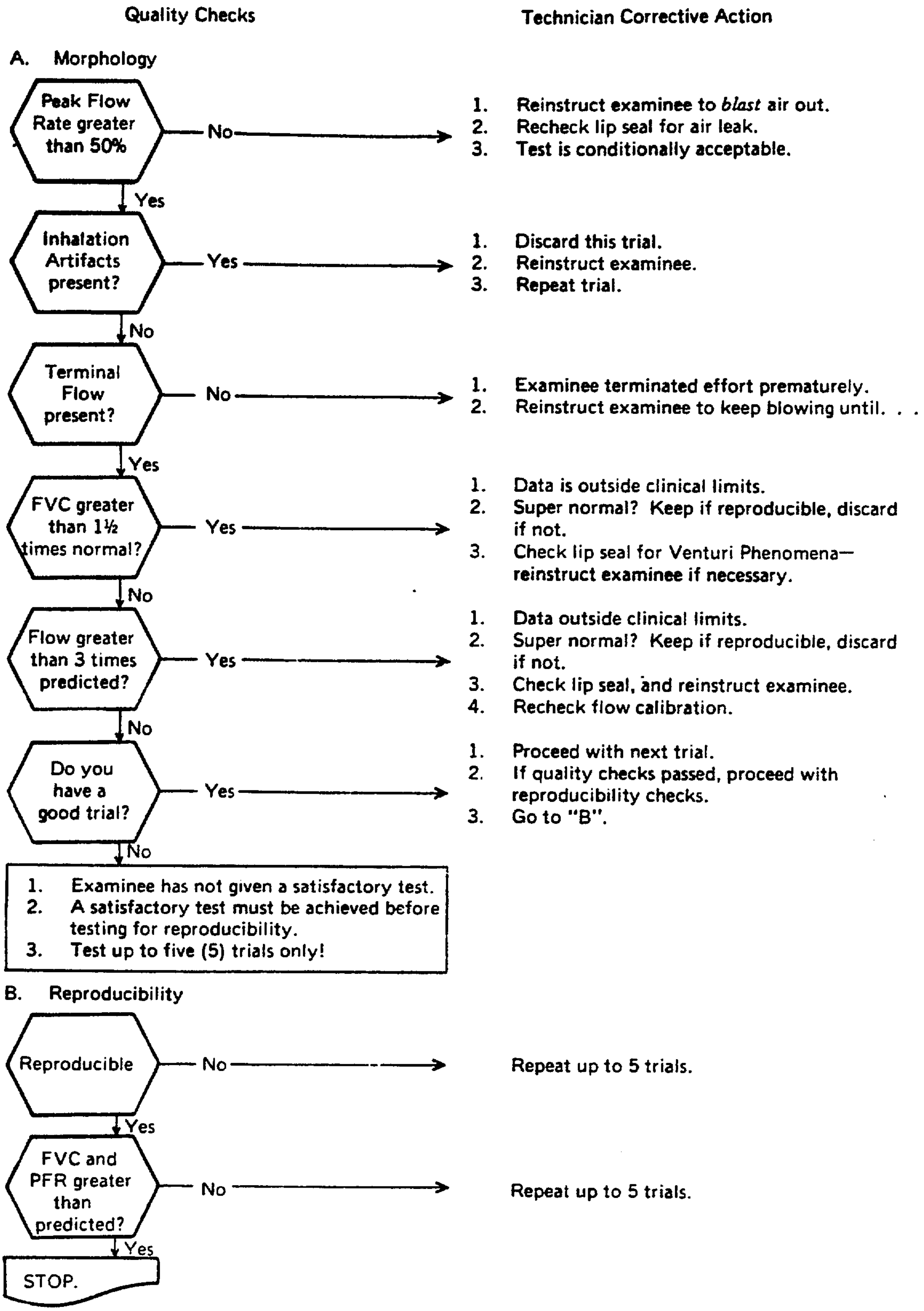


Blowing stopped too soon or mouthpiece removed before completed expiration - premature termination (PT)



Repeated efforts without removing mouthpiece - inhalation artifact (IH)

QUALITY CONTROL PROCEDURE



1. Reinstruct examinee to *blast* air out.
2. Recheck lip seal for air leak.
3. Test is conditionally acceptable.

1. Discard this trial.
2. Reinstruct examinee.
3. Repeat trial.

1. Examinee terminated effort prematurely.
2. Reinstruct examinee to keep blowing until. . .

1. Data is outside clinical limits.
2. Super normal? Keep if reproducible, discard if not.
3. Check lip seal for Venturi Phenomena—reinstruct examinee if necessary.

1. Data outside clinical limits.
2. Super normal? Keep if reproducible, discard if not.
3. Check lip seal, and reinstruct examinee.
4. Recheck flow calibration.

1. Proceed with next trial.
2. If quality checks passed, proceed with reproducibility checks.
3. Go to "B".

1. Examinee has not given a satisfactory test.
2. A satisfactory test must be achieved before testing for reproducibility.
3. Test up to five (5) trials only!

Repeat up to 5 trials.

Repeat up to 5 trials.

Calibration Procedures

1. Voltmeter calibration procedure

The instrument takes Volume and Flow Signals from a spirometer and freezes the display of the Volume voltage when the Flow voltage is zero. This freeze will happen at the maximum ***or*** minimum value of the Volume Signal when used in conjunction with a spirometer + Yoke Calibrator. The instrument will hold and display the first maximum or minimum after the RESET button is ***released.***

To operate, plug Volume and Flow cables from the spirometer into the appropriate front panel connectors or the rear panel connectors. Set the Front/Back switch located on the front panel to the appropriate position. Turn power on with ON/OFF switch. Allow 10 minutes warmup. Press RESET button and turn HOLD switch to hold position. The display will change as the Volume voltage changes until a minimum or maximum voltage is encountered. The display will hold the minimum or maximum voltage until the RESET button is pushed.

To get the desired maximum or minimum value, the release of the RESET must be synchronized with the movement of the spirometer piston. Look at volume display for indication of increasing volume. A similar procedure for holding minimums can be used.

For best results, it is advisable to take 12 maximum and 12 minimum readings. Discard the highest and lowest value from each set leaving 10 readings to average.

Subtract the means of the maximum and minimum sets, then add .003L (calculations done in Headquarters).

The voltmeter calibrator can be used as a voltmeter by placing the HOLD switch in the off or electronic position. The meter has a maximum range of ± 20V. The voltage displayed will be on the leads connected to the VOLTAGE terminals.

2. Spirometry calibration

A. Beginning of stand

(1) Preliminary setup

(a) Insert and connect the flow volume converter.

(b) Check to see if the bellows were completely open for transit, and remove the stopper from the portal.
(c) Engage the autonegator.
(d) Turn on all equipment and check the function of it.
(e) Put a tape on the digicorder labeled: Test tape - van number - beginning stand number - stand location - date - tech number. Proceed only after a minimum 20-minute warmup.

(2) EKG standard

(a) Turn the DRS lead selector switch to "STD" position.
(b) Turn the strip chart to "RUN" and use the "CENTERING CONTROL" on the main operator panel to position the baseline of the 10 mm. standardization pulse so that it coincides with the 25 mm. line on the chart.
(c) Check that the baseline is at the 25 mm. level on the DRS scope.
(d) Check that the height of the leading edge of the standardization pulse is 10 ± 1/2 mm.
(e) Code in on the DRS: stand number - tech number - temperature - barometric pressure; then 9's for the remaining six digits.
(f) Depress the "RECORD DATA" button on the DRS.

(3) Electronic spirometer calibration

(a) Code in all 9's on the DRS.
(b) Turn DRS lead selector switch to Aux Lo.
(c) Check that the BTPS on calibrator is on 000.
(d) Check that the BTPS on F/V converter is on 0.
(e) Place the "Front/Back" switch on the voltmeter calibrator to the "BACK" position.
(f) Turn the "HOLD" switch on the voltmeter calibrator to the "ELECTRONIC" position.
(g) Turn volume calibration dial (F/V converter) to "OPERATE."
(h) Record values (volts) from voltmeter calibrator on calibration form (Column 1).
(i) Depress the "RECORD DATA" button on the DRS.

(j) Turn volume calibration dial to "O."
(k) Record values on form.
(l) Depress "RECORD DATA" button.
(m) Turn volume calibration dial to "5."
(n) Record values on form.
(o) Depress "RECORD DATA" button.
(p) Turn volume calibration dial to "3."
(q) Record values on form.
(r) Depress "RECORD DATA" button.
(s) Turn volume calibration dial to "1."
(t) Record values on form.
(u) Depress "RECORD DATA" button.
(v) Repeat steps (g) through (u) (Record results in Column 2).
(w) Depress "RECORD DATA" button.
(x) Turn volume calibration dial (F/V converter) from "OPERATE" to: 0-5-0-5-0-5-0 (holding each for 1 second; the whole process should be completed in 9 seconds).
(y) Repeat steps (w) and (x) five times.
(z) Look at the storage scope for horizontal displacement, the large squares are equal to 1 liter/square horizontally.
(aa) Check the strip chart tracing for proper stylus placement; 0 should be 5 mm. from bottom of tracing and 5 should be 35 mm. from bottom of tracing.
(bb) Record in EKG/SPIRO data book the date and the values (volts) for "0" and "5" (e.g., 0 = 0.016 volts; 5 = +5.007 volts).
(cc) Compare these values to the values posted on the spirometer.
(dd) Report any differences greater than 0.020 volts to the supervisory technician.

(4) Mechanical spirometer calibration

(a) Turn volume calibration dial (F/V converter) to "OPERATE."
(b) Turn the DRS selector switch to AUX HI.
(c) Turn the "HOLD" switch on the voltmeter calibrator to the "HOLD" position.
(d) Attach the piston rod with the bolt on the shaft of the calibrator yoke.
(e) Turn on the calibrator.

(f) Check the circle on the storage scope for accuracy and any irregularity of the line while allowing the bellows to loosen up. Report any persistent irregularity of the line to the supervisory technician.

(g) Push the "RESET" button on the voltmeter calibrator between 2 → 6 liters (observe on the digital display) to obtain the maximum value (volts).

NOTE: Do not hold the reset button, just depress firmly and then release.

(h) Record the value (volts) from the reading on the voltmeter calibrator in column A of the form.

(i) Depress "RECORD DATA" button on DRS.

(j) Repeat steps (g), (h), and (i) five times.

(k) Push the "RESET" button on the voltmeter calibrator between 7 → 3 liters (observe on the digital display) to obtain the minimum value (volts).

(l) Record the value (volts) from the reading on the voltmeter calibrator in column B of the form.

(m) Depress "RECORD DATA" button on the DRS.

(n) Repeat steps (k), (l), and (m) five times.

(o) In the EKG/SPIRO data book record the date and the first values in columns A and B. Subtract the first number in column A from the first number in column B (e.g.: 8/27/73 7.328 - 2.189 = 5.139 volts)

(p) Compare the difference with the posted value.

(q) Report any differences greater than 0.020 volts to the supervisory technician.

(r) Remove the test tape and send it to Headquarters.

(s) Send the calibration form and tracing to the supervisory technician at Headquarters.

(5) Bellows leak check —beginning of stand

(a) Detach the bolt on the calibrator yoke.

(b) Open the bellows 5 → 7 liters (read on projection display).

(c) Insert the stopper in the portal.

(d) Turn the "HOLD" switch to the "ELECTRONIC" position.
(e) Turn volume calibration dial (F/V converter) to "OPERATE."
(f) Record values (volts) from voltmeter calibrator on calibration form.
(g) Turn off all equipment for 12 to 24 hours.
(h) Turn on all equipment (Dry Run Day).
(i) Allow minimum warmup period of 15 minutes.
(j) Record values (volts) from voltmeter calibrator on calibration form.
(k) Difference of 0.020 or greater is significant.
(l) Repeat leak test (steps (a) through (k)).
(m) Report any significant findings to the supervisory technician.
(n) If the second test also showed a leak of 0.020 volts or more:
  - Remove the front plate of the spirometer with rubber glove on.
  - Rub a light, smooth coat of talcum powder into the surface of the seal and the cylinder surface. Remove the excess dust.
  - Rub a light film of Vaseline on "0" ring.
  - Carefully reinstall the front plate. NOTE: See detailed instructions in Ohio Instrument Operations Manual, pages 39-40.
(o) Repeat leak test and report results to the supervisory technician; contact the engineer if difference is 0.020 volts or greater.

B. Daily calibration

A daily calibration is to be performed at the beginning of the first session of each examining day and at the beginning of both sessions when there is a break between; for example, morning and evening.

A MINIMUM OF A 20-MINUTE WARMUP OF THE EQUIPMENT MUST PRECEDE THE DAILY CALIBRATION.

(1) Turn the DRS lead selector switch to the STD position.
(2) Position the baseline of the 10 mm. standardization pulse to coincide with the 25 mm. line on the strip chart with the "Centering Control" on the DRS.

(3) Code in the stand number, technician number, temperature, and barometric pressure, then 9's for the remaining six digits on the digicorder.
(4) Depress "RECORD DATA" button on the DRS.
(5) Code in 9's for all 12 digits on the digicorder.
(6) Turn DRS lead selector switch to AUX LO.
(7) Check that the BTPS on calibrator is on 000.
(8) Check that the BTPS on F/V converter is on 0.
(9) Turn the "HOLD" switch to the "ELECTRONIC" position.
(10) Depress "RECORD DATA" button.
(11) Turn volume calibration dial (F/V converter) from "OPER" to: 0-5-0-5-0-5-0 (holding each for 1 second; the whole process should be completed in 9 seconds).
(12) Check reading on voltmeter calibrator at "0" and "5."
(13) Record in EKG/SPIRO data book; DAILY CALIB -- DATE-- TECH #-- VALUES (volts) for "0" and "5."
(14) Compare these values with the values posted on the spirometer.
(15) Report any differences greater than 0.020 to the supervisory technician.
(16) Turn the volume calibrator (F/V converter) to "OPERATE."
(17) Turn the DRS selector switch to AUX HI.
(18) Turn the "HOLD" switch on the voltmeter calibrator to the "HOLD" position.
(19) Attach the piston rod with the bolt on the shaft of the calibrator yoke.
(20) Turn on the calibrator.
(21) Check the circle on the storage scope for accuracy while allowing the bellows to loosen up.
(22) Push the "RESET" button on the voltmeter calibrator between 2 → 6 liters (observe on the digital display) to obtain the maximum value (volts).
(23) Record the value (volts) from the reading on the voltmeter calibrator in the EKG/SPIRO data book.
(24) Depress "RECORD DATA" button on DRS.
(25) Push the "RESET" button on the voltmeter calibrator between 7 → 3 liters (observe on the digital display) to obtain the minimum value (volts).

(26) Record the value (volts) from the reading on the voltmeter calibrator in the EKG/SPIRO data book.
(27) Depress "RECORD DATA" button on DRS.
(28) Repeat step (27) three times.
(29) Turn off the calibrator.
(30) Subtract the two values in the EKG/SPIRO data book (e.g.: 7.328-2.189 = 5.139 volts).
(31) Compare the difference with the posted value.
(32) Report any differences greater than 0.020 volts to the supervisory technician.
(33) Detach the piston rod from the shaft of the calibrator yoke.

C. End of stand

Perform the normal daily calibration on the last day of examinations.

Sterilization Procedure

Soak the spirometer hose for at least 1 hour in cidex aqueous solution and then wash it at the end of each stand and after an examination has been performed on an examinee who has TB or any other active respiratory tract disease.

Remove the front panel of the spirometer and wash the cylinder bore, seal, and piston with cidex aqueous at the end of every third stand. See the spirometer manual for further details.

FORCED VITAL CAPACITY
80 Percent of Predicted Value

Male

Age range in years	Height range in inches				
	59-63	64-68	69-73	74-78	79
20-25	3.24	3.85	4.47	5.08	5.33
26-30	3.11	3.73	4.35	4.96	5.21
31-35	3.00	3.62	4.24	4.85	5.10
36-40	2.89	3.51	4.13	4.74	4.99
41-45	2.78	3.40	4.02	4.63	4.88
46-50	2.67	3.29	3.90	4.52	4.77
51-55	2.53	3.18	3.78	4.38	4.66
56-60	2.45	3.07	3.68	4.30	4.56
61-65	2.34	2.96	3.57	4.19	4.44
66-70	2.23	2.85	3.46	4.08	4.33

Female

Age range in years	Height range in inches			
	55-59	60-64	65-69	70-72
20-25	2.33	2.74	3.16	3.49
26-30	2.26	2.67	3.09	3.42
31-35	2.19	2.61	3.02	3.35
36-40	2.13	2.54	2.96	3.29
41-45	2.06	2.48	2.89	3.22
46-50	2.00	2.41	2.82	3.16
51-55	1.93	2.34	2.76	3.09
56-60	1.86	2.28	2.69	3.02
61-65	1.80	2.21	2.63	2.96
66-70	1.73	2.15	2.56	2.89

Peak Flow Rate

Age range in years	80 percent of predicted		50 percent of predicted	
	Male	Female	Male	Female
20-29	9.46	6.14	5.91	3.84
30-39	9.08	5.90	5.68	3.68
40-49	8.71	5.65	5.44	3.53
50-59	8.33	5.40	5.21	3.38
60-69	7.96	5.16	4.97	3.22

CHAPTER 11

SINGLE-BREATH DIFFUSING CAPACITY

Equipment

Spirometer with 7-liter bell mounted in console
30-liter box-balloon system and single-breath 5-way valve mounted on console
Seven push-button panel with seven solenoid valves and associated electric circuitry
CO meter with relay-rack-mountable control panel with digital readout
Relay rack version of Collins helium meter with digital readout
Foot switch to activate kymograph
Tank of gas mixture containing approximately 10 percent helium, approximately 0.3 percent carbon monoxide, balance air, and reduction value
Stopwatch
Proportional divider
Metric ruler
Nose clip and foam inserts
Chart paper
Pens

General

The single-breath diffusing capacity will be measured on all examinees except those who the physician says are medically unable to do this part of the exam.

Because of the need for knowing the examinee's vital capacity before measuring the single-breath diffusing capacity, the spirometry measurements must be taken before the single-breath diffusing capacity.

Machine Maintenance and Calibration

1. Beginning of stand

   A. Tightly close the metal grid on the right half of the vent over the the Lung Analyzer machine and switch off the right half of the vent entirely. Adjust the metal grid on the left half so that the air is directed to the left side of the room.

   B. Adjust the following switches and leave them for the entire stand:

      (1) Turn the ON-OFF switch on the back of the machine to ON. Minimal warmup time is 6 hours.

      (2) Turn the CO switch on the CO readout panel to the .03 percent position.

      (3) Turn the He ON-OFF switch on the He readout panel to ON.

   C. Replace the drierite in the back of the machine.

   D. Replace the pens.

   E. Check the speed of the kymograph with the stopwatch; it should be 32 mm./sec.

   F. Fill the drum with distilled water to a level of 1 inch from the top.

2. Daily

   A. Replace the ascarite each morning.

   B. Check the drierite and replace when one-half inch is pink.

   C. Check the position of the pen and make sure it writes.

   D. Place paper on the the kymograph and cut sheets for the entire day's exams.

   E. See that the water level of the spirometer is 1 inch from the top.

   F. Check the pressure in the tank and notify the FOM if it's less than 500.

   G. Adjust the mercury level on the barometer base so that the ivory tip just kisses the top of the mercury.

3. Before each examination session

   A. Fill the balloon.

      (1) Push the RV-BB valve IN.
      (2) Connect the tubing from the gas tank to the CO-IN petcock.
      (3) Open the CO-IN and CO-OUT petcocks.
      (4) Slowly fill the balloon to between one-half and two-thirds full.
      (5) When the gas escape has stopped, close the CO-IN and then the CO-OUT petcock.
      (6) Pull the RV-BB valve OUT.

   B. Empty the balloon.

      (1) Turn the 5-way valve to Position 3. Raise the spiro bell.
      (2) Turn the valve to Position 2. Lower the spiro bell.
      (3) Repeat steps (1) and (2) until the bag is empty.

   C. Fill the balloon a second time.
   D. Fill and flush the system.

      (1) Turn the 5-way valve to Position 1, and remove the side stopper.
      (2) Raise the spiro bell and replace the stopper.
      (3) Turn the valve to Position 2, and lower the spiro bell.
      (4) Repeat steps (1), (2), and (3) until the bag is empty.

   E. Fill the balloon a third time.
   F. Check the zero and gain (span).

      (1) Set the pump on Position 4.
      (2) Push the AIR and GAS-PMP buttons.
      (3) Set the flow meter on 400.
      (4) Make all readings with the pump off.
      (5) Adjust the CO and He meter readings to zero and check again.

(6) Push the INSP and PMP buttons and let the machine run until the numbers stabilize.
(7) Stop PMP and then run again for a few seconds.
(8) If the helium reading is within ± 1 percent of the percentage certified on the tank, change it to read the same as the certified percentage.
(9) If the CO reading is within 10 percent of 100, change it to read 100.
(10) If both readings exceed the known percentages by ± 1 percent and ± 10 percent, respectively, flush the bag and start over.
(11) Record the zero, span, and gain settings in the log book.

The gain setting and helium reading should not change from day to day by more than ± 3 percent. Any larger deviation on any given day suggests that the balloon has been incompletely flushed or that there are leaks.

Criteria for an Acceptable Exam

Three trials must be done and all three must meet all the criteria listed below. A maximum of five trials may be done in order to meet the criteria. A small amount of CO is taken up by the bloodstream with each trial. Therefore, the blood carboxyhemoglobin saturation slowly rises. As a result, a significant "back pressure" develops after 5-10 trials, depending upon the depth of inspiration, which causes a measurable but small reduction in successive measures of "apparent diffusing capacity."

1. The volumes inspired must be at least 80 percent of vital capacity as determined by spirometry.
2. Inspiration time must not be longer than 3 seconds (9.6 cm.). The only exception to this rule is the case in which all three trials are identical with full examinee cooperation.
3. Breathholding time may not vary more than 1½ seconds from the ideal of 10 seconds (32 cm. ± 4.8 cm.).
4. There must be a minimum dead space washout of 500 ml.; the ideal is 1000 ml.
5. There must be a minimum volume collected of 500 ml.
6. The volume of inspired gas should not decrease more than 200 ml. during breathholding time. A decrease greater than 200 ml. indicates a defective J-valve in the machine.

7. Expired CO and He results from successive trials should not vary by more than 10 percent provided the inspired volumes were the same.
8. There should not be any inhalation artifacts.
9. There should be a smooth 1 second baseline before the onset of inspiration.

Examination Procedure

1. Apply paper to the kymograph drum; overlap right over left and tape. Place examinee label and technician number on the tracing.
2. Empty the balloon. This step is always required when gas has remained in the balloon for more than one-half hour. It should be done twice at the beginning of an examining session.

   A. Set the RV-BB valve in the BB position (OUT).
   B. See that the free-breathing valve is closed ("examinee to room air") (Position 1).
   C. Move the single-breath 5-way valve to Position 3.
   D. Gently raise the spirometer bell by hand to near the top. The bottom of the bell must remain in the water. During this maneuver air enters into the mouth opening of the single-breath 5-way valve, proceeds through Sidearm 3 to the box, thereby permitting gas to be removed from the box through the other orifice into the spirometer.
   E. Turn the single-breath valve to Position 2.
   F. Gently lower the spirometer bell by manual pressure. Note that air from the spirometer now enters the box and cannot escape the box. Therefore, the gas compresses the balloon which now empties via the one-way V-valve to Opening 2 of the single-breath valve and hence escapes through its mouthpiece. Each time the spirometer bell is depressed a decrease in the size of the balloon can be seen through the window in the box.
   G. Repeat Steps C, D, E, and F until the bag is empty, which is signaled by arrest of the descent of the spirometer bell.

3. Fill the balloon.

   A. Turn the RV-BB valve to the RV position (IN). This must always be done when filling the balloon; otherwise, the pressure within the box will force up the spirometer bell.

B. Turn the 5-way valve to Position 1.

C. Connect a rubber tubing from a CO tank to the CO-IN petcock on the front of the panel.

D. Open both petcocks—CO-IN and CO-OUT—during filling. Opening of the CO-OUT petcock permits air to escape from the box while the balloon is filled. If it is not opened the box or associated tubing will rupture.

E. Open the reduction valve on the CO tank. In order to avoid development of excessive pressure in the box, do not permit a flow of more than 10-15 liters/minute. At this rate it will require 2-3 minutes to fill the balloon about half full.The filling can be checked by inspection through the box window. The balloon should be filled until it is well-rounded and its edges touch the sides of the box. but it should not be filled to the point at which it appears to be under pressure. Meteorological balloons vary somewhat in size but, without pressure, they hold approximately 25-35 liters.

F. Close the CO tank valve.

G. Close the CO-IN valve and wait until no more gas escapes from the CO-OUT valve.

H. Close the CO-OUT valve.

I. Return the RV-BB valve to the BB position (OUT).

4. Zero the helium and CO meters.

A. Push the AIR button and GAS-PMP button. The sampling flow rate produced by the gas pump is now indicated on the flow meter.

B. Set the gas pump to Position 4.

C. Set the flow meter on 400.

D. Allow the GAS-PMP to run until the numbers stabilize.

E. Turn the GAS-PMP button off by pushing it a second time. (Note: During all analytical procedures associated with the single-breath procedures, meter readings should be taken while the sample pump is off. When the sample pump is running the pressure in the analytical cells is reduced slightly below atmospheric. Inasmuch as the CO meter reading indicates the total number of molecules of CO within the sample cell, the meter reading will be slightly lower while the pump is running.)

F. Set the helium meter reading to 00.00 ± 0.02 with the Zero Adjust knob.

G. Set the CO meter reading to 000.0 ± 0.2 with the Zero Adjust knob.

H. Push the GAS-PMP button and pump for another 30 seconds. Then repeat the zeroing process.

I. If the CO meter cannot be set to zero by means of the zero control, then the internal shutter on the control chamber within the CO meter will have to be changed. (See Section I and directions in the CO meter manual.) The CO meter and front panel CO readings should be the same.

5. Make the inspired gas adjustments (repeat before each trial).

A. Push the INS-GAS button and the GAS-PMP button and run the sample pump until the numbers stabilize (30-60 seconds). If the machine takes a long time to stabilize, check:

(1) The flow meter (it should be at 400).
(2) The ascarite sample tube (discoloration should not exceed a quarter of an inch), and hardening should not be present.
(3) The drierite sample tube (discoloration should not exceed one-half inch).

B. Stop the GAS-PMP and check the readings.

C. Repeat Steps A and B.

D. If the reading is within 1 or 2 percent of 100 unlock the CO meter gain control and turn it until the meter reads exactly 100. If not, repeat Steps 2-5.

E. If the reading is within 1 or 2 percent of the certified percentage unlock the helium meter gain control and turn it until the meter reads the same as the percentage certified by the gas supplier on the tank. If not, repeat Steps 2-5.

6. Now bring on the examinee.

A. Flush and fill the system with known gas:

(1) Turn the single-breath valve to Position 2.
(2) Lower the spirometer bell by manual pressure.

(3) Turn the valve to Position 1.

(4) Remove the rubber stopper from the examinee's side of the machine.

(5) Raise the spirometer bell and replace the rubber stopper. Note: Do not use weights on the spirometer. Always lift the spirometer bell by hand with the free-breathing valve open to permit gas to enter the spirometer. Opening the free-breathing valve and raising the spirometer bell do not contaminate the system because the spirometer is connected only to the box and not the balloon. Raising the spirometer bell as just described is an essential step. If the spirometer bell is not filled as high as possible with the base submerged in water, then, when the examinee suddenly makes a maximal inspiration the spirometer bell will hit bottom and may crack. Futhermore, water will be sucked into the interior of the spirometer and will run into the box because the RV-BB valve is in the BB position. Such a disaster necessitates the dismantling of the equipment.

B. Be sure the balloon has at least 15-20 liters of gas in it (even more if the examinee's vital capacity is greater than 4 liters).

C. Be sure the RV-BB valve is in the BB (OUT) position.

D. Be sure the spirometer pen is working properly on the paper. Start the pen just left on the paper overlap.

E. Set the paper speed to 1920 mm./minute.

F. Adjust the height of the single-breath valve to the level of the seated-examinee's nose; attach a rubber mouthpiece and insert it in the examinee's mouth. The examinee's neck will be extended upward slightly.

G. With the single-breath valve in Position 1 do a practice trial in the following way:

> Place the nose clip on the examinee's nose and instruct him thusly: "Take in a deep breath. Now blow all the air all the way out. When all the air is out, raise one hand or knock on the side of the machine. Now take in a great big breath as fast as you can and hold it. . .Hold it. . .hold it (10 seconds from start to expiration). Breathe out as fast and as much as you can. Now just breathe quietly."

If the examinee does the trial satisfactorily, proceed with the test.
If the examinee does not do the practice trial satisfactorily, repeat the trial.

H. Push the BAG-PMP button; wait until the sample bag is entirely evacuated and the tubing is flattened; and then push the OFF button to seal the vacuum in the bag.
I. Put one foot on the remote kymograph control, one hand on the single-breath valve, and the other hand around a stopwatch.
J. See that the single-breath valve is in Position 1 and that the examinee is breathing quietly.
K. Ask the examinee to take in a deep breath, blow all the air all the way out, and then signal (put a finger up or tap the side of the machine).
L. At the examinee's signal step on the foot switch to start the kymograph. There should be at least 1 inch of baseline before the onset of inspiration.
M. Turn the single-breath valve to Position 2; ask the examinee to take in a deep breath, and start the stopwatch.
N. When a full inspiration is achieved as rapidly as possible, encourage the examinee to hold his breath. (As long as the single-breath valve is in Position 2 expiration is prevented by the one-way J-valve near the side of the machine.)
O. When 9½ seconds have passed, quickly turn the valve to Position 3 and tell the examinee to exhale as fast and as fully as possible. After a minimum of 500 ml. has been expelled (a fraction of a sec.) turn the valve to Position 4 until the sample bag is nearly full or until expiration has ceased (whichever occurs first). Then return the valve to the original Position 1 and remove your foot from the foot switch.
P. Take the examinee off the mouthpiece, and remove the nose clip.
Q. Number the tracing of each trial at the beginning and the end.

Smooth operation of the valve requires some practice. With normal examinees the valve can be turned from Positions 2 through 4, and 4 back to Position 1 in one smooth motion. This will result in an adequate flush-out and bag sample because of the large volume and rapid flow rate. For examinees with obstructive disease the valve may have to be held in Positions 3 and 4 for one-half second or more. For examinees with a vital capacity of 1400-1500 ml., several trials may be required. If the vital capacity is less than 1400 ml., it will not be possible to do the test because a minimum of 500 ml. is required to flush out the dead space (of the mouthpiece and valve as well as the anatomical dead space of the examinee), and a minimum of 500 ml. more is required in the bag as a sample for flushing of the sampling circuit and the CO and helium cells. Usually an examinee whose vital capacity is less than 1000 ml. cannot hold his breath for 10 seconds anyway.

7. Analyze the sample.

   A. Push the GAS-PMP button and the BAG button once each and wait until the numbers stabilize.
   B. Stop the GAS-PMP and check the readings.
   C. Repeat Steps A and B.
   D. Press the OFF button for the seven solenoid valves.
   E. Record the CO meter reading on the examinee' s chart (Expired CO).
   F. Record the He meter readings on the examinee's chart (Expired Helium Percent).

   For examinees with a normal vital capacity, the bag will contain approximately 1500 ml. Since the pump is set for 400 ml./min., pumping for 1½ to 3 minutes should allow for several readings to be taken. If the sample is small (less than 1000 ml.) reduce the pump flow speed to about 300 ml./min. by regulating the flow control at the bottom of the sample CO and He flow tube. Since the volume of tubing, chemical tubes, and CO and He analysis cells is about 300 ml., a reliable reading will not be possible if there are less than 400 or 500 ml. of gas in the bag.

8. At least 5 minutes must be allowed between trials to assure complete removal of all CO and He from the lung gas. For examinees with marked obstructive disease, it may be wise to wait even a little longer.

Recording of Temperatures and Pressure

1. Record the Uncorrected Barometric Pressure to the nearest tenth of a mm. of mercury on the examinee's chart.
2. Record on the chart the room temperature to the nearest degree centigrade from the thermometer mounted on the barometer. Record it in the space to the right of the Uncorrected Barometric Pressure. This temperature is used to correct the barometric pressure.
3. Record on the chart the Small Spirometer Temperature to the nearest degree centigrade from the thermometer on the CO tubing system. This temperature is used in the calculation of the STPD factor.

Interpretation of Spirogram

1. Data from the spirogram

    A. Inspired volume (VC ATPS) in ml. This is calculated and recorded on the examinee's chart.
    B. Breathholding time. This is measured and recorded on the chart in cm. as the distance from the midpoint of inspiration to the onset of gas collection.
    C. Volume of dead space washout (minimum of 500 ml.). This is not recorded on the chart.

2. Procedure for getting this data

    A. Set the proportional divider for lines at a 1:2 ratio.
    B. Extrapolate the baseline of the tracing until it reaches a point where the line is perpendicular to the peak volume point of the spiro tracing. If there is a decrease in volume of more than 200 ml. during breathholding, use the mean of the volume from beginning to end of breathholding as the peak volume point. Small artifacts at the peak of inspiration should not be interpreted as peak volume as they are due to the inertia of the spirometer bell.
    C. Put one arm of the proportional divider on the baseline under the peak volume point and the other at the peak volume point of the spiro tracing.

D. Without altering the divider setting, remove the divider from the tracing, put one arm on a vertical line at zero and the other arm on the same line above it. Read the number of ml. at this point to get the inspired volume (VC ATPS) and record this volume on the tracing.

E. Without altering the divider setting, turn the divider over and place one arm on the baseline and the other on a point directly above.

F. Draw a horizontal line from this point parallel to the baseline until it intersects the spiro inspiration tracing. This intersection is the one-half inspiration point.

G. Draw a vertical line through the small artifact at the point where the washout tracing ends and the collection of gas in the sample bag begins.

H. Place one arm of the divider on the tracing at the point where gas collection begins and the other arm directly above at the level of the peak volume. This is the washout (dead space) measurement and must be at least 500 ml.

I. Place one arm of the divider on the tracing at the point where gas collection begins and the other arm on the extended baseline. This is the air sample collected and must measure at least 500 ml.

J. Measure and record the distance from the vertical line drawn through the artifact to the one-half inspiration point with a metric ruler. Make sure the ruler is parallel to the baseline. This is the breathholding time in cm. The computer program will divide the cm. results by 3.2 (the paper speed of the kymograph is 32 mm./sec.) to get breathholding time in seconds.

K. If inspiration time is hard to determine, extrapolate lines from the major (most constant, longest, and most obvious) section of the inspiration tracing. Zero inspiration is the point at which the extrapolated line intersects the baseline. Termination of inspiration is either the point at which the extrapolated line crosses a horizontal line tangent to the total volume peak, or the point at which a normal-looking inspiration curve reaches a peak and levels off with only a slight gradual increase.

Optional Check Procedures

1. Checking the tank gas against the balloon gas.

   A. Completely remove the sample bag from the single-breath valve and from the machine.
   B. Remove the rubber tubing from the CO-IN petcock and connect it to the sample bag.
   C. Open the CO tank valve and allow the gas to pass in and through the sample bag, thereby flushing the bag but still retaining a reasonably large sample.
   D. Close the CO tank valve.
   E. Seal off both ends of the sample bag.
   F. Remove the tubing from the sample bag.
   G. Reattach the sample bag to the machine; keep the other end sealed off.
   H. Push the BAG and GAS-PMP button and allow the machine to run until the numbers stabilize.
   I. Turn the GAS-PMP off.
   J. Repeat Steps H and I and then check the readings. The readings obtained with gas directly from the tank should be the same as the readings obtained from inspired gas in the bag.

2. Checking the gas in the tubing and system against the gas in the tank.

   A. Remove the end of the sample bag tubing from the machine inlet.
   B. Attach one end of a 2-3 foot piece of tubing of the appropriate size to this inlet.
   C. Attach the other end to the petcock on the 5-way valve, and open the petcock.
   D. Push the BAG and PMP buttons and allow the machine to run until the numbers stabilize.
   E. Turn the GAS-PMP off.
   F. Repeat Steps D and E and check the readings. They should be the same as the readings obtained from the gas in the bag (inspired gas).

Appendix B. Excerpt from NCHS Instruction Manual. Health Examination Survey—Cycle III, Field Staff Procedure Manual.

CHAPTER 18.

TREADMILL TEST

Restrictions

Subjects are to be restricted from performing the Treadmill Test

1. If there is a restriction on exercise or hard work imposed either by the parents or by the subject's physician.
2. If the subject is known to be pregnant.
3. If the subject has an oral temperature of 100° or greater.
4. If the examining staff physician feels there is any medical contraindication to performance of the test.

Restrictions will be noted on the Control Record. The examining staff physician is to be consulted on any questionable subject.

Equipment

3-Electrode kit (Beckman #350096).
Electrode adhesive collars (Beckman# 350177, 50 rolls, 100/roll).
Coupler Cable (15 ft. shielded connecting cable, Beckman Type 9806A for A-C Coupler).
Type RS or RP Dynograph, Beckman, Direct Writing Recorder, Simpson Model 29.
Littlefuse fuses, 3AG, 3-2/10, SB.
Ink (Offner Dynograph ink).
Paper (Beckman Instruments, Inc., #21707).
Keystone grease (#29, X-Light).
Keystone oil (S.R. #1 Specialized Lubricants Oil).
Floor wax (Powdered Cook's Dance Floor Wax).
2-levels.
Wooden template.
Grounding wire.
Timing clock.
Stop watch.
Thermometer.

January 1967

Hygrometer.
Humidifier.
Air conditioner.
Acetone.
4 x 4 gauze sponges.
Hemostat.
Kontax EKG Cream.
2 - 5 cc. syringes with needles of gauge # B-D 18.
Terry cloth wash rags.
Distilled water.
Smelling salts.

Exercise Room Procedure

Temperature and Relative Humidity

Ambient testing conditions 70-74^0 F. and 50-60 percent relative humidity. (The proper temperature range is more critical than relative humidity range.)

Before Subject Enters Exercise Room

1. Check level of treadmill at 0-percent grade and the achievement of the proper 10-percent grade by use of the wooden template once a week. Headquarters in D.C. must notify W. E. Quinton Instrument Co., Seattle, if any correction in the percentage grade is necessary. Correct any deviation in leveling.
    a. If possible, relevel by leveling the trailer.
    b. If for some reason the trailer cannot be made level, then level the treadmill with wooden sheets.
2. Check calibration of treadmill speed by timing the revolutions of the belt with a stop watch (10 rev. = 22.8 sec.).
3. Calibrate tachometer for 60-120-180-240 BPM

Subject in Exercise Room Before Test

1. Scrub electrode placement sites briskly with acetone soaked sponge on hemostat to decrease skin resistance.

2. Apply EKG paste to electrodes, using syringe (use enough, applied through center hole to fill the 3 holes and the well).
3. Apply electrodes to the following sites:
   a. Manubrium—white lead
   b. Anterior arc of fifth rib—red lead
   c. Xyphoid process—black lead
4. Connect three-prong connection to the coupler wire.
5. Have subject stand at front of treadmill with arms at side—take short resting EKG strip (25 mm/sec.—15-20 clear complexes).

Exercise Test

5 minutes—Speed 3½ mph. (Calibrate by belt speed at least two times per day; for 3½ mph, 10 belt revolutions = 22.8 seconds.)

Time (min.)	*Grade (%)*	*Speed (mph)*
0-2	0%	3½
2-5	10%	3½

Give following explanation while subject is standing on the mill:

"This is a short, easy test of physical fitness which we give to all the boys and girls so that we can tell what state of physical condition you are in. It consists of walking for 5 minutes on this treadmill and it is an easy test to do. I will stay right here beside you during the entire test."

In addition to the above, positive encouragement is essential for an optimal reaction on the part of the subject. "You are doing very well . . ." should be mentioned to the examinee frequently during the test.

Specific Instructions

1. Have subject hold onto front bar with both hands.
2. Warn subject that treadmill will start on the count of "3." Count "1-2-3."
3. Coach subject during walk while he (she) holds onto the bar.
   a. Natural strides—hips straight, head up.
   b. Land on heels.
   c. Bear body weight on the feet—not on the bar with hands.
   d. Release hands from bar and walk with natural arm swing.

Common Faults

1. Waiting after start of treadmill before starting to walk. Belt carries feet back and subject is off balance.
2. Short choppy strides.
3. Bending forward at waist.
4. Head down looking at feet.
5. Nervousness.

With proper coaching the subject should walk naturally and fully adapted within 1½ minutes.

1. Raise treadmill to 10-percent grade at 2:00 minutes.
2. Have subject hold front bar just before treadmill is stopped at 5 minutes.
3. Record directly on the EKG tracing:
   - A. A subjective evaluation of the walking gait viz "good" or "poor." Record this after 5-minute heart rate record
   - B. Any differences in the subject from the normal after the 5-minute heart rate record, e.g.,
     - a. Subject nervous
     - b. Subject held the bar frequently
     - c. Subject was awkward
     - d. Subject had balance problems
     - e. Subject repeated test

Electrocardiograms are sent weekly to
Dr. Jack Alexander
Lab. Physical Hygiene
Stadium Gate 27
University of Minnesota
Minneapolis, Minnesota 55414

Cardiotachometer Procedure

Calibration

1. Paper speed - 2.5 mm./sec.
2. Motor dial - to "operate" position.
3. Calibration dial - 60 BPM, 120 BPM, 180 BPM, 240 BPM.

Align stylus to 5 mm. above right side of paper with the "centering" dial to coincide with calibration dial indication of 60.

Set calibration dial to 120. Stylus should reach 25 mm. line—adjust with the "range" dial.

Set calibration dial to 180. Stylus should reach 45 mm. line—adjust with the "range" dial.

Set calibration dial to 240. Stylus should reach 65 mm. line—adjust with the range dial.

Establish linearity between 60 (e.g., 15 mm.) and 240 (65 mm.) using "centering" and "range" dials, respectively.

Operation

1. Leave motor dial on "operate" position.
2. Set calibration dial to "direct" position and adjust sensitivity.
   a. Adjust preamp. # 1 to 5 or 2 or 1 (MV/CM) position or until the qrs EKG spikes are approximately 10 mm. high.
   b. Adjust preamp # 2 to 10 or 5 or 2 (MV/CM) position or until the qrs EKG spikes are approximately 2 mm. high.
   Leave all other dials alone.
3. Set calibration dial to "RATE" position.

Record

1. Calibration no. 1 at 60-120-180-240 BPM (paper speed 2.5 mm./sec.). 15 mm. is sufficient for each calibration line.
2. Rest (standing) (25 mm./sec.).
3. At one minute check speed of treadmill with stop watch.
4. Recording times and paper speeds:

0:50 to 1:00 minute 1:50 to 2:00 minutes 2:50 to 3:00 minutes 3:50 to 4:00 minutes 4:45 to 5:00 minutes	Exercise at paper speed 5 mm./sec.
5:00 to 5:10 minutes	Exercise at paper speed 25 mm./sec.

5. Calibration no. 2 after finishing walk.

January 1967

NOTE.—Recording 4:45 to 5:10 minutes is the critical one. Make sure it is accurately timed, and include on the record the exact point at which the subject has walked in 5 minutes. This is done with a check mark through the EKG.

Care of Equipment

Electrodes

1. To detach the electrodes from the subject, proceed as follows.
   a. Disconnect the electrode leads from the input cable of the RS Dynograph.

   CAUTION: Never pull on a lead to remove an electrode or for any other reason. Always protect the lead-to-electrode junction from excessive stress.

   b. When the leads are free, detach the electrodes from the subject by pulling firmly and quickly on the edge of the adhesive collar.
   c. Use finger and thumb pressure directly on the electrode body to pull it clear of the adhesive collar (see caution above) and discard the collar.
   d. If the electrodes are to be used again immediately, wipe their faces on clean cloth and immerse in distilled water. If the electrodes are not to be used again immediately, clean them as soon as they have been detached from the subject.
2. Clean the electrodes as follows:
   a. Submerge the electrode in the beaker of distilled water.
   b. Use the syringe and beaker water to flush each hole in the submerged electrode several times. Avoid scratching the electrode sensor pellet with the needle.
   c. Remove the electrode from the water, pat it dry on clean absorbent cloth, and expose it to room air for 5 minutes. At the end of 5 minutes, the electrode will be ready to be reloaded with electrolyte or stored.
   d. Fit adhesive collars onto clean, dry electrodes.

NOTE:—The above is in part from "Biopotential Skin Electrode, Instruction Manual," published by Spinco Division of Beckman Instruments, Inc., Stanford Industrial Park, Palo Alto, California, # 0-TB-002, October, 1965.

January 1967

Treadmill

The treadmill is a sturdy piece of equipment which should need very few repairs and adjustments. The following items should be noted:

1. *Waxing.* The slipway (surface beneath belt) should be waxed once per month using two or three shakes from the wax can (powder). Do *not* put excess wax on the slipway.
2. *Lubrication.* It need be greased only once per year (Keystone Grease). There are three nipples located near the driveshaft about three-fourths to the rear of the treadmill. Oil (Keystone) need only be added when there is a leak from the crankcase. Oil is added through the small brass hex nut (hole in center) at the left rear side above the drive chain guard box.
3. *Belt tracking.* The belt should be adjusted only when it slips loosely over the drums or slips (tracking) to one side of the slipway surface. To correct slipping on the drums, use a key and lengthen both screw adjustors located on either side of the front drum. To correct tracking, lengthen the screw adjustor on the side to which the belt has tracked.
4. *Transportation.* The treadmill should be bolted down with the bolts provided.

Cardiotachometer

At the end of each Stand the ink should be drained from the cardiotachometer, and it should be flushed with distilled water. The cardiotachometer should then be stored in the special crate provided.

Chapter **3**

Assessments of Body Composition, Dietary Patterns, and Nutritional Status in the National Health Examination Surveys and National Health and Nutrition Examination Surveys

Dorothy Blair, Ph.D.
The Pennsylvania State University

Jean-Pierre Habicht, M.D., Ph.D.
Cornell University

Lee Alekel, M.S., R.D.
The Pennsylvania State University

Introduction

This chapter presents a developmental and methodological perspective on the National Health Examination Surveys (NHES) and the National Health and Nutrition Examination Surveys (NHANES), carried out from 1959 to 1985. It focuses on the measuring and monitoring of the Nation's nutritional status using dietary, biochemical, and body composition variables. The issues addressed are the following: How the survey series has evolved over the last 25 years and the factors that have defined and limited the methodologies chosen to assess nutritional status in a National Nutrition Surveillance Survey (NNSS); what the data from these surveys reveal about the Nation's nutritional health; and finally, measurement and design issues that reduce the sensitivity, specificity, and reliability of our estimates of nutritional status and the means by which these might be improved.

Tables 1–3 present in detail the body composition, dietary, physical fitness, and biochemical variables that were collected for each survey of the NHES–NHANES series from 1959 to the present. The tables reflect the original protocol for each survey. They include both those variables that are actually available for analysis and those that were included in the survey's published protocol but were disregarded because of methodological, reliability, or analysis difficulties. Variables that are not available on public use data tapes at the time of this writing are presented in brackets.

Table 1. Anthropometric and x-ray measures used in the National Health Examination Survey (NHES) and National Health and Nutrition Examination Survey (NHANES): 1959–80

				NHANES I			
Item	NHES I, 1959–62	NHES II, 1963–65	NHES III, 1966–70	NNSS,[1] 1971–74	NHES IV, 1971–74	Augmentation, 1974–75	NHANES II, 1976–80
Age in years	18–79	6–11	12–17	1–74	25–74	25–74	6 mo.–74
Number examined	6,672	7,119	6,773	[2]16,895	3,854	3,059	20,322
Percent of sample examined	86.5	96.0	90.1	74.0	70.0	71.3	73.1
Anthropometric measurement							
Height, standing	[3] ****	____ ****	____ ****	____ ****	____ ****	____ ****	____ ****
Weight	****	____ ****	____ ****	____ ****	____ ****	____ ****	____ ****
Number of:							
Heights	6	4	10				
Lengths	2	6	3				
Breadths-depths	2	6	12	Bitrochanteric, elbow	****	____ ****	____ ****
Diameters	1	3	0				
Girths	3	6	6	Upper arm	____ ****	____ ****	____ ****
Circumferences	0	0	0	Chest (1–7 yr)	Chest: Inspiration	____ ****	____ Chest: Erect (2–7 yr) Chest: Supine (0–3 yr)
				Head (1–7 yr)	Expiration	____ ****	Head (0–7 yr)
Skinfolds	[3]Triceps	____ ****	____ ****	____ ****	____ ****	____ ****	____ ****
	[3]Subscapular	____ ****	____ ****	____ ****	____ ****	____ ****	____ ****
			Lateral chest wall ****				
			Suprailiac				
			Medial calf				
X ray							
Bone:							
Age		Hand-wrist	____ ****	____ **** (1–17 yr)	____ ****		
Cortical thickness					Phalanx	____ ****	
Density					Phalanx	____ ****	
					Radius	____ ****	
Joint:							
Chondrocalcinosis					Hip, knee	____ ****	
Osteoarthritis/rosis	Hands, feet				Hip, knee & Sacroiliac	____ ****	[cervical & lumbar spine (25–74 yr)]
Rheumatoid arthritis	Hands, feet						
Sacroiliitis					Hip	____ ****	

[1] National Nutrition Surveillance Survey.

[2] The total number of those persons receiving the nutrition portion of NHANES I is 20,749, which includes the NNSS and NHES IV samples. The NHES IV sample also received a detailed examination component.

[3] Triceps referred to as "upper arm" and subscapular as "infrascapular" in NHES I and II.

Notes: **** indicates that the measure was collected for that survey.
____indicates continuity between surveys.
[ ] indicate that the data have not yet been released on public use data tapes. For a few specific indexes, data may be obtained with permission from the National Center for Health Statistics.

Data for certain variables in NHANES were collected on subsamples of the total examined population. The age ranges for the subsamples are noted in the tables next to the variable. In some cases, the subsampling techniques are based on more extensive information than age and are therefore explained in the text.

Unbroken lines in tables 1–3 indicate continuity of data collection across surveys. Although all the surveys have a certain continuity of core variables, the concept of surveillance of the changes in the Nation's health was not a consideration in the original NHES design. Maintaining data continuity among surveys is a more recent concern, beginning with the mandate for national nutrition surveillance in the late 1960's.

Readers should refer to tables 1–3 throughout the following discussion of the evolution of the NHES–NHANES series. These tables will also be useful references for the subsequent discussion of the factors that

Table 2. Dietary-nutrition information and physical fitness-activity parameters used in the National Health Examination Survey (NHES) and National Health and Nutrition Examination Survey (NHANES): 1959–80

				NHANES			
Item	NHES I, 1959–62	NHES II, 1963–65	NHES III, 1966–70	NNSS,[1] 1971–74	NHES IV, 1971–74	Augmentation, 1974–75	NHANES II, 1976–80
Age in years	18–79	6–11	12–17	1–74	25–74	25–74	6 mo.–74
Number examined	6,672	7,119	6,773	[2] 16,895	3,854	3,059	20,322
Percent of sample examined	86.5	96.0	90.1	74.0	70.0	71.3	73.1
			Dietary-nutrition information				
Eating habits-patterns (qualitative, descriptive):		****____	****____	****			
Alcohol				Consumption (12–74 yr)			Liquor, wine consumption Liquor, beer frequency
Coffee-tea (frequency)				****____	****		****
Questionnaires:							
Food programs				****____	****		****
24-hour dietary recall				****____	****		****
Dietary frequency				****____	****		****
Water usage [3]						****	
Supplementary dietary questions				****____	****	****____	**** (12–74 yr)
Vitamin-mineral usage				****____	****		****
			Physical fitness parameters				
Cardiovascular function:							
Blood pressure	****____	****____	****____	****____	****____	****	
Cardiotachometer (pulse rate)			****				
Electro-cardiogram	****____	****____	****		****____	****____	[****] (25–74 yr)
Exercise tolerance:							
Bicycle ergometer		****					
Treadmill			****				
Pulmonary function-characteristics:							
Carbon monoxide (pulmonary diffusion)					****____	****	
Chest x rays	****____	****____	****		****____	****____	**** (20–74 yr)
Spirometry		****____	****____	****(6–24 yr) ____	****____	****____	[****] (6–24 yr)
Goniometry (range of motion) [4]					****		
Grip strength:							
Dynamometer		****____	****				
Smoking behavior			****		****____	****____	**** (12–74 yr)
Physical activity-exercise behavior (subjective)				**** (12–74 yr)____	****____	****____	**** (12–74 yr)

[1] National Nutrition Surveillance Survey.

[2] The total number of those persons receiving the nutrition portion of NHANES I is 20,749, which includes the NNSS and NHES IV samples. The NHES IV sample also received a detailed examination component.

[3] Water usage data were collected but are not available because of methodological problems.

[4] Goniometry data were collected only for the first 36 stands. These data are available.

Notes: ****indicates that the measure was collected for that survey.

____ indicates continuity between surveys.

[ ] indicate that the data have not yet been released on public use data tapes. For a few specific indexes, data may be obtained with permission from the National Center for Health Statistics.

Table 3. Biochemical indexes used in the National Health Examination Survey (NHES) and National Health and Nutrition Examination Survey (NHANES): 1966–80

		NHANES I			
Item	NHES III, 1966–70	NNSS,[1] 1971–74	NHES IV only, 1971–74	Augmentation, 1974–75	NHANES II, 1976–80
Age in years	12–17	1–74	25–74	25–74	6 mo–74
Number examined	6,773	[2] 16,895	3,854	3,059	20,322
Percent of sample examined	90.1	74.0	70.0	71.3	73.1
Whole blood					
	Hematocrit	****	****	****	****
		Hemoglobin	****	****	****
	Blood type				
	Protein-bound iodine				
	Red cell antigens				
		[3] RBC morphology	****	****	[****]
			Platelet estimate	****	
		[4] RBC count	****	****	****
		WBC count	****	****	****
		[3] WBC differential	****	****	[****]
		Sed. rate	****		
					Lead
					Protoporphyrin
					[3] [RBC folate]
					Methemoglobin
					Carboxyhemoglobin (3–74 yr)
					[Abnormal hemoglobin]
Urine					
	[Testosterone (boys)]				
	Bacteriuria (girls)				
		Creatinine	****		
		Iodine	****		
		Riboflavin	****		
		Thiamin	****		
					[Specific gravity & urinary sediments (20–74 yr)]
					[Gonorrhea cultures] (12–40 yr)
					[Pesticide residues & metabolites (12–74 yr)]
		N-Multistix (dipstick) (6–74 yr):			
		Acidity (pH)	****	****	[****]
		Albumin protein	****	****	[****]
					[Bacteriuria (nitrite)]
				Bilirubin	[****]
		Glucose	****	****	[****]
		Hematuria	****	****	[****]
				Ketones	[****]
				Urobilinogen	[****]
Serum/plasma					
	Cholesterol	(4–74 yr)	****	****	(20–74 yr)
					[Triglycerides & HDL (20–74 yr)]
	Uric acid		Uric acid	****	
		Protein, total &	****		
		Albumin (4–74 yr)	****		Albumin (3–74 yr)
					[Vit B_{12}]

Table 3. Biochemical indexes used in the National Health Examination Survey (NHES) and National Health and Nutrition Examination Survey (NHANES): 1966–80—Continued

		NHANES I			
Item	NHES III, 1966–70	NNSS,[1] 1971–74	NHES IV only, 1971–74	Augmentation, 1974–75	NHANES II, 1976–80
			Serum/plasma—Con.		
		Iron (4–74 yr)	****		Iron
		Transferrin sat &	****		Transferrin sat
		TIBC (4–74 yr)	****		TIBC
					[3] [Ferritin (3–74 yr)]
					Copper & zinc (3–74 yr)
		[5] Folates	[5] ****	Folates	****
			Calcium	****	
		Magnesium (4–74 yr)	****	****	
			Phosphate	****	
				Potassium	
				Sodium	
		Vit A			Vit A (3–11 yr)
		[5] Vit C			Vit C (3–74 yr)
					[GTT (20–74 yr)]
				Kidney function tests:	
				BUN	
				Creatinine	(12–74 yr)
			Liver function tests:		
			Alkaline phosphatase	****	[6] [****]
			Bilirubin, total	****	[6] [****]
			SGOT	****	[6] [****]
					[Bile acids (35–74 yr)]
			Thyroid function tests:		
			T_3 & T_4		
					[Syphilis, pesticide residues, & metabolites (12–74 yr)]

---

[1] National Nutrition Surveillance Survey.

[2] The total number of those persons receiving the nutrition portion of NHANES I is 20,749, which includes the NNSS and NHES IV samples.

[3] Tests were performed on 10% of the sample and on those individuals with abnormal hemoglobin criteria, as explained in references 7 and 8.

[4] Approximately 20–25% of the RBC count data were discarded because of methodological difficulties.

[5] Folate and vitamin C analyses are not available on the NHANES I data tapes.

[6] Tests were performed only on those individuals with elevated bile acids.

Notes: **** indicates that the measure was collected for that survey.
____ indicates continuity between surveys.
[ ] indicate that the data have not yet been released on public use data tapes. For a few specific indexes, data may be obtained with permission from the National Center for Health Statistics.

determine variables chosen to be representative of nutritional status in the United States.

Evolution of the NHES–NHANES series

Survey history

The National Health Survey Act of 1956 (Public Law 84–652) authorized the creation of an ongoing National Health Survey and special studies to secure "accurate and current statistical information on the amount, distribution and effects of illness and disability in the United States." The focus of the first of the special studies was on obtaining information about debilitating diseases such as heart disease, cancer, diabetes, arthritis, and rheumatism.

In the first through third cycles of NHES (NHES I through NHES III) the population of the United States was surveyed by age group, and the information collected was tailored to the health concerns of the age group being examined. From 1959 through 1970, special mobile examination caravans toured the United States almost continuously, interviewing and examining a systematic stratified sample of persons from households within systematically chosen geographic areas representative of the United States.

The core information that was common to each survey included medical and dental examinations, blood pressure and an electrocardiogram, hand-wrist and chest x rays, weight, height, girth and skinfold thicknesses, visual and auditory acuity, household composition, and demography. In NHES I, adults 18–79 years of age were examined, with the focus on degenerative diseases; the biochemical analyses performed were limited to whole blood glucose, a hematocrit, a serologic test for syphilis, serum cholesterol, urine albumin (males only), and glucose (1). In NHES II, children 6–11 years of age were examined, with the focus on growth; physical, intellectual, and psychological development; and fitness. No blood or urine analyses were performed, but a limited qualitative assessment of food consumption and eating habits was added to the original NHES protocol (2). In NHES III, youths 12–17 years of age were examined; the format was similar to that of NHES II, with information on sexual maturation added. Blood and urine were more extensively analyzed in NHES III than in NHES I (3). Because so few biochemical measures were collected in NHES I and II, these surveys were omitted from table 3.

The data from the NHES series have been published in 103 primarily descriptive reports, each dealing with one particular health, educational, or developmental parameter, usually cross-tabulated by race, sex, age, income, location, or other descriptive variables. Regression techniques have been used sparingly to delineate disease or developmental relationships. The purpose of these reports has been to make health statistics available, rather than to interpret these statistics for policy purposes (4).

The original mandate for creating NHES did not include nutritional health as an examination priority. However, by the mid-1960's, reports of malnutrition existing in the United States shocked the public and raised congressional consciousness. In 1968 the documentary film "Hunger in America" appeared on national television. The report of the Citizens' Board of Inquiry into Hunger and Malnutrition, *Hunger-USA*, was released (5), documenting cases of malnutrition among Mexican Americans, American Indians, black people in Mississippi, and Appalachian white people. These and other inquiries, plus widespread media attention, led to the creation of the Senate Select Committee on Nutrition and Human Needs (6) and spurred the congressional mandate for a national nutrition survey to determine the "incidence and location of serious hunger and malnutrition in the U.S." (7).

The Nutrition Program of the National Center for Chronic Diseases, Department of Health, Education, and Welfare (DHEW), took responsibility for the survey. With an inadequate budget and political time pressures, a random sample of the U.S. public could not be obtained. The Ten-State Nutrition Survey (TSNS), as it came to be called, was carried out from 1968 to 1970 in a sample of 10 States thought to have a high proportion of low-income residents. Sampling was heaviest in areas with high concentrations of migrant laborers, southwestern Spanish-speaking people, and inner-city slumdwellers. Sampling was based on 8–10-year-old census data (8).

Before it was completed, the deficiencies of the TSNS were clear. TSNS exemplified the difficulties of assessing nutritional status with existing methodologies and of obtaining statistically reliable assessments of the nutritional status of the U.S. population. In 1969 President Nixon gave the job of a continuing nutrition surveillance program to DHEW. The National Center for Health Statistics (NCHS) formed a task force on nutrition surveillance that determined that a National Nutrition Surveillance Survey could be conducted in conjunction with NHES. Thus, NHES IV became part of the first National Health and Nutrition Examination Survey, or NHANES I (9).

Planning for NHES IV was almost complete when the NNSS was added. Therefore, NHES IV retained a separate identity within NHANES I. A subsample of the persons participating in NHANES I received the more extensive NHES examination. Unfortunately, in NHANES I, a substantial decrease from the previous surveys occurred in the percent of sample persons electing to participate in the examination portion of the survey. The 3,854-person sample size achieved for NHES IV during the 1971–74 period of data collection was not sufficient. A 14-month Augmentation Survey was required so that the detailed examination component could be performed on 3,059 more individuals, for a total of 6,913 NHES examinees (10, 11).

NHES IV was designed to assess the health, perceived health care needs, and the use of health care in adults ages 25–74. The survey focused on chronic and debilitating disease conditions, especially arthritic, respiratory, and vascular disease. The NHES IV protocol was shortened to accommodate the NNSS. The NHES IV Augmentation Survey was based on approximately the same protocol, although dermatologic and ophthalmologic tests were dropped. In order to test procedures for NHANES II, other variables were added, such as serum calcium and the mineral content of drinking water.

For the nutrition phase of NHANES I, 20,749 individuals ages 1–74 received the nutrition examination component. The elderly, women of childbearing age, children, and persons at or below the poverty level were sampled at rates that were substantially higher than their representation in the U.S. population. These persons were considered to be at higher risk of nutritional deficiencies than the general population and therefore were oversampled to improve the reliability of the statistical analysis of their data.

The choice of nutrition variables to be included in the NNSS portion of NHANES I was made only after consultation with many private individuals and persons from agencies within the Government. Those at the Centers for Disease Control who had been involved with the TSNS and those within the Department of Agriculture who had participated in the planning of the Nationwide Food Consumption Survey were particularly valuable resources. Tests, measurements, and clinical assessments were specifically tailored for the purpose of assessing nutritional status. The skin, mouth and gums, musculoskeleture, and organ examinations concentrated on uncovering lesions associated with the deficiencies of essential nutrients. Anthropometric indexes in NHES I and II had been designed in part to measure body frame lengths, widths, and diameters for clothing and seating manufacturing needs. Anthropometric indexes in the NNSS for all ages were gauges of obesity, arm muscle mass, and body build. They were also measures of early protein-calorie malnutrition in children. The selected biochemical indexes of nutritional status reflect the nutrition concerns of the early 1970's: Urinary riboflavin, thiamin, and iodine; serum vitamin A, vitamin C, folacin, magnesium, protein, and albumin; multiple indexes of iron status; and fluoride (by microbiopsy of tooth enamel). Later, most folacin data, the vitamin C data, and the red blood cell count were discarded because of methodological problems. For the first time, quantitative measures of dietary intake for estimating nutrient consumption were added to the NHES series, as well as a food frequency questionnaire and questions as to Federal food program utilization (10). (See tables 2 and 3.)

NHANES II: Continuity versus flexibility

A major question in the development of successive surveys in the NHANES series is whether the survey's major function is to monitor changes in the nutritional status of the population over time with maximum continuity or whether the surveys should be flexible enough to respond to newly defined nutritional concerns and methodological improvements that allow more precise determinations of nutrition parameters. Each survey in the series has provided the proving grounds for subsequent surveys, allowing modifications in survey protocol that enhance the utility of the data collected (12).

Flexibility is desirable to improve the survey, but excessive flexibility erodes the surveillance function of the survey. Samples must remain comparable, and designated core variables should continue to be measured in a comparable manner over time. A good example of these considerations is the elimination of vitamin A as a variable for adults in NHANES II after it was demonstrated in NHANES I that low serum vitamin A levels were not a public health problem. Some people have deplored the deletion of a variable that did not currently seem important because of the possible beneficial relationship of provitamin A with cancer. This relationship became apparent shortly after NHANES II was implemented. These people feel that NHANES lost an important opportunity to provide baseline data on the relationship among dietary provitamin A, serum vitamin A, and disease.

Careful examination of the dietary interview schedule and of biochemical variables collected reveal that substantial changes would have had to be made in both for either useful monitoring of provitamin A ingestion or relating provitamin A ingestion to morbidity or mortality in individuals. The conclusion from this example is that it is not useful to collect data before their use and the level of appropriate statistical analysis are defined. Thus the elimination of vitamin A in adults was justified. The questions are, How will the data be used and what are the best means to collect information about provitamin A for those uses?

The trend in the planning for subsequent NHANES surveys is to try to achieve the goals of both continuity and flexibility, with the emphasis on flexibility. In NHANES II, a nutritional status assessment protocol similar to that of NHANES I continued to be used, but many variables were added that reflect new nutrition concerns. The scope of our concept of nutrition health changed in the 1970's. In 1974, the Senate Select Committee published their dietary goals, which reflected the increasing concern about the role of overconsumption of macronutrients in the aggravation of degenerative diseases (6). Knowledge of the interactions between vitamins and minerals and disease states was beginning to blossom. In response, the scope of NHANES II became correspondingly broader, still maintaining a core interest in undernutrition and obesity but also focusing on nutrient-disease relationships, environmental contaminants, and overconsumption of vitamins and minerals. Anemia, which was shown to be a particular health problem in NHANES I, was afforded a more indepth assessment in NHANES II.

Increased collaboration between NCHS and other Federal agencies has provided pressure for flexibility rather than continuity. With the added resources available from collaborative data collection and analysis efforts, the scope and complexity of the data collected could increase dramatically. Specialized offices such as the Office of Pesticides and Toxic Substances (Environmental Protection Agency), the Bureau of Laboratories (Centers for Disease Control), the Bureau of Foods (Food and Drug Administration), the National Heart, Lung, and Blood Institute and the National Institute of Arthritis, Metabolism, and Digestive Diseases (National Institutes of Health) were able to design the survey protocol related to their special interest area, implement the analysis, and in some cases provide special funding. NHANES II could then be a source of high-quality information to be used as baseline data for further studies.

A major change in data analysis expectations has also influenced the scope of NHANES II, compared with that of past surveys in the series. Data tapes from the NHES I–III series were not released to the public until 1984. Data were analyzed in house and released to the public in the form of descriptive publications on the incidence of single health or developmental variables. NHANES I data tapes were released to the public only after certain in-house descriptive studies were published. Relative to other surveys, NHANES II data have been rapidly released to the public without waiting for the publication of in-house descriptive studies. The expectation is that most of the data analysis will be done on a contractual basis with agencies and universities or by individuals with their own particular health and nutrition interests. Greater public participation in data analysis has slanted the analysis emphasis toward the definition of the interrelationships among nutrition variables and between nutrition variables and measures of health status. NHANES II has been adjusted to provide the variables that make this type of analysis more feasible.

Modifications of NHANES II

By the time NHANES II was initiated in 1976, the objectives of NHES and the NNSS were completely integrated. The examination focused on four new target conditions: Diabetes, kidney disease, liver disease, and allergy. Data were also collected related to nutrition, cardiovascular conditions, lung function, osteoarthritis, hearing and speech pathology, and environmental pollutants such as lead and pesticides. There was no detailed examination component for a subsample of the examinees in NHANES II. Rather, the overall survey protocol varied by age group. Spirometry and allergy tests were reserved for persons aged 6–24 years; only adults aged 25–74 received x rays. In some cases the sample was split in two. Half of the group aged 20–74 years received serum and urine pesticide tests, and the persons in this group aged 35–74 years received a postprandial liver bile acid test. The other half of the group aged 20–74 years underwent a 2-hour glucose tolerance tests and received triglyceride and high-density lipoprotein determinations (12).

New nutrient variables added to NHANES II included serum zinc and copper. The importance of these trace minerals as part of enzyme systems and the extensive nature of their interaction with disease processes was only fully realized only by the 1970's. Vitamin C was analyzed successfully in NHANES II. Other new biochemical determinations included pesticide residues in serum and urine in all persons aged 12–19 years and in those aged 20–74 years who received the glucose tolerance test. Whole blood lead, carboxyhemoglobin plus methemoglobin determinations were performed on odd- and even-numbered examinees, respectively (12).

Data from NHANES I had shown that anemia was a significant public health problem in the United States. Therefore, anemia was accorded special detailed attention in NHANES II through biochemical assessments of blood, medical history questionnaires, and physician's examinations. A complete red blood cell morphology was performed as well as an assessment of iron stores and iron-binding capacity. (See table 3.) Serum folate, ferritin, and vitamin B_{12} assessments were performed on those individuals who were found to be anemic and also on a subsample of all examinees. Such an indepth analysis allows a more precise determination of the prevalence of various types of anemia in the United States than was possible using NHANES I variables and allows relationships between measurements of iron status to be characterized (12).

Some aspects of the sampling protocol were revised. In NHANES II, the oversampling of females of childbearing age was eliminated because of their already adequate representation in an unweighted sample. NHANES II included infants ages 6 months to 1 year.

A second modification of the sampling was the use of health-related variables in addition to population density to determine the 16 superstrata in each of the 4 regional areas in the United States from which the 64 primary sampling units would be chosen. Prior surveys had relied on U.S. Bureau of the Census data emphasizing only population density as a criterion variable. Population density continued to be used, but regression analysis was also employed to determine which health-related variables were most predictive of five specific health conditions (infant mortality rates and percent with kidney trouble, heart trouble, hypertension, and high serum cholesterol) in each of the four regional areas. Based on this analysis, median income was chosen as the second major predictor of health problems. The third variable chosen differed for each regional area: Northeast—percent below poverty line; Midwest—rate of growth; South and West—races other than white plus Hispanics (12). These sampling changes did not impair the nutrition monitoring function of NHANES.

The number of Hispanic persons included in NHANES I and II was representative of their numbers in the general U.S. population but was insufficient for separate data analysis. In an effort to sample Hispanic persons in sufficient numbers for surveillance and monitoring purposes, a special addition to NHANES II, the Hispanic Health and Nutrition Examination Survey (Hispanic HANES), was conducted from 1982 to 1984. The mobile caravans sampled Hispanics ages 6 months to 74 years in geographic regions with high densities of persons of Puerto Rican, Mexican, and Cuban extraction or with other Latin American ethnic backgrounds. The protocol for the Hispanic HANES is not identical to NHANES II. Some new measures were included that may be particularly reflective of Hispanic nutrition and health problems. New measures were also added to test their feasibility for inclusion in NHANES III. Table 4 shows the new dietary, anthropometric, and biochemical variables included in the Hispanic HANES (13).

Practical considerations and constraints in variable choice for NHANES

The mobile examination caravan travels from one primary sampling unit to the next with its field staff and administers hundreds of examinations per month. The mobile examination caravan has been adapted to collect a tremendously wide variety of data from the NHANES examinees, but this field examination situation carries its own constraints, which have shaped the protocol for NHANES. Measurements chosen for inclusion in NHANES must meet certain practical criteria. They must be acceptable to the examinees, not overly taxing for the technical staff, and adaptable to the field situation.

In less than 3 hours an immense amount of data is collected from one examinee undergoing the NHANES protocol. To maintain subject compliance, the whole

Table 4. Dietary, anthropological, biochemical and other physical parameters included in the Hispanic Health and Nutrition Examination Survey, 1982–84, that were not included in the second National Health and Nutrition Examination Survey, 1976–80

Category	Age	Description
Dietary intake	6 mo–74 yr	Food frequency adapted for ethnic food use
Anthropometry	6–74 yr	Biacromial breadth Biiliac breadth Medial calf skinfold Iliac crest skinfold Medial calf circumference
	<3 yr	Crown-rump length
	>2 yr	Sitting height
Biochemistry (serum)	4–74 yr	Vitamin A Vitamin E
	20–74 yr	Calcium Phosphorus Sodium Chlorine Potassium Uric acid
	4–11 yr	Tetanus
Other physical parameters	12–74 yr	Hair collection for mineral analysis, including a questionnaire on hair care practice (performed on a subsample)
	20–74 yr	Ultrasound examination of the gall bladder
	6 mo–74 yr	Tympanic impedance

process must flow smoothly and not be overly demanding, embarrassing, or uncomfortable. The numbers of sample persons actually complying with the examination component has decreased from a high of 96% during NHES II to 73% during NHANES II, even with the institution of a $20 remuneration. The examination duration and rigorousness was limited in NHANES II to ensure maximum participation. Consideration was also given to the elimination of procedures that might create participation bias.

A second constraint is the need to train and maintain a highly motivated field staff. Methodologies chosen should not require high levels of technical proficiency. Measures need to be precise, which means readily repeatable among examiners and by the same examiner on multiple tries, without requiring excessive training. Methods must lend themselves to quality control procedures that are not overly burdensome.

Field methods require the use of the minimum of sensitive equipment, which is prone to malfunction. Techniques should be completely validated and pretested for field conditions. Adequate support systems must exist to expedite the biochemical analysis of blood and urine. Measures must be accurate reflectors of the conditions being measured, but the field situation presents many practical considerations where accuracy may be reduced for expediency.

Throughout NHANES, the 24-hour dietary recall method of collecting food consumption data has been used to give a quantitative assessment of nutrient intake. The suitability of the 24-hour recall method for this purpose has been hotly debated (14). Its continued inclusion in NHANES II and the Hispanic HANES serves as an example of the methodological considerations and constraints that dictate the planning of these surveys.

Ideally, dietary data should be collected over at least 3 days (15), but logistically it would be difficult in NHANES to adopt a method of longer than 24-hour duration. Only 1 day's food intake can be recalled during the 3-hour examination period. Hiring and training special dietitians to remain in the area after the examinations to collect dietary information on successive days would substantially increase survey costs. Dietary interviews by telephone are a possibility (16) but would constitute a change in methods and introduce new error. Standardization, and thus quality control, of successive dietary recalls would be difficult. The extra burden on the subjects may decrease participation in the overall examination protocol.

Given the existing constraints on dietary data collection within the context of a mobile health examination caravan, at the present time the 24-hour recall method provides the only possibility for quantitative assessment of nutrient consumption. However, it is feasible to perform a repeat 24-hour recall on a smaller percent of the sample. The use of less than optimal assessment techniques requires that the data analysts be aware of the limitations of field techniques.

In each NHANES, a dietary frequency survey has complemented and reinforced the data collection of the 24-hour dietary recall. This frequency survey registers the number of times (daily, weekly, monthly) foods from 18 food groups have been eaten. The logic of grouping the food affects the usefulness of these data. For instance, provitamin and vitamin A are found in large quantities in only two of the food groups, and calcium predominates in only one. If consumption is low in these groups, vitamin A and calcium intakes are presumably low. The certainty is far greater than that afforded by 24-hour recall data. However, intakes of fat and many of the vitamins are distributed over many food groups so that one cannot identify excessive or low intakes. Identifying high intakes from the food frequency intake data is difficult unless the foods grouped together in a particular food group all contain similar amounts of the nutrient of interest. This method of grouping has not been utilized in NHANES, and in the future more thought should be given to the impact of specific food groupings.

NHANES I and NHANES II: Survey findings on the Nation's nutritional health

One of the surveillance functions of NHANES is to provide information on the national prevalence of important health conditions and nutritional status indicators in the population by sex, age, race, income level, and region. NHANES I was designed so that early release of half the data could be analyzed by those who make public policy decisions that affect the health and nutrition of the community. Because NHANES I was based on what was considered to be a representative sample of the U.S. population, the NHANES I preliminary data were the basis for designing programs for those who were malnourished or at nutritional risk. NHANES I also facilitated improvement of monitoring and surveillance in the United States. During the past 15 years, there has been an expansion of nutrition programs, greater emphasis on surveillance and monitoring, and evidence of action directed at reducing malnutrition.

The NHANES I data revealed that malnutrition in the United States was associated with diseases and socioeconomic factors in subgroups identified at risk. NHANES I results identified these groups as the elderly, preschool children, low-income groups (particularly black persons), and women of childbearing age.

Clinical findings, NHANES I

The overall prevalence of single clinical signs possibly related to nutritional deficiencies or excesses was low (<15%), with the exception of gum signs in persons with chronic gingivitis. The prevalence of many clinical signs showed an increase from ages 1–18 and after age 45 years. There were marked sex differences in prevalences of various signs, as well as higher prevalences in black than white persons. The greater prevalence of most clinical signs in black persons cannot be explained merely on the basis of income level (17).

In NHANES I, only 3% of all persons examined were free of the 48 clinical signs used for assessment (17). This is not surprising because very few clinical signs are specific indicators of malnutrition, particularly when dealing with the milder forms, as seen in industrialized countries (18). The signs and symptoms of nutritional deficiency overlap with many other diseases and may affect various organ systems. Clinical symptoms are nevertheless helpful when used in conjunction with biochemical and anthropometric parameters in making a complete nutritional assessment or diagnosis of malnutrition (19).

Biochemical indicators, NHANES I

The major laboratory finding of NHANES I was the relatively high prevalence of anemia among infants, adolescents, females during the reproductive years, and pregnant women, as defined by standard cutoff points for hemoglobin. This was especially true for the poor and even more so for black persons. The prevalence of low hemoglobin was 19.3% for males and 12.6% for females at age 1 year and approximately 9% for both sexes at 2–3 years of age. After early childhood, the prevalence rarely exceeded 5%, except for males 65–74 years of age, for whom it rose to almost 15%. At most ages black males and females were 2–10 times more likely to have low hemoglobin concentrations than white persons. Males showed a greater prevalence of low hemoglobin values than females in most age groups, with pronounced differences seen in infancy, ages 12–14 years, and ages 55 years and over. Approximately 32% of the pregnant women in the survey had hemoglobins below 12 g/dl, with a prevalence of 28.4% among white women and 52.9% among black women (17). Physiologically low hemoglobin concentrations because of hemodilution in pregnant women have been widely reported. Therefore, lower cutoffs are in common use.

Further analyses of NHANES I data resulted in major reinterpretations of these data. The commonly reported differences in hemoglobin levels between black and white persons are apparently genetically determined. At equivalent transferrin saturation levels, black persons have consistently lower hemoglobin concentrations than white persons. These findings indicate that a racial variation exists for hemoglobin values (20). Thus, traditionally used hemoglobin cutoff values to indicate anemia in white individuals may not be applied to black individuals. Using the example of black and white women, when appropriate statistical methods are used (21), most of the differences in the prevalence of iron-deficiency anemia between the races disappear.

Low serum iron concentration, high total iron-binding capacity (TIBC), and low transferrin saturation were widespread. These findings were considerably more common than low hemoglobin values and equally prevalent among black and white persons. TIBC is a measure of the total amount of transferrin available for iron binding. Transferrin saturation (%) is calculated from the formula:

$$\frac{\text{serum iron } (\mu\text{g/dl})}{\text{TIBC } (\mu\text{g/dl})} \times 100\%$$

A change in transferrin saturation can thus result from a change in either measurement but is a more useful diagnostic indicator of iron deficiency than either value alone. A transferrin saturation less than 16% in adults is considered to be indicative of iron-deficient erythropoiesis.

Groups identified in NHANES I as being at the greatest risk of iron deficiency were children 1–5 years of age, adolescents 12–14 years of age, and women 18–44 years of age. Although iron-deficiency anemia (low hemoglobin concentration with low transferrin saturation) seemed to be a problem among 1-year-old children, only a small proportion of those over 2 years were classified as having iron-deficiency anemia. Despite the prevalence of low transferrin saturation, according to NHANES I data, frank iron-deficiency anemia was not found to be widespread when appropriate statistical methods were used (22).

The finding of widespread low values of transferrin saturation in NHANES I should be treated with caution because low serum iron and high TIBC may be caused by a variety of chronic disorders. Moreover, because of methodological problems, serum iron values were frequently imputed and therefore may not be valid reflectors of iron status. It must be noted that there are several sources of error and variability in the measurement of serum iron and TIBC. These problems were recognized in the planning stages of NHANES II, and more specific indicators of iron status were included in the NHANES II survey protocol.

Biochemical indicators, NHANES II

Results from the selected hematological and nutritional biochemical determinations from NHANES II are presented by race, sex, and age in tables 5–8, with iron-deficiency indexes presented first. The data for children ages 6 months–2 years are not presented because the problems of drawing blood from this age group resulted in a large percentage of missing data. The biochemical results chosen include prevalences of low serum zinc, copper, vitamin A, and vitamin C, and hematological parameters related to iron-deficiency anemia that have been statistically analyzed and published. Reports on zinc, iron, folacin, and vitamin A have been prepared by The Federation of American Societies of Experimental Biology (FASEB) for the Food and Drug Administration (23–26). The current publication of NHANES II serum copper data (27) presents means, standard deviations, and cumulative percent distributions but does not contain standards for normal values. Therefore, the cutoff value for suboptimal serum copper levels used in table 8 has been selected from the literature (28). It is difficult to make direct comparisons of NHANES I and II results because a number of laboratory tests, procedures, and quality control measures changed between the two surveys.

Hematological variables and iron nutriture

NHANES II afforded a more indepth assessment of the relationship between iron nutriture and hematological-biochemical variables. NHANES II biochemical analyses new to the survey series included erythrocyte protoporphyrin, a measure of the availability of iron for heme synthesis; serum ferritin, an index of iron stores; and mean corpuscular volume (MCV), a direct measure of microcytic anemia. Serum ferritin values were obtained for 5,157 persons aged 3–74 years. Despite its smaller size, the ferritin sample yields national probability estimates similar to those of the larger NHANES II population.

The NHANES II data on iron status were analyzed by a group of experts and approved by the Life Sciences Research Office Advisory Committee under the authority of FASEB (24). Two models were developed to estimate the prevalence of impaired iron status. The ferritin model included ferritin, transferrin saturation, and erythrocyte protoporphyrin as variables. The MCV model included MCV, transferrin saturation, and erythrocyte protoporphyrin. Thus, the second model contained a measure of frank anemia assumed to be specific for iron deficiency. The ferritin model is more sensitive because it can identify mild depletion of iron stores. Each model required that at least two of the three iron-status variables be abnormal. A hemoglobin percentile shift model was also defined but is not reported here.

Table 5. Cutoffs used for prevalence analyses in the second National Health and Nutrition Examination Survey

Age	Serum ferritin (ng/ml)	Transferrin saturation (percent)	Erythrocyte protoporphyrin (ug/dl RBC)	Mean corpuscular volume (fl)
1–2 years	—	<12	>80	<73
3–4 years	<10	<14	>75	<75
5–10 years	<10	<15	>70	<76
11–14 years	<10	<16	>70	<78
15–74 years	<12	<16	>70	<80

Source: Pilch, S.M., and F.R. Senti, editors: Assessment of the iron nutritional status of the U.S. population based on data collected in the second National Health and Nutrition Examination Survey, 1976–1980. Prepared for the Center for Food Safety and Applied Nutrition, Food and Drug Administration, under Contract No. FDA 223–83–2384 by the Life Sciences Research Office, Federation of American Societies for Experimental Biology, Bethesda, MD. Available from: FASEB Special Publications Office, Bethesda, MD. 1984.

The rationale for choosing a model with three iron-status indicators was to provide a better measure of iron status and to increase the specificity for diagnosing iron deficiency. Taken alone, none of the measures is diagnostic for iron deficiency. Transferrin saturation is highly labile, responding to diurnal variations in serum iron and generally declining in response to infection or inflammation. Although erythrocyte protoporphyrin determination does not distinguish between iron deficiency and infection and is also elevated in lead poisoning, it provides a more stable index of iron-deficient erythropoiesis than transferrin saturation. Serum ferritin values tend to increase in response to infection, malignancy, and liver disease, but serum ferritin is the most sensitive indicator of storage iron depletion. The occurrence of two or three abnormal values for iron-status indicators in an individual cannot be used solely as the basis for a definitive diagnosis of iron deficiency but is an improvement on past attempts to determine the existence of low iron stores in otherwise asymptomatic individuals.

Table 5 presents the cutoffs used for the analysis of iron nutrition status by FASEB. Precise definitions of normal and abnormal values for iron-status indicators in infancy, childhood, and adolescence are still evolving. The expert committee based their cutoffs on statistical analyses, specificity considerations, and cutoff values in current use.

Abbreviated results of the analysis of the prevalence of two or three abnormal values using the ferritin model and the MCV model for black and white subjects aged

3–74 years are shown in table 6. Data are presented for males and females together through age 10 years because there were no significant differences in values for iron-status indicators by sex up to this age. For most age-sex groups, the prevalences of abnormal values obtained using the ferritin model were slightly higher than those using the MCV model. The greatest differences were seen in males aged 11–14 years (not shown), females aged 15–19 years, and females aged 20–44 years. The ferritin model yielded slightly lower prevalence estimates than the MCV model in the elderly, possibly reflecting elevated serum ferritin levels caused by inflammatory disease in older groups.

The association of low hemoglobin concentrations with abnormal values for iron-status indicators was also examined by the FASEB committee. The prevalence of low hemoglobin values increased with an increasing number of abnormal values for iron-status indicators. This finding supports the usefulness of the ferritin and MCV models for identifying populations with impaired iron status.

The analyses of NHANES II data by FASEB suggest that the prevalence of impaired iron status was fairly low for most age-sex groups. Black persons tended to have slightly higher prevalences of abnormal values than did white persons, with statistically significant differences for children aged 3–4 years (using the MCV model), females aged 15–19 years (using both models), and persons of both sexes aged 45–74 years (using both models). Several groups in the population exhibited relatively high prevalences of inadequate iron status and deserve further consideration: Children aged 1–2 years, males aged 11–14 years, and females aged 15–44 years.

The ranges of prevalence estimates for impaired iron status tend to be lower than those from some other recent surveys of North American populations, including NHANES I. One interpretation is that the change is methodological in origin. Comparison of absolute values is almost impossible. Different parameters were used, the choice of cutoff values changed, and the survey and laboratory analytical methodologies were modified. Another interpretation is that iron status has actually improved in the United States. The authors speculated that the range of prevalence for impaired iron status is lower because of the availability of high-quality food through feeding programs, availability of the iron used in fortification, use of iron supplements, use of ascorbic acid supplements or dietary ascorbic acid, and use of oral contraceptive agents, which decrease menstrual blood loss. In spite of differences in numbers deficient, relative trends of impaired iron status among various groups in NHANES II remained similar to those from earlier surveys.

The extensive NHANES II data on iron nutriture allowed an opportunity to generate new reference standards for hemoglobin concentrations and to reassess the

Table 6. Percent of persons with 2 or 3 abnormal values for iron status indicators using the MCV and ferritin models, by race: Second National Health and Nutrition Examination Survey, 1976–80

		MCV model [1]			Ferritin model [1]		
Sex	Age	White	Black	P-value [2]	White	Black	P-value [2]
		Percent			Percent		
Both sexes	1–2 yr	8.4	10.9	NS			
Both sexes	3–4 yr	3.2	8.5	<0.05			
Both sexes	5–10 yr	3.8	4.9	NS			
Male	11–14 yr	3.5	2.8	NS			
Female	11–14 yr	3.0	6.7	NS			
Male	15–19 yr	0.8	1.6	NS	0.0	0.9	
Female	[3]15–19 yr	3.8	12.6	<0.10	11.7	31.3	<0.05
Male	20–44 yr	0.8	1.1	NS	0.6	0.4	
Female	[3]20–44 yr	5.0	5.7	NS	9.2	11.9	
Both sexes	45–74 yr	2.9	5.9	<0.05	2.7	4.9	<0.10

[1] The MCV model employs mean corpuscular volume, transferrin saturation, and erythrocyte protoporphyrin as variables. The ferritin model employs serum ferritin, transferrin saturation, and erythrocyte protoporphyrin as variables.

[2] Based on 2-tailed t-test.

[3] Pregnant women are excluded.

Notes: All statistics are weighted to represent the U.S. population at the midpoint of the survey (March 1, 1978) by a method that accounts for the complex survey design.

NS = not significant, $p \geq 0.1$.

Source: Adapted from: Pilch, S.M. and F.R. Senti, editors: Assessment of the iron nutritional status of the U.S. population based on data collected in the second National Health and Nutrition Examination Survey, 1976–1980. Prepared for the Center for Food Safety and Applied Nutrition, Food and Drug Administration, under Contract No. FDA 223-83-2384 by the Life Sciences Research Office, Federation of American Societies for Experimental Biology, Bethesda, MD. Available from: FASEB Special Publications Office, Bethesda, MD. 1984.

Table 7. Prevalence of anemia in persons of all races and white persons: Percent below the 95th percentile reference ranges,[1] by age and sex in the second Health and Nutrition Examination Survey, 1976–80

Sex and race	1–2 years	3–5 years	6–8 years	9–11 years	12–14 years	15–17 years	18–24 years	25–44 years	45–64 years	65–74 years
Both sexes										
All races	5.7	3.5	2.3	2.8						
White		2.3	1.5	2.5						
Male										
All races					2.9	2.6	2.7	2.9	3.8	4.4
White					2.2	2.2	3.0	2.9	3.4	4.5
Female										
All races					3.6	5.9	3.3	5.8	3.9	3.9
White					3.7	4.3	3.0	4.6	3.7	3.5

[1] Ranges in hemoglobin concentrations by age, sex, and race were derived after excluding pregnant women, subjects with hemoglobinopathies, and individuals with transferrin saturation <16%, MCV <80 fl, or erythrocyte protoporphyrin >75 µg/dl red blood cells. The 95th percentiles of these ranges are used as the cutoffs for normal hemoglobin for each age, sex, and race group.

Source: Dallman, P.R., R. Yip, and C. Johnson. Prevalence and causes of anemia in the United States, 1976 to 1980. *Am. J. Clin. Nutr.* 39:437–445, 1984. Copyright © Am. J. Clin. Nutr.

prevalence of anemia in the United States based on these new hemoglobin reference standards. This work was undertaken by Dallman, Yip, and Johnson (29).

To develop the reference standards for hemoglobin, all persons with hemoglobinopathies and abnormal values of transferrin saturation, MCV, or erythrocyte protoporphyrin were eliminated from the sample. Criteria for elimination tended to be stringent, resulting in the exclusion of some healthy individuals. Also, pregnant women were removed from the sample. The 95% range around the mean was calculated in each age, sex, and race group in which the sample size exceeded 200. The lower hemoglobin value in the 95% range was considered the cutoff point for anemia.

The hemoglobin reference standards that resulted from this analysis of NHANES II data correspond well to the results of large sample surveys in Western countries. Hemoglobin concentrations peaked in adults in their early twenties. Thereafter, hemoglobin concentrations gradually declined, and the range of values broadened among elderly males. In adult women, hemoglobin concentrations rose slightly after age 40.

Prevalences of anemia based on the hemoglobin reference standards are shown in table 7. Prevalences ranged from 1.5% in white children aged 6–8 years to a high of 5.8% in females of all races aged 25–44 years. Anemia prevalences were considerably lower than those obtained in NHANES I using standard cutoff values for hemoglobin concentrations. Because of the small sample sizes necessitated by the use of ferritin as a variable, data for black persons could not be disaggregated. There was a tendency for the prevalences to be higher in the "all races" groups than in the white groups, especially in children ages 1–11, but this trend was not consistent in teenagers or adults of either sex.

Thus, the consistently higher incidences of anemia among black persons originally reported in NHANES I were not borne out by NHANES II data, but the two data sets are congruent when analyzed by the same methods. One unexplained finding, probably resulting from differences in methodology, is a systematically lower level of hemoglobin in NHANES II than NHANES I. There is no evidence that this is related to iron nutrition status.

Other biochemical variables

Table 8 shows the prevalences of low laboratory values for serum copper, zinc, vitamin C, and vitamin A in NHANES II (26, 27, 30). The prevalence of low serum zinc values was greater in white females than white males 15–64 years of age and greater in black females than black males beginning with age 6 years. Mean serum zinc values were higher in males than in females for ages 12–74 years (30). The major differences in serum zinc levels seem to be between the sexes, particularly in white persons.

On the other hand, the prevalence of low serum copper levels was greater in white males than white females for each age category (whereas conclusions could not be drawn for black persons because of small sample sizes). For each sex, black persons had higher observed serum copper levels than white persons. Thus, mean serum copper levels were higher in black than in white persons and were higher in females than in males (27).

The reported high prevalence of low serum zinc values must be carefully interpreted. Circulating zinc levels in plasma and serum do not necessarily provide an accurate reflection of body zinc stores and therefore can only be suggestive of zinc nutritional status. Circulating zinc levels closely correlate with albumin and thus may reflect depressed plasma protein binding of zinc as well as body zinc deficiency. In addition to nutritional factors, many diseases and medical treatments may produce zinc deficiency. Measures of plasma or serum zinc involve the determination of zinc levels in a body fluid and do not indicate zinc-dependent metabolic functional status. Functional tests that include zinc-dependent en-

Table 8. Percent of population below standard values for serum copper, zinc, vitamin C, and vitamin A, by race, age, and sex: Second National Health and Nutrition Examination Survey, 1976–80

	Serum copper [1, 2]			Serum zinc [3]		Serum vitamin C [3]		Serum vitamin A [3–5]	
Race and age	Male	Female	Age	Male	Female	Male	Female	Male	Female
White									
3–5 yr	—	0.2	3–5 yr	3.1	3.8	0.1	—	2.8	6.1
6–11 yr	0.6	0.1	6–11 yr	1.9	1.2	0.4	—	2.0	1.6
12–14 yr	0.8	0.8	12–17 yr	1.0	0.8	1.6	1.7		
15–17 yr	1.6	0.3							
18–24 yr	1.9	0.5	18–24 yr	0.2	3.6	2.1	1.9		
25–44 yr	0.9	0.4	25–54 yr	1.0	3.0	5.4	2.9		
45–64 yr	0.9	0.2							
65–74 yr	0.4	0.1	55–74 yr	1.9	2.4	6.5	1.8		
Black									
3–5 yr	—	—	3–5 yr	2.7	1.0	—	—	5.9	3.0
6–11 yr	—	1.4	6–11 yr	3.0	3.9	—	1.0	2.9	1.6
12–14 yr	—	—	12–17 yr	1.3	3.2	1.4	0.6		
15–17 yr	1.3	—							
18–24 yr	—	1.1	18–24 yr	2.1	5.9	3.0	2.5		
25–44 yr	—	—	25–54 yr	1.4	5.5	9.6	4.7		
45–64 yr	—	—							
65–74 yr	—	—	55–74 yr	3.3	3.8	16.2	5.3		

[1] Source: National Center for Health Statistics, R. Fulwood, C.L. Johnson, J.D. Bryner, E.W. Gunter, C.R. McGrath: Hematological and nutritional biochemistry reference data for persons 6 months–74 years of age: United States, 1976–80. *Vital and Health Statistics.* Series 11–No. 232. DHHS Pub. No. (PHS) 83–1682. Public Health Service. Washington. U.S. Government Printing Office, Dec. 1982.

[2] Cutoff value of <70 µg/dl from Sauberlich, H.E. Current laboratory tests for assessing nutritional status. *Surv. Synth. Path. Res.* 2:120–133, 1983.

[3] Source: U.S. Department of Health and Human Services and U.S. Department of Agriculture: *Nutrition Monitoring in the United States—A Report from the Joint Nutrition Monitoring Evaluation Committee.* DHHS Publication No. (PHS) 86–1255. Public Health Service. Washington. U.S. Government Printing Office, July 1986, pp. 102, 140, 180, 309, 311, 315. Cutoff values used:

Zinc: <70 µg/dl—a.m., fasting; <65 µg/dl—a.m., nonfasting; <60 µg/dl—p.m.
Vitamin C: <0.25 µg/dl.
Vitamin A: <0.20 µg/dl.

[4] Vitamin A status was assessed only in children 3–11 years of age.

[5] Source: Pilch, S.M., editor: Assessment of the vitamin A nutritional status of the U.S. population based on data collected in the National Health and Nutrition Examination Surveys. Prepared for the Center for Food Safety and Applied Nutrition, Food and Drug Administration, under Contract No. FDA 223–84–2059 by the Life Sciences Research Office, Federation of American Societies for Experimental Biology, Bethesda, MD, 1985, p. 31.

Note: — = insufficient sample size for analysis.

zyme activities would be more sensitive and specific to zinc nutritional status (31).

Low serum copper levels did not seem to be a problem, according to NHANES II data, although fewer studies on which to base cutoff values are available for copper than for zinc. Normal ranges for circulating copper are very similar to the ranges for zinc, although there is indication that they may be age specific (25). Decreased copper levels are found in patients with Wilson's disease (a copper accumulation disorder), nephrosis, kwashiorkor, and cystic fibrosis. Serum copper levels must also be interpreted with caution because elevated levels accompany most infections, leukemia, Hodgkin's disease, various anemias, hyperthyroidism, hemochromatosis, collagen disorders, and myocardial infarction (31). Perhaps serum copper values should be examined in terms not only of percent of population below but also percent above the standard.

Mean levels of serum vitamin C were within the acceptable range for males and females classified by age, race, or poverty status, although they differed by age and sex. Values of less than 0.30 mg/dl indicate low or inadequate intakes with inadequate reserves, although serum vitamin C is sensitive to recent intake. The percent of males below the standard tended to increase with age, with the percent of black males 65–74 years being notably highest. Males had lower mean levels of serum vitamin C than females for all the adult age groups (30).

Other factors, such as cigarette smoking and vitamin supplementation, can affect levels of serum vitamin C. Adults in NHANES II who smoked cigarettes and seldom or never used vitamin-mineral supplements had a much higher percent of low serum vitamin C values than would be expected. Although the overall proportion of persons aged 3–74 years with low serum vitamin C was only about 3%, some adult subpopulations considered to be at high risk for poor vitamin C nutritional status were identified by NHANES II data analysis. These include consumers of diets that were low in vitamin C, smokers of cigarettes, irregular consumers or nonconsumers of vitamin-mineral supplements, and the poor (30).

Vitamin A status was assessed only in children 3–11 years of age because, in NHANES I, low levels of vitamin A had been found in less than 1% of all persons 12–74 years of age. The Expert Panel on Vitamin A Nutriture stated that serum vitamin A levels alone are inadequate to provide estimates of the prevalence of vitamin A deficiency and toxicity in the U.S. population (26). Table 8 shows that about 5% of children 3–5 years of age had low levels of serum vitamin A, and this decreased to about 2.0% in the age group 6–11 years. There do not seem to be consistent patterns with regard to race in the prevalence of low values. Mean serum vitamin A levels were within normal ranges regardless of race, sex, or poverty status (26).

Because the overall prevalence of low serum albumin in NHANES II was less than 0.1% and did not differ significantly by race or poverty status, the data are not presented. These findings are consistent with the dietary data and indicate that protein nutriture is not a public health problem (30).

Although cutoff values have been used in this discussion of serum nutrient status, we recognize that they merely represent guidelines for expressing the biochemical levels of specific parameters rather than quantifying deficiency status. Such cutoff points lack sensitivity and, above all, specificity when used as diagnostic criteria for assessment of nutritional or health status.

Functional indicators of nutritional status have potential benefits over static indicators and are emerging as a new class of diagnostic tools. Five areas of functional competence likely to be affected by malnutrition have been identified by a group under the auspices of the National Academy of Sciences (32). They include cognitive ability, disease response, reproductive competence, physical activity and work performance, and social-behavioral performance. Solomons and Allen amplified this concept and suggested that functional performance can be used as an index of nutritional status (33). However, more nutrient-dependent physiological functions need to be explored and diagnostic procedures developed before functional indicators can be of widespread use or applicable to field surveys such as NHANES. The sensitivity-specificity relationship (34) of those measures is probably even worse than that of more traditional indicators of nutritional status.

Dietary intake in NHANES: Findings and problems with data interpretation

NHANES I dietary data indicate that iron, calcium, and vitamin A consumption were substandard. Mean iron intake was about 50% of the Recommended Dietary Allowances, or RDA's (35), in children ages 1–4 years and in women during the reproductive period, 12–54 years of age (36). The range of mean calcium intake was 65–80% of the 600-mg standard for black women and 95–114% of that standard in white women (20–74 years of age). The median vitamin A intake was 46–75% of the 3,500 IU standard in black women in most age groups and in black males aged 12–34 years (17, 33). NHANES II dietary data have not yet been analyzed in terms of substandard intakes.

Three major problems arise in the interpretation and utilization of nutrient intake data. The first problem is the variability and error inherent in the data. The second is the correct application of standards in assessing the adequacy of quantified dietary intakes. The third lies in the appropriate uses of dietary intake variables.

Inherent variability and error

Beaton (37) divides the variability observed in group dietary intakes into two categories: True interindividual variation resulting from actual subject differences in usual intake and intraindividual variation, which includes any methodological error plus true variability in day-to-day food consumption.

Methodological errors in 24-hour recall data include error in subject recall of actual food consumed and error in the quantification of nutrient composition from dietary data. The subject may have difficulty remembering exactly what and estimating how much was eaten. The 24-hour recall may result in overreporting intakes below the mean and underreporting intakes above the mean, the "flat-slope syndrome" (38). Nutrient calculations from food data lack accuracy because of the inherent variability in the nutrient content of foods caused by cultivar differences, food handling, and processing.

Another level of error in the 24-hour recall data is the inability to measure day-to-day variability in dietary intake by this method. The intake of certain nutrients such as vitamin A varies greatly from day to day and may require many days for an accurate assessment of habitual consumption. Short-term intake data may tend to overestimate the prevalence of habitually low nutrient intakes (39).

Application of standards for assessing the adequacy of intakes

Beaton has recently demonstrated the folly of using the RDA's as standards for assessing the nutrient intakes of populations without knowledge of the nature of the distribution of requirements for a specific nutrient. Because the RDA is usually set two standard deviations above the mean requirement for a nutrient, individuals at any particular intake level can be assigned only a probability of risk for inadequate intake. Thus, RDA's do apply to individuals, but definitive statements cannot be made about the actual adequacy or inadequacy of individual nutrient intake. An estimate of the number of persons in a population group who have inadequate intakes may be calculated with knowledge of the probability of inadequacy at each level of intake and the actual number of persons in the sample at each level of intake. Such an analysis requires that the nature of the distribution of requirements for each nutrient used in generating the RDA's be published along with the other RDA information (40).

Beaton makes a second important point in his analysis of correct RDA usage. The RDA can be applied only to group data that are estimates of the usual intakes of

Table 9. Regression analysis of biochemical values on nutrient intake and sociodemographic variables: First National Health and Nutrition Examination Survey, 1971–74

	Dependent variable (biochemical value)	Sample size	Independent variable (nutrient intake)	Unadjusted		Adjusted[1]	
				Slope	R^2	Slope	R^2
1.	Serum protein	18,159	Protein intake	**0.0037	0.001	0.0010	0.080
2.	Serum albumin	18,159	Protein intake	**0.0124	0.011	0.0023	0.097
3.	Serum vitamin A	18,159	Vitamin A intake	**0.0322	0.004	0.0264	0.265
4.	Urinary thiamin/creatinine ratio	16,533	Thiamin intake	*1.9702	0.002	**2.4585	0.031
4a.	Log transformed urinary thiamin/creatinine ratio	16,533	Thiamin intake	**0.0024	0.012	**0.0026	0.092
5.	Urinary riboflavin/creatinine ratio	16,677	Riboflavin intake	**1.2703	0.003	**1.5152	0.023
5a.	Log transformed urinary riboflavin/creatinine ratio	16,677	Riboflavin intake	*0.0019	0.025	**0.0019	0.092
6.	Hb	17,238	Iron intake	** −0.0151	0.010	−0.0009	0.098
7.	Hematocrit	17,238	Iron intake	** −0.0195	0.023	−0.0011	0.115
8.	Serum iron	17,200	Iron intake	** −0.2968	0.044	** −0.0699	0.156
9.	Transferrin saturation	17,200	Iron intake	** −0.0846	0.006	** −0.0385	0.096

[1] Adjusted for race, sex, age, income above or below the poverty level, and urban or rural residence.

Notes: * $p < 0.05$. (Refers to slope values significantly different from zero.)
** $p < 0.01$.

Source: Kerr et al.: Relationships between dietary and biochemical measures of nutritional status in HANES I data. *Am. J. Clin. Nutr.* 35:303, 1982. Copyright © Am. J. Clin. Nutr.

individuals. The RDA's are not standards for daily requirements but rather are standards for the average of usual intakes over time. Because single-day assessments overestimate the number of habitually low intakes in a population, use of the RDA as a standard for the 24-hour recall data in NHANES grossly overestimates the number of persons truly at risk of deficiency (40).

Relationships between dietary and biochemical measures in NHANES I

Cross-sectional nutrition survey data should not be used to indicate cause-and-effect relationships. However, a link between dietary intake of a particular nutrient and its respective biochemical measure may be found. To assess the predictive value of nutrition survey data, Kerr et al. (41) examined the relationships between dietary and biochemical indexes (of individual subjects) for several nutrients evaluated by NHANES I.

Kerr et al. considered those nutrients for which both dietary and biochemical values were available from NHANES I data tapes. These included the intake of protein, vitamin A, thiamin, riboflavin, and iron; serum protein, albumin, and vitamin A; urinary thiamin-creatinine and riboflavin-creatinine excretion ratios; hemoglobin, hematocrit, percentage transferrin saturation, and serum iron.

All dietary and biochemical data were expressed as a percentage of the appropriate values for age and sex on smoothed curves derived from commonly used reference standards. The reference standards for 24-hour dietary recall data were derived from the 1980 RDA and those for biochemical and hematological data were adapted from minimal "acceptable" standards previously used in national nutrition surveys (36).

The results from the Kerr et al. study are presented in table 9. Relationships between biochemical and dietary values were examined by regression analysis. First, the effect of the nutrient intake (independent variable) on the biochemical value (dependent variable) was examined under the heading "unadjusted." In a second analysis, five additional independent sociodemographic variables (race, sex, age, income below or above the poverty level, and urban or rural residence) were added to the regression equation. These results are under the heading "adjusted." The R^2 values indicate that nutrient intake as assessed by a 24-hour recall is a poor predictor of an individual's biochemical values. The inclusion of the five sociodemographic factors into the equation improved the R^2 values considerably. These results imply that the sociodemographic profile is a better predictor of biochemical status than a 1-day reported dietary intake.

These results demonstrate the difficulty of using 24-hour recall data to predict biochemical parameters. The correct assumption may not be that no relationship exists within the population, but rather that the methods used to assess dietary intake are not sufficiently reliable to truly estimate intraindividual variability.

Given a sufficiently large sample size, good estimates of group mean nutrient intakes can be achieved using 24-hour recall nutrient data. However, a large group size does not eliminate the problems of using data with a high intraindividual variance to perform correlational or regression analysis. The largest source of intraindividual variation is the day-to-day variability in nutrient intake (14). An increase in the number of days measured is necessary to increase the reliability of the estimate of usual intake. Replicate examination of 24-hour recalls in a subsample of the population would at least provide the information to calculate the magnitude of intraindividual covariance (34).

Analyses using data from the dietary frequency interview concerning food groups specific for certain food components have revealed interesting associations (42), and development of this method is worth pursuing.

Table 10. Comparison of the percent of men and women aged 20–74 years in the first National Health Examination Survey (NHES I) (1960–62) and first National Health and Nutrition Examination Survey (NHANES I) (1971–74) 10 and 20 percent or more above desirable weight, by sex and age: United States

	Percent deviating from desirable weight[1]—			
	10 percent or more		20 percent or more	
Sex and age	1960–62	1971–74	1960–62	1971–74
Men				
20–25 years	22.2	18.5	9.6	7.4
25–34 years	28.7	30.3	13.3	13.6
35–44 years	31.8	39.1	14.9	17.0
45–54 years	36.9	35.7	16.7	15.8
55–64 years	36.4	34.0	15.8	15.1
65–74 years	33.5	32.5	14.6	13.4
20–74 years	32.1	32.1	14.5	14.0
Women				
20–25 years	18.8	19.4	9.1	9.6
25–34 years	24.3	25.2	14.8	17.1
35–44 years	34.6	36.6	23.2	24.3
45–54 years	43.4	42.9	28.9	27.8
55–64 years	56.2	50.2	38.6	34.7
65–74 years	56.2	49.0	38.8	31.5
20–74 years	38.1	36.4	25.1	23.8

[1] Desirable weight estimated from regression equations of weight on height for men and women aged 20–29 years, obtained from NHANES I and used as the base for the findings in NHES I.

Source: Adapted from: National Center for Health Statistics, S. Abraham, C.L. Johnson, and M.F. Najjar: Weight and height of adults 18–74 years of age, United States, 1971–1974. *Vital and Health Statistics*, Series 11–No. 211, DHEW Pub. No. (PHS) 79–1659. Public Health Service. Washington. U.S. Government Printing Office, May 1979.

Table 11. Percent of persons in first National Health and Nutrition Examination Survey (1971–74) who would be considered obese at selected percentiles based on triceps plus subscapular skinfolds, by age and sex

	Percent obesity[1] at selected percentiles (cutoff for skinfold sums at each percentile, male/female ages 20–29)		
Sex and age	85th percentile or more (38/52mm)	90th percentile or more (43/58.5mm)	95th percentile or more (51.1/68mm)
Men			
20–24 years	12.7	9.1	4.5
25–34 years	20.8	13.6	6.6
35–44 years	21.2	11.1	3.8
45–54 years	20.8	14.1	6.3
55–64 years	18.5	12.0	4.2
65–74 years	16.6	10.0	3.9
20–74 years	19.4	12.0	5.1
Women			
20–24 years	13.2	8.1	3.6
25–34 years	21.0	14.4	8.0
35–44 years	30.7	19.9	10.3
45–54 years	37.1	23.8	11.0
55–64 years	36.4	23.2	10.6
65–74 years	26.1	23.2	6.3
20–74 years	27.7	17.8	8.5

[1] Obesity measure is based on the sum of triceps and subscapular skinfolds and is defined as the sex-specific 85th, 90th, and 95th percentile for persons aged 20–29 years.

Source: Adapted from: National Center for Health Statistics, S. Abraham, C.L. Johnson, and M.F. Najjar: Weight and height of adults 18–74 years of age, United States, 1971–1974. *Vital and Health Statistics.* Series 11–No. 211. DHEW Pub. No. (PHS) 79–1659. Public Health Service. Washington. U.S. Government Printing Office, May 1979.

NHES and NHANES: Body composition and obesity

Prevalence of obesity and overweight in NHANES

The ranges of weight, height, and body fatness found in NHES and NHANES I have been extensively analyzed (43–50) and compared (51–53). Tables 10 and 11 and figure 1 present some of the important comparisons that have been published in the literature. These examples are chosen for the extent to which they condense information relevant to the following discussion.

Analyses comparing NHES I data with NHANES I data show a distinct trend toward taller and heavier bodies and larger arm muscle areas in NHANES I (52, 53). These increases were evident across all age and sex groups except women ages 55–64, who were actually 1 pound lighter in 1971–74 despite a 4-cm increase in stature. Subscapular skinfolds increased in all but women ages 55–64 years; triceps skinfolds decreased in men ages 25–55 (figure 1).

These data indicate that Americans have been enjoying a continually favorable environment for growth over the period from 1960 to 1974. The authors have also interpreted the increase in body size to imply that the samples from NHES I and NHANES I are not anthropometrically comparable and therefore are unsuitable for monitoring changes in obesity in the U.S. population over time. By the same line of reasoning, these authors propose that the old standards for obesity not be applied to the new surveys.

It is important to know that height has increased before interpreting the meaning of increased body weights in NHANES I relative to NHES I. However, a change in body dimensions does not alter the validity of using relative measures of obesity such as the body mass index (BMI) or skinfold thickness to predict body fat. If this were otherwise, relative measures could not be used to compare obesity within samples for which body dimensions varied widely. An increase in arm circumference and arm muscle circumference without a change in skinfold thickness implies an increase of absolute body fat as well as an increase in bone diameter and muscle thickness. The amount of fat as a percent of total body weight remains constant. The prediction of percent

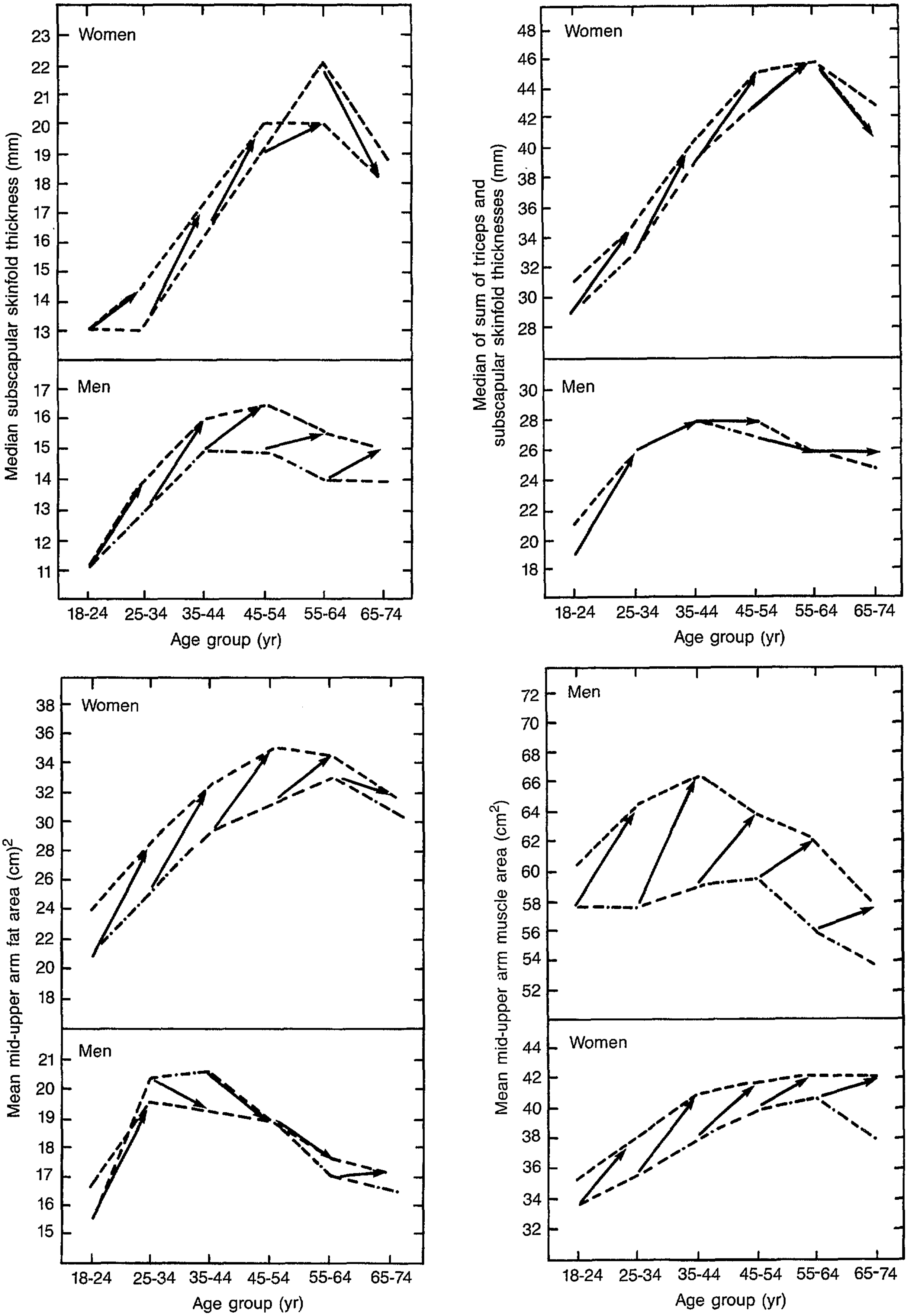


Figure 1. Comparison of median subscapular skinfold thickness, sum of triceps and subscapular skinfold thicknesses, median upper-arm muscle area from first National Health Examination Survey (dot-dashed lines) and first National Health and Nutrition Examination Survey (dashed lines). Arrows indicate change in the median for each age group as it aged 10 years. Source: Bowen, P.E., and P.B. Custer. Reference values and age related trends for arm muscle area, arm fat area, and sum of skinfolds for United States adults. *J. Amer. Col. Nutr.* 3:357–376, 1984. Copyright; used by permission.

body fat by skinfold thicknesses is improved only slightly and inconsistently by inclusions of height, weight, or arm circumferences as predictor variables (54).

A second interpretation of the trend toward larger body dimensions and skinfold thicknesses from NHES I to NHANES I is that the differences are in part methodological. Weight and height are relatively easy measures to replicate, but methods for obtaining skinfolds must be highly standardized. Without constant training and retraining of the technical staff, skinfold measurements may vary among staff members. Expert trainers change between surveys and may vary in their technique. Springs in skinfold calipers wear out with constant use and thus must be frequently calibrated. The skinfold measurements in NHANES II are slightly higher than those in NHANES I without a compensatory increase in weight (55), indicating a slight change in technique between surveys or equipment malfunction.

It is clear from the analysis of NHANES I and the comparison of NHANES I with NHES I that greater than ideal weights and larger than desirable skinfold measurements exist in a substantial portion of the adult population, particularly in middle-aged females and males and especially among the poor. Obesity, as measured by overweight and overfatness, is the most prevalent form of malnutrition in the United States (tables 10 and 11).

The NHES and NHANES data sets support the well-recognized fact that obesity is a major public health problem. A particular strength of such data sets should be their ability to provide us with standards for desirable levels of skinfolds or weight standards based on the minimization of disease risk and other negative functional outcomes associated with overfatness. Such standards would allow us to quantify the degree of functional obesity, as opposed to cosmetic obesity. Such standards should have a high level of specificity and sensitivity to aid in the survey's functions of surveillance, screening, and causal determination.

NHANES I data have provided the opportunity to compare measures of body fatness with blood pressure (42, 56, 57), serum urate and serum cholesterol (58), and prevalence of diabetes (59). These analyses demonstrate fairly linear relationships between indicators of health risk on the one hand and increasing body weight, BMI as measured by weight over height squared, and skinfold thickness on the other. Individuals in the lowest 20th percentile of weight, BMI, or skinfold thickness have consistently lower indicators of diabetes and cardiovascular disease risk than those with higher percentiles. Thus, any increment in body fatness can be construed as risk. NHANES II data will allow further assessment of the association between obesity and indicators of health risk or actual disease states such as osteoporosis and hyperglycemia.

Based on the existing data from NHANES I, standards for obesity derived from negative functional outcomes would need to be set far below the mean weights or skinfolds of the U.S. population. Even a small amount of excess fat presents a health risk. There is little medical rationale at this time for stating that 20% overweight should be the cutoff for weight loss therapy intervention with otherwise healthy adults, as did a recent National Institutes of Health consensus panel (60). The disease risk is less in all but excessively lean individuals, i.e., triceps skinfold <5 mm or relative weight <75% of desirable weight (59, 60). The 20% overweight cutoff rationale is essentially methodological. Overweight is not specific for obesity at weights less than 20% above desirable weight and may result from lean body mass. Therefore, the cutoff of 20% above desirable weight is more a function of the crudeness of our measurement of obesity than a reflection of the cutoff for risk. NHANES III should provide alternatives to relative weight estimates of obesity.

A second asset of a large surveillance survey, such as NHANES, in which multiple indexes of fatness are measured is the opportunity to determine which parameters of obesity are most specifically correlated with disease outcomes. Delineating the best criteria for obesity is particularly important for diagnostic and screening purposes.

In the NHANES analyses published to date, the BMI has been the single highest predictor of serum cholesterol, uric acid, and blood pressure (42, 58). Weight, triceps or subscapular skinfolds alone, and the sum of triceps and subscapular skinfolds are significantly, but not as highly, related to disease risk. The relationship between skinfolds and blood pressure tends to be nonlinear at the extremes of skinfold thickness. The linearity of the relationship is not improved when skinfolds are logged (57, 61).

Because the BMI is based on simple and readily accessible measurements, the BMI is gaining in popularity as an obesity indicator. Its linear correlation with disease outcomes is another strong point. However, the meaning of the BMI is difficult to interpret. The BMI explains, on the average, only 50% of the variation in skinfolds (61). Therefore, it is not synonymous with adiposity. The BMI explains less of the variation in body density than skinfolds (61). The optimal BMI is not correlated to height, but the power to which height must be taken to achieve the optimal BMI differs with age and sex (50). The commonly used weight/height^2 is optimal only for middle-aged males; thus, the meaning of the BMI is not the same across age and sex groups unless the power to which height is taken varies optimally with the group. Part of the association of the BMI with the negative functional outcomes of obesity is its association with aging, a stronger association than for skinfolds (57).

Skinfold thicknesses are less precise measures than height and weight but are more direct assessments of actual adiposity than the BMI. In the case of hypertension in persons aged 30–59 years, the relationship between skinfolds and blood pressure is not age dependent (57). Furthermore, all skinfolds are not equal in their ability to predict the aspect of obesity related to disease. By summing skinfolds, important predictive information may be lost.

Stepwise regression analysis of NHANES I data shows that triceps skinfold has no ability to predict either systolic or diastolic blood pressure after removing the variance in triceps explained by subscapular skinfold. However, subscapular skinfold has significant predictive power beyond that explained by triceps (57). This analysis is supported by other research indicating that fat deposited on the trunk is the best predictor of blood pressure, diabetes, and elevated lipids (62–65). In some cases, triceps skinfold may even be a negative predictor variable (63). Subscapular skinfold, although relatively reliable and easily obtained, is probably not the skinfold of choice for predicting centralized obesity. A better estimate would be less highly correlated to peripheral body fat. Such a skinfold might be obtainable from an abdominal site such as the iliac crest.

Two new skinfold sites, the iliac crest and calf skinfolds, were included in the Hispanic HANES protocol. This trend toward increasing the diversity of measures for estimating obesity should continue in NHANES III. When anemia was identified as a significant public health problem from NHANES I data, the sophistication and number of measurement tools for iron status increased in NHANES II. Likewise, the tools used to measure obesity should be sharpened in NHANES III so that health outcomes may be clarified. It is particularly important to include measures that distinguish between peripheral and centrally located body fat and those that are sensitive and specific indicators at lower levels of adiposity. By the criteria of precision, the calf skinfold is properly not included in NHANES III, although it does give important information about peripheral body fat. The biceps skinfold is a more precise alternative. The addition of simple circumference measures, such as the minimum waist and maximum hip circumferences or diameters, would meet the criteria both of precision and of identifying centralized obesity. More sophisticated techniques for measuring adiposity should be included in NHANES as these become practical for field studies.

Relationship of obesity to caloric consumption and activity levels

Obesity results from excess caloric consumption relative to need. Factors that influence caloric requirements are age and growth rates, lean body mass, genetic variations in metabolic efficiency, less commonly metabolic abnormalities and disease states, and levels of physical activity. NHANES I data provide some evidence that caloric intake is related to reported levels of physical activity (table 12), although reported recreational activity levels accounted for, at most, 6.2% of the variance in caloric intake. A corollary finding is that obesity as measured by the sum of triceps and subscapular skinfolds is negatively associated with reported caloric consumption. These two findings support the conclusion that obesity in the United States is a function of underactivity rather than overeating.

Within samples of persons of similar sex and weight, activity levels predict actual energy needs. Among these groups, activity levels should be considered important nutrition variables. However, among persons of widely differing weights, the major predictors of energy needs are body weight and lean body mass (66). These variables predict both the resting metabolic rate and the caloric cost of physical exertion. Except among those engaged in intensive bouts of activity, the resting metabolism accounts for the largest percent of a day's energy needs.

Table 12. Correlation coefficients among the sum of triceps and subscapular skinfolds, age, caloric intake, and recreational and nonrecreational activity, by race and sex, in persons aged 12–74 years participating in the first National Health and Nutrition Examination Survey, 1971–74

	Correlation coefficients [1]			
	Male		Female	
Item	White (n = 5,141)	Black (n = 1,095)	White (n = 7,339)	Black (n = 1,766)
Skinfold sum by:				
Recreational activity [2]	−0.165	−0.198	−0.216	−0.245
Nonrecreational activity [2]	−0.140	−0.078	−0.121	−0.137
Reported kcal intake	−0.117	−0.049	−0.187	−0.168
Age	0.162	0.184	0.233	0.284
Reported kcal intake by:				
Recreational activity	0.190	0.248	0.126	0.181
Nonrecreational activity	0.180	0.228	0.074	0.096
Age	−0.394	−0.368	−0.247	−0.326
Age by:				
Recreational activity	−0.390	−0.556	−0.294	−0.360
Nonrecreational activity	−0.253	−0.329	−0.151	−0.190

[1] Significant at $p < .05$ if $r \geq .062$; at $p < .01$ if $r \geq .081$.

[2] 3-point self-reported exercise score: Little or none = 1, moderate = 2, much = 3.

Source: Blair, D. The energy expenditure and activity levels of lean, adult-onset, and child-onset obese women. Unpublished dissertation data. Cornell University, 1980.

The data that indicate obesity as a function of inactivity rather than caloric intake may be an artifact of reporting differences between the obese and lean. Measures of both reported dietary intake and self-reported activity levels depend highly on subject self-perception. In these measures, no reporting bias between the obese and lean is assumed. Existing data imply that such reporting biases may exist and may skew the results.

First, it was assumed in NHANES I that a given amount of inactivity means the same for an obese person as for a lean person. In a small study of energy expenditure in obese and lean women (66), the self-rated activity scores used in NHANES I were compared with actual measured energy expenditure and time spent in activities by middle-aged females. In lean women, the correlation between the NHANES I recreational activity scale and the actual measured time per day spent at energy expenditure levels >6 kcal/min was 0.912 (n = 8). For obese women, the correlation was 0.174 (n = 15). The obese women tended to rate themselves as inactive whether or not they were active. The NHANES I nonrecreational activity score showed no correlation in either group to the actual time per day spent in medium to heavy work ($r = 0.002$, $n = 23$). These data indicate that self-reported recreational activity scores underrate activity in the obese (67).

Second, the 24-hour recall of dietary intake is a situation that favors underreporting of large intakes (38). Underreporting may be more predominant in the obese than in the lean. Several studies have shown that the accuracy of recall methods of estimating dietary intake in obese subjects depends on the extent of the probing techniques. Caloric intakes reported by 24-hour recall methods were originally lower in the obese than lean adults but were significantly higher in the obese when dietary history techniques were employed (68, 69).

The scope of nutrition surveillance includes the provision of information that will "contribute to the analysis of causes and associated factors and so permit a selection of preventative measures which may or may not be nutritional" (70). Obesity has been delineated as a major public health problem by NHANES and other large surveys such as the Framingham Study. To fulfill the surveillance objective of causative and associative analysis, measures that attempt to delineate differences between the obese and lean must be quantitative, must put less emphasis on subject perception, and should take advantage of existing probing techniques.

Implications for development of indicators of body composition, dietary patterns, nutritional status, and physical fitness

In this review, we have attempted to summarize the most salient findings from the NHANES series. Some of these findings are clear cut. Others are clouded by interpretive difficulties. Differences observed in indicators may result from real changes in the underlying phenomenon or may be artifacts of changes in measurement or statistical methods. Limitations in the interpretability of NHANES data restrict the usefulness of the data for policymakers and program planners. We suggest the following eight steps as possible remedies:

1. Define the objectives of the data collection and analyses

NHANES data are used to ascertain the prevalence of states of health and of health risk factors in the U.S. population, to estimate differences in these prevalences among population subgroups, to identify potentially important determinants of these differences, and to estimate the impacts of these determinants on the observed differences in prevalence. NHANES data are also used in setting normative standards such as "growth standards." These objectives cannot be met by measuring only the outcomes of concern (e.g., health, nutritional status, physical fitness). They also depend on the design of the surveys and on the design of the statistical analyses of the data. For all these objectives, except for estimating the absolute prevalences, other kinds of data need to be collected.

It is critical to identify population groups for comparisons (e.g., targeting of programs) and to collect the appropriate data to meet these specific objectives. If the data are used to investigate the existence and impact of determinants, then the determinants must themselves be correctly measured. For example, studying the effect of food stamps requires knowledge not only about eligibility and present use of food stamps but also about the extent and timing of past use. Other determinants of outcome may confuse interpretation of the results unless they are identified and correctly measured. (For example, racial effects on hemoglobin, independent of income, must be taken into account because they will otherwise exaggerate the impact of poverty on anemia rates.) Often one will wish to measure factors (such as food intake) that are thought to covary with both a determinant (such as poverty) and an outcome (such as anemia). None of these needs for ancillary data collection is usually considered in discussions about indicators, nor will we consider them at greater length here. Nonetheless, they are as important to meeting NHANES objectives as are correct choice and measurement of the indicators (71).

2. Define the phenomenon that the indicator is supposed to reflect

The papers commissioned for this compendium, *Assessing Physical Fitness and Physical Activity in Population-Based Surveys*, have helped to define those underlying qualities of physical fitness that are desirable to measure. A similar process has been undertaken with dietary intake and body composition. When the phenomena of interest are carefully defined, it is possible to ascertain the match between indicators and the actual phenomena of interest.

The clarification of phenomenon-indicator relationships is a process that needs to be continued with other nutrition variable categories, such as nutritional status, in the context of the specific purposes for which NHANES is designed. Developing such a consensus will require efforts similar to this one on physical fitness and the NHANES III Dietary Survey Methodology Workshop in March 1986.

3. Develop and validate responsive and specific indicators

Once the phenomena are defined and operationalized, indicators that are responsive to changes in the specific underlying phenomenon can be developed and validated. This usually proceeds by identifying indicators in clinical settings that are responsive in some range of the underlying phenomenon and then ascertaining the extent of the range of responsiveness. For instance, hemoglobin is responsive to improvement in very deficient iron nutriture but not in moderate deficiency; ferritin, an indicator of iron stores, is not responsive in the very deficient range because the stores remain empty as iron is incorporated into hemoglobin. Ferritin becomes more responsive as deficiency decreases. Once the responsiveness has been ascertained in clinical settings, it must be examined in free-living populations, where the responsiveness is usually attenuated (71).

NHANES planners are often urged, for the reasons given in this review, to adopt functional indicators of nutrition status such as cognitive development and disease response. However, these indicators have been shown to be responsive only in severe to moderately severe malnutrition, levels that are more sensitively and specifically picked up by clinical, anthropometric, and biochemical indicators already in use in NHANES. These are also levels of malnutrition so rare in the United States that the prevalences are too low to be picked up by NHANES.

The indicator should be responsive to changes in the underlying phenomena of concern but should not be responsive to other determinants. Optimally, the indicator should be specific to the phenomenon of concern. In reality, the goal should be to choose indicators that are least likely to be affected by determinants that confuse or confound the interpretation of the findings.

This goal requires an examination of the kinds of findings one expects from NHANES. For instance, if NHANES data are to be used to determine whether or not poverty is an important determinant of malnutrition, indicators of nutritional status should not be strongly affected by other factors associated with poverty. In this case, cognitive development would be a particularly poor indicator for assessing the nutritional status of the poor.

Few indicators are perfectly specific. Sometimes a number of indicators can be found for the same underlying phenomenon that are affected by different confounding factors. Appropriate analysis of these combinations can then give reasonable estimates of the underlying phenomenon (21).

Sometimes the major confounding factor is known and can be quantified. Low serum albumin, an indicator of severe protein-energy malnutrition, may also be caused by nephrosis. Proteinurea, a diagnostic test for nephrosis, could be used as a marker for eliminating suspect cases from the analysis.

4. Document the reliability of indicators

Factors that affect the indicators independently of the underlying phenomenon affect not only the specificity of the indicator but also its responsiveness. Responsiveness can be thought of as the degree to which a signal can be differentiated from background noise. The signal is the actual change of the indicator in response to a change in the underlying phenomenon. The background noise is the random responses of the indicator to all other determinants. The more the random changes resulting from other factors or determinants, the lower is the responsiveness of the indicator to the signal in question. Some of this noise created by peripheral factors can be quantified by repeating measurements of an indicator over time while the underlying phenomenon remains constant. An indicator that varies only slightly from measurement to measurement relative to its overall range in a population is said to be reliable.

Knowledge about the reliability of NHANES data is essential for its interpretation, as is discussed under point 5. Therefore, in NHANES, replicate measures should be collected in a small subsample of subjects over a few weeks, as was done for anthropometry in NHES II (72) and NHANES II (73), and then the results should be published.

The noise that contributes to unreliability is made up of measurement variability and of variability of the indicator itself, which is independent of changes in the underlying reality. The measurement variability is called imprecision, and the variability of the indicator itself can be named undependability (74). Unreliability, imprecision, and undependability can be quantified and related to each other (74).

5. Document the precision of an indicator and improve it if necessary

Imprecision is estimated by repeated measurements of a stable indicator performed without knowledge of the previous measurements. For indicators that can be expected to fluctuate over a timespan, the repeat measurements must be done over shorter timespans. Such data are currently collected through NHANES for anthropometry and could be collected for almost all measured (in contrast to interview) data. This imprecision should be documented, the range of imprecision over measurings and over time should be analyzed, and the results should be examined to see if greater precision is necessary.

If reliability (73) is poor (for instance, <90% in anthropometry) and if imprecision is a large component of unreliability, then increases in precision should be sought either through improving the measurement methodology or through repeated, independent measures that are then averaged (74). For example, triceps skinfolds have a reliability of 86% in men in NHANES II (73) and imprecision makes up 96% of the unreliability. The average of two measures of triceps skinfold increases reliability to 93%. Thus it is worth continuing the present NHANES practice of measuring skinfolds twice and using the average.

Another important concern about precision is whether or not there are biases in measurement between the groups being compared. For example, different technicians may measure different groups or the obese may appear to eat less than others because they underreport food ingestion. Biases in measurement may also occur over time (unintentional changes in methodology). The biases in measurement among groups can be caused by inadequate randomization of measurers across groups; the extent to which this can occur needs some attention and documentation from NCHS analysts, although it is unlikely to be a serious source of error in NHANES. Differences in measurement resulting from characteristics of those being measured should be an object of scientific research fostered by NCHS.

Secular changes in methodology are the curse of any monitoring survey. The methodology to identify such shifts needs to be developed and implemented over all the important NHANES core variables. An example of the beginning of such work is the attempt to estimate the extent of the method-related shifts in values for hemoglobin and skinfolds between NHANES I and II. This includes inspection of the means and distributions over time and within surveys. It also includes looking at changes in factors that are important covariants of the indicator variable. For instance, weight and skinfolds usually vary together, but weight has not increased between NHANES I and II, as have skinfolds. Once a methodological shift is identified, it is important that data users be made aware of the fact and that appropriate statistical corrections be performed.

6. Improve dependability if necessary and possible

Undependability is calculated from unreliability and imprecision (73, 74). If reliability is poor and undependability is a large component of unreliability, an attempt should be made to increase dependability. This can be done by one of the following methods.

- Identify determinants that cause the random variability in the indicator and either stratify the measurements so as to control for this variation, or measure the determinant and control for it in the statistical analysis. For instance, the time of day is often a powerful determinant of the level of an indicator, such as transferrin saturation. Measuring everybody at the same time of day would reduce this undependability. Where this is impractical, recording the time of day and introducing this variable into the statistical analysis can achieve the same end (75).
- The mean of repeated measurements taken over a long enough timespan will be more reliable because both undependability and imprecision are reduced. This is the solution usually prescribed for improving the usefulness of dietary intake data. These data are notoriously unreliable in that the within-person day-to-day variability is very high relative to the phenomenon of concern, the between-person variability in usual intake.
- For some uses, estimates of the unreliability based on replicate measurements, as recommended in point 4, would be sufficient. These uses include but are not limited to estimating the likely quantitative relationships among underlying phenomena through regression and correlation coefficients (see point 7) or estimating the prevalences of the phenomenon of concern—for example, using dietary survey data to estimate the prevalence of dietary intakes inadequate to meet nutritional requirements (76).

It is particularly important to note that better measuring methodology, which only improves precision, is not the answer to remedying unreliability, which results from undependability.

7. Consider the impact of unreliability in estimating prevalence and differences in prevalence

NHANES data are used to estimate the prevalences of various phenomena and to identify and estimate differences in these prevalences within and across NHANES.

It is now becoming clear that estimating prevalences of underlying phenomena through counting persons who fall above or below a specified level (a cutoff point) of an indicator is fraught with major difficulties (34), especially if the indicator is unreliable (77). The results often yield unacceptable numbers of false positive cases. In theory these mistakes could be corrected, but in practice this seems unfeasible. Other methods have been suggested, such as comparing distributions of populations containing nonhealthy individuals with healthy populations (21, 75, 78) or with other standards (76). Unreliability will affect these alternative methods for estimating prevalence, but the extent of error has not yet been investigated. In fact, absolute prevalences are rarely required for any important decisions. A twofold mistake is often acceptable, but not the fivefold to tenfold errors sometimes being made.

More exactitude is required in identifying differences in prevalence. Often the most powerful test for identifying differences in prevalence is the statistical test for the difference in indicator means (77). The advantage of this test is that data on the indicator's precision and dependability can be used directly to ascertain how much they affect the statistical test (74). The estimate of the magnitude of the differences can then proceed using less powerful methods, where again the exact magnitude is not usually important.

8. Take reliability data into account in multivariate analysis

Examining NHANES data to identify determinants of outcomes and to estimate the impact of these determinants involves comparing different levels of the determinants with different levels of the outcomes. Because of the need to take many determinants into account at the same time, one usually uses multivariate analysis techniques. If ordinary least-squares techniques are employed, the reliability of the indicator can determine whether the indicator can be used for these analyses. For instance, it is useless to compare NHANES 24-hour recall data on dietary intake of vitamin A with vitamin A nutritional status because the reliability of 24-hour vitamin A intake is close to zero in the United States, and therefore the best r^2 that could be expected would be statistically insignificant. Sempos et al. (79) show how one calculates the number of times one needs to measure vitamin A intake in each individual to be able to use such data in multivariate analyses. This is clearly impossible in the NHANES context. In this case, another solution must be sought. For instance, a recall of the frequency of consumption of foods rich in vitamin A is feasible and would certainly be more reliable than the 24-hour recall. Such a solution, however, remains to be developed and validated.

If the indicator's reliability is adequate, then one can use the reliability data to estimate the maximum association (maximum r^2) that can be expected (74). It is this maximum expected proportion of the r^2, and not the absolute value of the r^2, that is important in interpreting the data. For some indicators, the maximum expected r^2 is 10%, in which case an actual r^2 of 5% would indicate a strong association. Reliability data can also be used to see how much unreliability attenuates the estimate of the magnitude of impact.

If the eight points outlined here were followed during the appropriate phases of the NHANES III design, implementation, and analysis, the resulting information would be readily interpretable and usable by policy planners. The survey represents a monumental and expensive effort of data collection and processing. Its continued funding rests on the data's usefulness, which is in turn dependent on the specificity and sensitivity of its indicator variables and the components of its reliability. Every step must be taken to promote the continued improvement of the reliability of the NHANES data so that results of high quality, continuity, and applicability may be assured and so that the survey's diverse functions can be achieved.

References

1. National Center for Health Statistics: Plan and initial program of the Health Examination Survey. *Vital and Health Statistics.* PHS Pub. No. 1000–Series 1–No. 4. Public Health Service. Washington. U.S. Government Printing Office, July 1965.
2. National Center for Health Statistics: Plan, operation, and response results of a program of children's examinations. *Vital and Health Statistics.* PHS Pub. No. 1000–Series 1–No. 5. Public Health Service. Washington. U.S. Government Printing Office, Oct. 1967.
3. National Center for Health Statistics: Plan and operation of a Health Examination Survey of U.S. youths 12–17 years of age. *Vital and Health Statistics.* PHS Pub. No. 1000–Series 1–No. 8. Public Health Service. Washington. U.S. Government Printing Office, Sept. 1969.
4. Panel of the National Academy of Public Administration: *Improving the Health and Nutrition Examination Survey.* Washington, D.C.: National Academy of Public Administration, 1980.
5. Citizens' Board of Inquiry into Hunger and Malnutrition: *Hunger-USA.* Washington, D.C.: New Community Press, 1968.
6. U.S. Senate: Legislative History of the Select Committee on Nutrition and Human Needs. Staff report, 1976.
7. Public Law 90–174. Partnership for Health Amendments, Sec. 14, December 5, 1967.
8. *Ten-State Nutrition Survey, 1968–1970.* DHEW Pub. No. (HSM) 72–8134, 1972.
9. Ostenso, G.L. National nutrition monitoring system, a historical perspective. *J. Amer. Diet. Assoc.,* 84:1181–4, 1984.
10. National Center for Health Statistics, H.W. Miller: Plan and operation of the Health and Nutrition Examination Survey, United States, 1971–1973. *Vital and Health Statistics.* Series 1–Nos. 10a and 10b. DHEW Pub. No. (PHS) 79–1310. Public Health Service. Washington. U.S. Government Printing Office, Jan. 1977.
11. National Center for Health Statistics, A. Engel, R.S. Murphy, K. Maurer, and E. Collins: Plan and operation of the HANES I Augmentation Survey of adults 25–74 years, United States, 1974–1975. *Vital and Health Statistics.* Series 1–No. 14. DHEW Pub. No. (PHS) 78–1314. Public Health Service. Washington. U.S. Government Printing Office, June 1978.
12. National Center for Health Statistics, A. McDowell, A. Engel, J.T. Massey, and K. Maurer: Plan and operation of the Second National Health and Nutrition Examination Survey, 1976–80. *Vital and Health Statistics.* Series 1–No. 15, DHHS Pub. No. (PHS) 81–1317. Public Health Service. Washington. U.S. Government Printing Office, July 1981.
13. National Center for Health Statistics, Data collection forms of the Hispanic HANES, United States, 1982–1984. DHEW, Public Health Service (unpublished document).
14. National Research Council, Coordinating Committee on Evaluation of Food Consumption Surveys. *National Survey Data on Food Consumption: Uses and Recommendations,* Washington, D.C., National Academy Press, 1984.
15. Fidanza, F. Sources of error in dietary surveys. *Bibl. Nutr. Diet,* 20: 105–113, 1974.
16. Posner, B.M., C.L. Borman, J.L. Morgan, W.S. Borden, and J.C. Ohls. The validity of a telephone administered 24-hour dietary recall methodology. *Am. J. Clin. Nutr.* 36:546–553, 1982.
17. Lowenstein, F.W. Major nutritional findings from the first Health and Nutrition Examination Survey in the United States of America, 1971–1974. *Biblthca Nutr. Dieta.* 30:1–16, 1981.
18. Isaksson, B. Recommended methods used in the nutritional status assessment: Clinical signs and symptoms. *Biblthca Nutr. Dieta.* 20:52–61, 1974.
19. Christakis, G. How to make a nutritional diagnosis without really trying. A. Adult nutrition diagnosis. *J. Florida M.A.* 66:349–356, 1979.
20. Meyers, L.D., Habicht, J.P., and C.L. Johnson: Components of the difference in hemoglobin concentrations in blood between black and white women in the United States. *Am. J. Epidemiology,* 109:539–549, 1984.
21. Meyers, L.D., Habicht, J.P., Johnson, C.L., and C. Brownie: Prevalences of anemia and iron deficiency anemia in black and white women in the United States estimated by two methods. *Am. J. Public Health,* 79:1042–1049, 1983.
22. National Center for Health Statistics, J.D. Singer, P. Granahan, N.N. Goodrich, L.D. Meyers, C.L. Johnson: Diet and iron status, a

study of relationships. *Vital and Health Statistics.* Series 11–No. 229. DHHS Pub. No. (PHS) 83–1679. Public Health Service. Washington. U.S. Government Printing Office, Dec. 1982.

23. Pilch, S.M., and F.R. Senti, editors: Assessment of the zinc nutritional status of the U.S. population based on data collected in the second National Health and Nutrition Examination Survey, 1976–1980. Prepared for the Center for Food Safety and Applied Nutrition, Food and Drug Administration, under Contract No. FDA 223–83–2384 by the Life Sciences Research Office, Federation of American Societies for Experimental Biology, Bethesda, MD. Available from: FASEB Special Publications Office, Bethesda, MD. 1984.
24. Pilch, S.M., and F.R. Senti, editors: Assessment of the iron nutritional status of the U.S. population based on data collected in the second National Health and Nutrition Examination Survey, 1976–1980. Prepared for the Center for Food Safety and Applied Nutrition, Food and Drug Administration, under Contract No. FDA 223–83–2384 by the Life Sciences Research Office, Federation of American Societies for Experimental Biology, Bethesda, MD. Available from: FASEB Special Publications Office, Bethesda, MD. 1984.
25. Senti, F.R., and S.M. Pilch, editors: Assessment of the folate nutritional status of the U.S. population based on data collected in the second National Health and Nutrition Examination Survey, 1976–1980. Prepared for the Center for Food Safety and Applied Nutrition, Food and Drug Administration, under Contract No. FDA 223–83–2384 by the Life Sciences Research Office, Federation of American Societies for Experimental Biology, Bethesda, MD. 1984.
26. Pilch, S.M., editor: Assessment of the vitamin A nutritional status of the U.S. population based on data collected in the National Health and Nutrition Examination Surveys. Prepared for the Center for Food Safety and Applied Nutrition, Food and Drug Administration, under Contract No. FDA 223–84–2059 by the Life Sciences Research Office, Federation of American Societies for Experimental Biology, Bethesda, MD. 1985.
27. National Center for Health Statistics, R. Fulwood, C.L. Johnson, J.D. Bryner, E.W. Gunter, C.R. McGrath: Hematological and nutritional biochemistry reference data for persons 6 months–74 years of age: United States, 1976–80. *Vital and Health Statistics.* Series 11–No. 232. DHHS Pub. No. (PHS) 83–1682. Public Health Service. Washington. U.S. Government Printing Office, Dec. 1982.
28. Sauberlich, H.E. Current laboratory tests for assessing nutritional status. *Surv. Synth. Path. Res.* 2:120–133, 1983.
29. Dallman, P.R., R. Yip, and C. Johnson. Prevalence and causes of anemia in the United States, 1976 to 1980. *Am. J. Clin. Nutr.* 39:437–445, 1984.
30. U.S. Department of Health and Human Services and U.S. Department of Agriculture: Nutrition Monitoring in the United States—A Report from the Joint Nutrition Monitoring Evaluation Committee. DHHS Pub. No. (PHS) 86–1255. Public Health Service. Washington. U.S. Government Printing Office, July 1986.
31. Jacob, R.A. Zinc and copper. Symposium on Laboratory Assessment of Nutritional Status. *Clinics Lab. Med.* 1:743–766, 1981.
32. Report of Study Team IX, World Food and Nutrition Study, National Academy of Sciences. Washington, D.C. 1977.
33. Solomons, N.W., and L.H. Allen. The functional assessment of nutritional status: Principles, practice and potential. *Nutr. Review* 41:33–50, 1983.
34. Habicht, J.P., Meyers, L.D., and C. Brownie: Indicators for identifying and counting the improperly nourished. *Am. J. Clin. Nutrition* 35:1241–1254, 1982.
35. Committee on Dietary Allowances, Food and Nutrition Board. *Recommended Dietary Allowances.* 9th ed. Washington. National Academy of Sciences, National Research Council, 1980.
36. National Center for Health Statistics, Abraham, S., M.D. Carroll, C.L. Johnson, and C.M.V. Dresser: Caloric and selected nutrient values for persons 1–74 years of age: United States, 1971–1974. *Vital and Health Statistics.* Series 11–No. 209. DHEW Pub. No. (PHS) 79–1657. Public Health Service. Washington. U.S. Government Printing Office, June 1979.
37. Beaton, G.H., J. Milner, P. Corey, V. McGuire, M. Cousins, E. Stewart, M. de Ramos, D. Hewitt, V. Grambsch, N. Kassim, and J.A. Little. Sources of variance in 24-hour dietary recall data: Implications for nutrition study, design, and interpretation. *Am. J. Clin. Nutr.* 32:2546–2559, 1979.
38. Gersovitz, M., J.P. Madden, H. Smiciklas-Wright. Validity of the 24-hour dietary recall and seven-day record for group comparisons. *J. Am. Diet. Assoc.* 73:48–55, 1978.
39. Block, G. A review of validations of dietary assessment methods. *Am. J. Epidemiol.* 115:492–505, 1982.
40. Beaton, G.H. Uses and limits of the use of the Recommended Dietary Allowances for evaluating dietary intake data. *Am. J. Clin. Nutr.* 41:155–164, 1985.
41. Kerr, G.R., E.S. Lee, M.-K. Lam, R.J. Lorimer, E. Randall, R.N. Forthofer, M.A. Davis, S.M. Magnetti. Relationships between dietary and biochemical measures of nutritional status in HANES I data. *Am. J. Clin. Nutr.* 35:294–308, 1982.
42. National Center for Health Statistics, W.R. Harlan, A.L. Hull, R.P. Schmouder, F.E. Thompson, F.A. Larkin, J.R. Landis, Univ. Michigan: Dietary intake and cardiovascular risk factors, Part I. Blood pressure correlates, United States, 1971–1975. *Vital and Health Statistics.* Series 11–No. 226. DHHS Pub. No. (PHS) 83–1676. Public Health Service. Washington. U.S. Government Printing Office, Feb. 1983.
43. National Center for Health Statistics, H.W. Stoudt, A. Damon, R.A. McFarland, and J. Roberts: Weight, height, and selected body dimensions of adults, United States, 1960–62. *Vital and Health Statistics.* Series 11–No. 8. DHEW Pub. No. (HRA) 76–1074. Health Resources Administration. Washington. U.S. Government Printing Office, June 1965.
44. National Center for Health Statistics, H.W. Stoudt, A. Damon, R.A. McFarland, and J. Roberts: Skinfolds, body girths, biachromial diameter, and selected anthropometric indices of adults, United States, 1960–62. *Vital and Health Statistics.* Series 11–No. 35. DHEW Pub. No. (HRA) 74–1281. Health Resources Administration. Washington. U.S. Government Printing Office, August 1973.
45. National Center for Health Statistics, F.E. Johnston, P.V.V. Hamill, S. Lemeshow: Skinfold thicknesses of children 6–11 years. *Vital and Health Statistics.* Series 11–No. 120. DHEW Pub. No. (HSM) 73–1602. Health Services and Mental Health Administration. Washington. U.S. Government Printing Office, October 1972.
46. National Center for Health Statistics, F.E. Johnston, P.V.V. Hamill, S. Lemeshow: Skinfold thicknesses of youths 12–17 years, United States. *Vital and Health Statistics.* Series 11–No. 132. DHEW Pub. No. (HRA) 74–1614. Health Resources Administration. Washington. U.S. Government Printing Office, January 1974.
47. National Center for Health Statistics, S. Abraham, C.L. Johnson, and M.F. Najjar: Weight and height of adults 18–74 years of age, United States, 1971–1974. *Vital and Health Statistics.* Series 11–No. 211. DHEW Pub. No. (PHS) 79–1659. Public Health Service. Washington. U.S. Government Printing Office, May 1979.
48. Cronk, C.E., and A.F. Roche. Race- and sex-specific reference data for triceps and subscapular skinfolds and weight/stature. *Am. J. Clin. Nutr.* 35:347–354, 1982.
49. Bishop, C.W., and S.J. Ritchey. Evaluating upper arm anthropometric measurements. *J. Amer. Diet. Assoc.* 84:330–335, 1984.
50. Stavig, G.R., A.L. Leonard, A. Igra, and P. Felten. Indices of relative body weight and ideal weight charts. *J. Chron. Dis.* 37:255–262, 1984.
51. National Center for Health Statistics, S. Abraham and C.L. Johnson: Overweight adults in the United States, 1971–1974. *Vital and Health Statistics.* Series 11–No. 230. DHEW Pub. No. (PHS) 83–1680. Public Health Service. Washington. U.S. Government Printing Office, February 1983.
52. Bishop, C.W. Reference values for arm muscle area, arm fat area, subscapular skinfold thickness and sum of skinfold thickness for

American adults. *J. Parenteral and Enteral Nutr.* 8:515–522, 1984.

53. Bowen, P.E., and P.B. Custer. Reference values and age related trends for arm muscle area, arm fat area, and sum of skinfolds for United States adults. *J. Amer. Col. Nutr.* 3:357–376, 1984.
54. Womersley, J., and J.V.G.A. Durnin. The assessment of obesity from measurements of skinfold thickness, limb circumference, height and weight. In: The regulation of the adipose tissue mass. *Proceedings of the International Meeting of Endocrinology,* Vague J., and Boyer, J., eds. Marseilles. Exerpta Medica. Amsterdam, 1974.
55. Johnson, C.L. Personal communication, March 28, 1985.
56. National Center for Health Statistics, J. Roberts and M. Rowland: Hypertension in adults 25–74 years of age, United States, 1971–75. *Vital and Health Statistics.* Series 11–No. 221. DHHS Pub. No. (PHS) 81–1671. Public Health Service. Washington. U.S. Government Printing Office, April 1981.
57. Blair, D., J.-P. Habicht, E.A.H. Sims, D. Sylwester, and S. Abraham. Evidence for an increased risk for hypertension with centrally located body fat and the effect of race and sex on this risk. *Am. J. Epidemiol.* 119:526–540, 1984.
58. National Center for Health Statistics, W.R. Harlan, A.L. Hill, R.P. Schmouder, et al. Dietary intake and cardiovascular risk factors, part II. Serum urate, serum cholesterol, and correlates: United States, 1971–75. *Vital and Health Statistics.* Series 11, No. 227. DHHS Pub. No. (PHS) 83–1677. Public Health Service. Washington. U.S. Government Printing Office, March 1983.
59. Bray, G.A., ed. *Obesity in America.* NIH Pub. No. 79–359. Public Health Service and National Institutes of Health, DHEW. Washington. U.S. Government Printing Office, 1979.
60. National Institutes of Health. Health Implications of Obesity. National Institutes of Health Consensus Development Conference Statement 5(9):1–6. U.S. Government Printing Office, Feb. 1985.
61. Keys, A., F. Fidanza, M.J. Karvonen, N. Kimura, and H.L. Taylor. Indices of relative weight and obesity. *J. Chron. Dis.* 25:329–343, 1972.
62. Kissebah, A.H., N. Vydelingum, R. Murray, et al. Relation of body fat distribution to metabolic complications of obesity. *J. Clin. Endocrinol. Metab.* 54:254–260, 1982.
63. Damon, A., S.T. Damon, H.C. Harpending, et al. Predicting coronary heart disease from body measurements of Framingham males. *J. Chronic Dis.* 21:781–802, 1969.
64. Feldman, R., A.J. Sender, A.B. Sigelaub. Difference in diabetic and nondiabetic fat distribution patterns by skinfold measurements. *Diabetes,* 18:478–484, 1969.
65. Butler, W.J., L.D. Ostrander, Jr., W.J. Carman, et al. Diabetes mellitus in Tecumseh, Michigan: Prevalence, incidence and associated conditions. *Am. J. Epidemiol.* 116:971–980.
66. Blair, D., and E.R. Buskirk. Habitual daily energy expenditure and activity levels of learning and adult-onset and child-onset obese women. *Am. J. Clin. Nutr.* 45:540–550, 1987.
67. Blair, D. The energy expenditure and activity levels of lean, adult-onset, and child-onset obese women. Unpublished dissertation data. Cornell University, 1980.
68. Beaudoin, R., and J. Mayer. Food intakes of obese and nonobese women. *J. Amer. Diet. Assoc.* 29:29–33, 1953.
69. Bray, G.A., B. Zachary, W.T. Dahms, R.L. Atkinson, and T.H. Odie. Eating patterns of massively obese individuals. *J. Amer. Diet. Assoc.* 74:24–27, 1978.
70. Habicht, J.-P., J.M. Lane, and A.J. McDowell. National nutrition surveillance. *Federation Proc.* 37:1181–1187, 1978.
71. Sahn, D.E., Lockwood, R., and Scrimshaw, N.S. Methods for the Evaluation of the Impact of Food and Nutrition Programs, United Nations University, Tokyo. 1984.
72. Malina, R.M., Hamill, P.V.V., and Lemeshow, S. Selected body measurements of children 6–11 years. *Vital and Health Statistics.* Series 11, No. 123. 1973.
73. Mueller, W.H., Slater, C.S., and Habicht, J.P. Reliability, Dependability and Precision of Anthropometric Measurements—HANES II. (In preparation)
74. Habicht, J.P., Yarbrough, C., and Martorell, R. Anthropometric Field Methods: Criteria for Selection. In Jelliffe, D.G., and Jelliffe, E.F.P. (editors). *Human Nutrition* Volume 11, Nutrition and Growth, Plenum Publ. Corp., pages 365–387, 1979.
75. Wen Harn Pan. Estimating the prevalence of iron-deficiency anemia in U.S. populations on the basis of changes in hemoglobin distribution at low transferrin saturation levels. Ph.D. Thesis, Cornell University, 1983.
76. Chairman—L.J. Filer; members—J.-P. Habicht (chairman, present committee), G.H. Beaton (vice chairman, present committee), J.J. Feldman, H.A. Guthrie, P. Havlik, D.M. Hegsted, K.K. Stewart, H. Smiciklas-Wright, and A.A. Tsiatis. *Nutrient Adequacy: Assessment Using Food Consumption Surveys.* Report of the Subcommittee on Criteria for Dietary Evaluation, Coordinating Committee on Evaluation of Food Consumption Surveys, Food and Nutrition Board, National Academy Press, Washington, D.C., 1986.
77. Brownie, C., and Habicht, J.P. Selecting cut-off point of diagnostic criterion for comparing prevalences of disease. *Biometrics,* 40: 675–684, 1984.
78. Brownie, C., Habicht, J.P., and Robson, D.S. An estimation procedure for the contaminated normal distributions arising in clinical chemistry. *J. Am. Stat. Assoc.* 78:228–237, 1983.
79. Sempos, C.T., Johnson, N.E., Smith, E.L., and Gilligan, C. Effects of intraindividual and interindividual variation in repeated dietary records. *Am. J. Epidemiology,* 121:120–130, 1985.

Part II

Fundamental Perspectives on Health-Related Physical Fitness and Cardiopulmonary Health

Chapter **4**

Design Issues and Alternatives in Assessing Physical Fitness Among Apparently Healthy Adults in a Health Examination Survey of the General Population

Jack H. Wilmore, Ph.D.
University of Texas, Austin

Introduction

In 1980, the U.S. Public Health Service published a report entitled *Promoting Health/Preventing Disease: Objectives for the Nation* (1). This document presented a series of goals and objectives formulated in each of 15 priority areas of concern for health promotion and disease prevention. These goals and objectives represented the consensus of more than 500 individuals and organizational representatives from the public and private sector, each with expertise in one or more of these 15 designated areas. The areas addressed included: High blood pressure control, family planning, pregnancy and infant health, immunization, sexually transmitted disease control, toxic agent and radiation control, occupational safety and health, accident prevention and injury control, fluoridation and dental health, surveillance and control of infectious diseases, smoking and health, prevention of misuse of alcohol and drugs, improved nutrition, physical fitness and exercise, and control of stress and violent behavior.

For each of the 15 priority areas, the report briefly outlined the nature and extent of the specific problem area, described and evaluated available health promotion and disease prevention measures, and identified national objectives. The objectives were identified in sets under the following categories: Improved health status, reduced risk factors, increased public or professional awareness, improved services or protection, and improved surveillance or evaluation systems. The specific background information and recommendations in the area of physical fitness and exercise are presented in appendix A.

Since the publication of this report, much progress has been made in implementing plans for attaining these national objectives (2). This has been particularly evident in the specific area of physical fitness and exercise, with increased Government funding of basic research on the role of physical fitness and exercise in promoting

---

Note: Parts of this chapter were adapted with permission from: Wilmore, J.H., and D.L. Costill. *Training for Sport and Activity: The Physiological Basis of the Conditioning Process.* 3rd edition. Boston: Allyn and Bacon, 1987.

health and preventing disease. However, in the establishment of the objectives in the area of physical fitness and exercise, there were few data sets that provided accurate information on the existing fitness levels and exercise habits of the American population. Although national estimates of the physical fitness status of children and youth have been collected for more than 25 years, data on the physical fitness status of American adults drawn from national probability samples do not exist. Without a representative data base, it would be impossible to assess changes in the population as a result of various intervention strategies. Thus, staff of the National Center for Health Statistics (NCHS) have recently added more appropriate questions on exercise to their population-based surveys and have begun to explore how national assessments of physical fitness might be developed within the context of ongoing NCHS data systems.

The purpose of the present paper is to provide a broad overview of design issues and alternatives in assessing physical fitness among apparently healthy individuals in a health examination survey of the general population. The term "apparently healthy individual" refers to those individuals with no known disease who are asymptomatic for any disease. Physical fitness assessment in special populations, such as children, the elderly, and patients with specific diseases, presents unique design problems that are not addressed in this chapter but will be addressed in other chapters in this book. Further, a recent book edited by Skinner, *Exercise Testing and Exercise Prescription for Special Cases*, provides considerable detail in this area (3). This paper will initially attempt to answer the question, What is physical fitness? The specific components that combine to form the entity "physical fitness" will then be isolated, described, and discussed. This will be followed by a discussion of both laboratory and field assessment techniques for assessing each of the identified components of physical fitness, with an emphasis on the strengths and weaknesses of each technique, including administrative concerns. Finally, on the basis of this review, a recommendation will be made regarding the appropriate battery for assessing physical fitness in the context of an examination survey of the general population.

Physical fitness: A definition and an analysis of its components

Physical fitness is a term that is used frequently, but often in quite different contexts. The drill sergeant in the Marine Corps has his or her own unique definition of physical fitness, which is based on a perception of what attributes Marine recruits need to prepare themselves for a state of military emergency. The high school track coach, who has the responsibility of optimizing the performance of highly skilled athletes, has a definition of physical fitness that is likely to differ considerably from that of the drill sergeant. In fact, the track coach may define fitness differently for each athlete on the basis of his or her event because each event has its own specific requirements. The physical education specialist in the elementary school and the pediatrician will also differ in their definitions of physical fitness based on their different professional perceptions of the growing child.

A definition of physical fitness

Because of these different definitions, physical fitness would appear to be an amorphous term, defying precise definition. However, a detailed analysis of the definitions of the drill sergeant, the track coach, the physical education specialist, and the pediatrician would reveal that their differences are more a matter of emphasis than substance. This becomes clear after reviewing the classic definitions of physical fitness that have been reported in the literature over the years. In 1963, Hunsicker (4) stated that there were several operational definitions of physical fitness, which included: (a) Freedom from disease and disability, (b) the ability of the organism to maintain the various internal equilibriums as closely as possible to the resting state during strenuous exertion and to restore promptly after exercise any equilibriums that have been disturbed, and (c) the ability to meet the physical demands of one's daily routine and possess sufficient reserve for additional activities. Getchell (5) defines physical fitness as the capability of the heart, blood vessels, lungs, and muscles to function at optimal efficiency, with optimal efficiency defined as the most favorable health needed for the enthusiastic and pleasurable participation in daily tasks and recreational activities.

Throughout the years, classic tests of physical fitness, such as those developed and promoted by the American Alliance for Health, Physical Education, Recreation, and Dance (AAHPERD) (6) and the President's Council on Physical Fitness and Sports (7), have contained a mixture of items, such as the pullup or flexed-arm hang, situp, standing long jump, shuttle run, 50-yard dash, 600-yard run, and softball throw for distance. There has been considerable debate, however, over the appropriateness of a number of these test items for a physical fitness test battery. Pate (8) has made an important distinction between motor fitness, or athletic fitness, and health-related physical fitness test items. Health-related physical fitness includes the components of cardiorespiratory endurance, muscular strength and endurance, body composition, and flexibility. Motor fitness includes each of the health-related physical fitness components in addition to the components of agility, power, speed, and balance. He defines health-related physical fitness as the ability to perform strenuous physical activity with vigor and without excessive fatigue and a demonstration of physical activity traits and capacities that are consistent with minimal risk ofdeveloping hypokinetic diseases, i.e., diseases associated with physical inactivity.

The distinction between motor fitness and health-related physical fitness is an important one when contemplating the design of a physical fitness test battery for a national examination survey of the general population. From the examples used in the introduction to this section, the drill sergeant, track coach, and physical education specialist would be most concerned about motor fitness, with a different emphasis being placed on each of the specific components. Conversely, although the pediatrician recognizes the importance of proper development of motor fitness in the growing child, his or her major concern would be with health-related physical fitness.

Health-related physical fitness components

For the purpose of this paper, the sole emphasis will be on the health-related aspects of physical fitness. Health-related physical fitness includes only those fitness components that directly relate to disease prevention, health promotion, or both (8). These components have been recently defined by AAHPERD as cardiorespiratory endurance, muscular strength and endurance, flexibility, and body composition (8). Each of these components will be defined, described, and briefly discussed. Other components such as balance and reaction time become important in the aging population, but these will not be discussed.

Cardiorespiratory endurance

Cardiorespiratory endurance can be defined as the ability to perform high-intensity activity for a prolonged period of time without experiencing fatigue or exhaustion (9). It differs from muscular endurance in that it involves total body activity, which is limited by cardiovascular and respiratory factors. Muscular endurance, on the other hand, refers to the ability of a single muscle or a muscle group to sustain prolonged exercise, either of a rhythmical and repetitive nature, e.g., pullups, or of a static nature, e.g., sustained isometric contraction. A highly developed cardiorespiratory endurance capacity is typified by the marathon or cross-country runner or cross-country skier, who is able to run or ski long distances at a fast pace.

Exercise scientists agree that the best measure of cardiorespiratory endurance capacity is one's maximal oxygen uptake (10). In the research laboratory, oxygen uptake is measured directly while the individual exercises at increasing intensities on either a treadmill or cycle ergometer. Although other types of ergometers can be used, the treadmill and cycle ergometer are by far the most common. As the intensity of exercise increases, the oxygen uptake increases in direct proportion. Eventually, the individual will reach his or her maximum ability to deliver oxygen to the active tissue; i.e., the oxygen uptake will plateau as the rate of work continues to increase. This is illustrated in figure 1 using two examples, a highly fit athlete and an unfit individual. The individual is at or near exhaustion when oxygen uptake plateaus. The value achieved at the point of the plateau is referred to as the maximal oxygen uptake ($\dot{V}O_2$max).

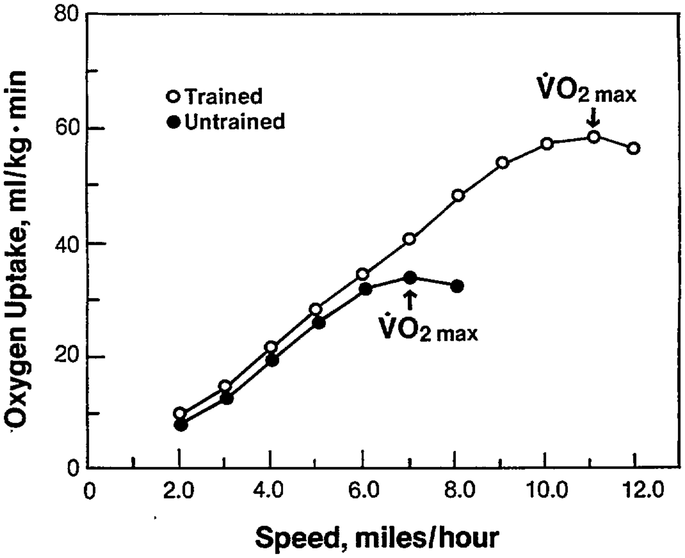


Figure 1. Maximal oxygen uptake in a trained and untrained individual, determined on a treadmill using increases in speed with the grade held constant at 0.0 percent.

With respect to the two examples in figure 1, the higher $\dot{V}O_2$max of the trained athlete allows him to run on the treadmill at substantially higher speeds than the untrained individual. This is an important point, for several of the field tests used to estimate maximal oxygen uptake are based on this linear relationship between speed and oxygen uptake capacity.

Because oxygen uptake represents the product of cardiac output and arteriovenous oxygen difference, $\dot{V}O_2$max represents the ability of the individual to maximize both oxygen delivery, i.e., cardiac output, and oxygen extraction, i.e., arteriovenous oxygen difference. Maximal cardiac output is determined by the interaction of heart rate and stroke volume. Maximal heart rates do not usually differ significantly between highly fit and unfit individuals of the same age. In fact, highly fit individuals tend to have slightly lower maximal heart rates than low-fit individuals (9). High-fit individuals also have, however, considerably higher stroke volumes at the point of maximal exercise than unfit individuals. Arteriovenous oxygen difference is determined by one's ability to extract oxygen at the active site of muscular contraction. Again, the more fit individual appears to have achieved adaptations through training that allow greater extraction of oxygen at the tissue level. The combination of a higher stroke volume and an increased ability to extract oxygen in the muscles provides the highly trained individual with a greater cardiorespiratory capacity, which is reflected in a high $\dot{V}O_2$max compared with the untrained individual.

Maximal oxygen uptake values relative to fitness category and age are presented in table 1 and figure 2. Values of athletes trained in different sports are presented in appendix B, table B–1. $\dot{V}O_2$max is a particularly good index of cardiorespiratory endurance capacity, as it is highly responsive to changes in the activity level of the individual. With endurance training, $\dot{V}O_2$

Table 1. Maximal oxygen uptake values, by age and fitness level

	Maximal oxygen uptake, ml·kg^{-1}·min^{-1}				
Age	Low	Fair	Average	Good	High
10–19 years	Below 38	38–46	47–56	57–66	Above 66
20–29 years	Below 33	33–42	43–52	53–62	Above 62
30–39 years	Below 30	30–38	39–48	49–58	Above 58
40–49 years	Below 26	26–35	36–44	45–54	Above 54
50–59 years	Below 24	24–33	34–41	42–50	Above 50
60–69 years	Below 22	22–30	31–38	39–46	Above 46
70–79 years	Below 20	20–27	28–35	36–42	Above 42

Note: Because values for females are generally 20 percent lower, on the average, than values for males, normal values for females can be obtained by shifting over one category to the right; e.g., the "fair" category for males would be considered "good" for females.

Source: Adapted from: Wilmore, J.H. *Training for Sport and Activity: The Physiological Basis of the Conditioning Process.* Second Edition. Boston: Allyn and Bacon, 1982, p. 274.

max will increase in proportion to the training stimulus, i.e., intensity, duration, and frequency of exercise (11). Likewise, with physical inactivity or bed rest, $\dot{V}O_2$max will decrease accordingly (12). A word of caution must be raised at this point. There is a strong genetic component that influences one's $\dot{V}O_2$max. This was first demonstrated by the work of Klissouras (13), who concluded that 93.4% of variation in $\dot{V}O_2$max was genetically determined. In a recent review of literature on the genetics of physiological fitness, Bouchard and Malina (14) indicate that the genetic component is not as dominant as was initially proposed by Klissouras. They conclude, on the basis of their analysis of the literature in this area, that heritability accounts for not more than 40% to 60% of the overall variability in $\dot{V}O_2$max, and possibly less. (Refer also to chapter 19.)

The concept of heritability of any physiological variable is of particular importance to cross-sectional epidemiological studies. Using $\dot{V}O_2$max as an example, several studies have conducted cross-sectional examinations of large populations and have made the assumption that $\dot{V}O_2$max is an accurate index of general physical activity level (15–17). Saris et al. (18) and Noonan (19) have recently demonstrated that there is not a close relationship between physical performance capacity or $\dot{V}O_2$max and daily physical activity levels, as estimated by 24-hour heart rate monitors. Saris et al. (18) reported that children ages 4–6 and 8–12 years with high physical performance capacities spent fewer minutes with their heart rates in an active zone of 125 to 176 beats per minute than those with low physical performance capacities. Noonan (19) found no difference in physical activity levels across two schooldays and one weekend day between a low-fit (mean $\dot{V}O_2$max of 32.0 ml·kg^{-1}·min^{-1}) and high-fit (mean $\dot{V}O_2$max of 57.9 ml·kg^{-1}·min^{-1}) group of 13- to 15-year-old boys. Ashton (20) reported no significant difference in estimated $\dot{V}O_2$ peak or in other measures of cardiorespiratory fitness between highly active and inactive girls 9–10 years of age, as determined by an activity questionnaire. Thus, it is possible to have a high $\dot{V}O_2$max value as a result of genetic endowment yet have low to moderate levels of daily physical activity. Likewise, an individual could have very high levels of daily physical activity yet have a normal or below normal value for $\dot{V}O_2$max. Although this does not discredit the importance of assessing $\dot{V}O_2$ max in cross-sectional epidemiological studies, it does emphasize the importance of recognizing the limitations associated with any one variable. $\dot{V}O_2$max is an acceptable measure of cardiorespiratory endurance capacity, but it is not necessarily an appropriate index of daily levels of physical activity. The heritability of $\dot{V}O_2$max may well account for at least part of this discrepancy.

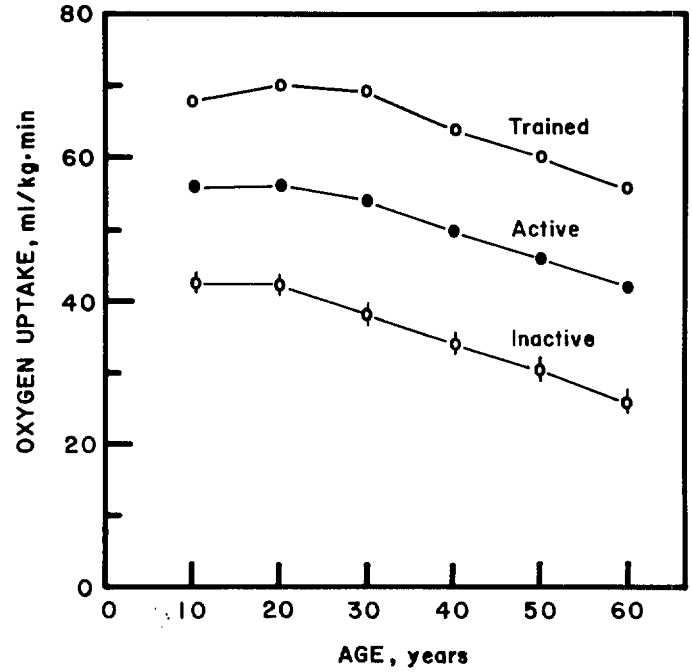


Figure 2. Maximal oxygen uptake in highly trained, active, and inactive individuals with aging. Adapted from Wilmore, J.H. *The Wilmore Fitness Program.* New York: Simon and Schuster, 1981.

Muscular strength and endurance

There are three different types of muscular contraction: Isometric, where the muscle length remains unchanged; eccentric, where the muscle performs a lengthening contraction, e.g., a controlled lowering of the weight in a two-arm curl; and concentric, where the muscle performs a shortening contraction. Isometric contractions are also referred to as static contractions, and eccentric and concentric contractions are referred to as dynamic contractions.

Strength refers to the ability of the muscle or a group of muscles to exert or apply force. Usually, the term strength is used in the context of one's maximum force-producing capabilities. The individual who can successfully bench press a maximum of 200 pounds is twice as strong as the individual who can bench press a maximum of only 100 pounds. In the purest sense, however, this illustration is not totally accurate, as the result of a bench press is work, i.e., work = force × distance, not just force alone. A static or isometric contraction provides a more precise estimation of strength. In fact, Atha (21), in his extensive review of the strength literature, defines strength very specifically as the ability to develop force against an unyielding resistance in a single contraction of unrestricted duration. However, dynamic tests of functional strength are generally considered acceptable.

For the athlete, and possibly for the average person as well, power is the key component for most physical activity. Power is defined as work per unit time, i.e., power = (force × distance)/time. Two individuals who can bench press a maximum of 200 pounds have identical functional strength. However, if one individual is able to execute his maximum bench-press strength in half of the time, e.g., 1.0 sec compared with 2.0 sec, he would have twice the power of the other individual. Although absolute strength is important for many activities, power is probably of even greater functional significance.

Muscular endurance refers to the ability of the muscle or group of muscles to sustain repeated contractions or to sustain a fixed or static contraction for an extended period of time. Muscular endurance is frequently equated with resistance to local muscular fatigue and is not to be confused with cardiorespiratory endurance associated with total body fatigue.

There is no universally accepted standard measure that is considered representative of muscular strength in the same manner that $\dot{V}O_2$max is considered representative of cardiorespiratory endurance capacity. First, there appears to be a high degree of specificity associated with strength. A person with a high level of grip strength does not necessarily have high levels of upper or lower body strength. Clarke (22) has conducted possibly the most extensive analysis of the specificity versus generality of muscular strength. Reporting the results of extensive testing on males and females from fourth grade through college, he concluded that generally the strength of various muscle groups throughout the body did not correlate highly with one another. While one intercorrelation reached $r = 0.91$, seldom did intercorrelations reach $r = 0.80$. Using cable-tension, static strength testing procedures on 25 selected sites, he reported that the highest correlations among strength tests were between muscle groups in the same joint area, frequently antagonists. Using the average of all 25 strength tests as the strength criterion, shoulder extension strength had the highest correlation with the strength criterion in the upper elementary and junior high school ages for both boys and girls, and for boys in senior high school.

As a result of the high degree of specificity of strength, some test batteries will include several estimates of strength representing different regions of the body. Clarke (22), using factor analysis, reported five cluster areas, including arm-shoulder strength, leg-back lift strength, lower leg strength, grip strength, and general strength. Jackson et al. (23) conducted a similar factor analytic study using one-repetition maximum strength assessments of 12 selected sites. They identified three factors, which included upper and lower extremity as well as trunk strength. Thus, strength assessment in at least three areas, i.e., upper body, trunk, and lower body, would seem highly desirable for any national population survey.

Muscular power, although probably of greater functional significance than the single component of strength, has received relatively little attention relative to its assessment. Most scientists, clinicians, and practitioners would agree on the importance of muscular power to performance of physical activities, but the inability to obtain simple measurements of power has greatly restricted the development of an adequate data base. With advancements in technology, new testing systems interfaced with dedicated computers have recently provided the researcher with accurate and rapid means of power assessment. Within the next several years, it should be possible to define a representative power testing battery, which will then be logically followed by testing of the general population to establish normative data.

Muscular endurance appears to be highly specific to the muscle or muscle group tested. In addition, few pure tests of muscular endurance have been used in testing the general population. Most testing of muscular endurance has been achieved through field tests. In fact, many of these field tests could be classified as combination strength-muscular endurance tests. Examples from existing test batteries would include the pullup, flexed-arm hang, and situp from the AAHPERD Youth Physical Fitness Test (6) and the President's Council on Physical Fitness and Sports' Youth Physical Fitness Test (7). Use of laboratory test equipment, such as computerized dynamometers, allows for much more accurate assessments of muscular endurance as a distinct component. As an example, such devices allow the expression of maximal force on each repetition. Having the individual perform a series of maximal contractions at a constant rate, e.g., one contraction every 2 seconds, for a set time period, e.g., 1 minute, allows the calculation of the total work performed during the test period, the decrement in power output from the initial few contractions to the final few contractions, the variation in performance at varying speeds of contraction, or other measures that would reflect the muscular endurance of that muscle or muscle group. Because the use of these more advanced laboratory testing devices has been greatly limited, normative data are not available. It is of interest to note, however, that even with these standard tests of muscular endurance, the resulting values may reflect both the absolute strength of the individual and his or her genotype. Clarkson et al. (24) reported that plantar flexion rate of fatigability was correlated highly ($r = 0.84$) with the percentage of slow-twitch muscle fibers in the gastrocnemius. Komi et al. (25) had demonstrated previously that there is a predominant genetic influence on the skeletal muscle fiber composition in man, on the basis of their studies of monozygous and dizygous twin pairs.

Flexibility

Flexibility refers to the range of motion, or to the looseness or suppleness of the body or specific joints, and reflects the interrelationships among muscles, tendons, ligaments, and the joint itself (9). Most young people have relatively high levels of flexibility, but this is rapidly lost with aging. Too much flexibility can lead to joint laxity, which is associated with joint instability and increased susceptibility to injury. At the other extreme, joint tightness also predisposes one to increased risk for injury, particularly sports-related injuries. Several factors limit the range of motion at a joint, including the structure of the joint, the interface between two articulating surfaces, soft tissues surrounding the joint (e.g., muscles, tendons, ligaments, fascia, and skin), and certain pathologies or previous injuries (26).

Flexibility is not an easy component to measure accurately. In the laboratory, flexibility is typically assessed by an electrogoniometer or flexometer. Both of these devices are used to measure the degrees of rotation through the range of joint motion. The Leighton flexometer has a weighted 360-degree dial and a weighted pointer mounted in a case. The dial and pointer move independently and are both controlled by gravity. Both can be locked in position independently of each other. The segment to be measured is usually positioned at one extreme in the range of motion, the dial is locked in position, and then the segment is moved through the full range of motion. The pointer follows the movement of the segment, thus indicating the extent of movement, or degrees of movement. The electrogoniometer is a protractor-like device that is used to measure the joint angle at both extremes in the range of movement. A potentiometer provides an electrical signal proportional to the angle of the joint. Most of the major joints in the body can be measured using either of these devices.

Field testing of flexibility has generally been limited to the sit-and-reach test. In this test, the individual sits on the floor with the legs extended forward. Bending forward at the waist with the knees locked and pressed against the floor, the individual reaches forward as far as possible, attempting to touch the toes or as far beyond the toes as possible. The extent of forward reach is then measured. This test is used as an index of lower back and hamstring flexibility.

Unfortunately, flexibility is similar to strength (discussed previously) in that it is highly specific to the joint and joint action (26). Harris (27) investigated the structure of certain measures of flexibility by the technique of factor analysis in 147 college women. Using 42 joint action measures of flexibility and an additional 13 composite measures of flexibility, it was concluded that no one composite test or joint action measure provides a satisfactory index of individual flexibility. Thus, to develop an appropriate battery to test flexibility, it would first be necessary to define those joints or composite movements considered important to the study being conducted.

Body composition

The whole body is comprised of a number of different components, and these can be categorized into chemically or anatomically defined compartments. Possibly the simplest and probably the most widely used classification system differentiates total body weight into two components, fat weight and lean weight. Fat weight is defined as the weight of the body's total fat stores and includes what some have termed the essential fat. Lean weight, or fat-free body weight, is comprised of the skeletal mass, muscle mass, and the organs of the body, including the skin. Technically, the lean and fat-free weights are different entities, the lean weight including that fat essential for life. However, the two terms are now used interchangeably, and lean weight will be used in this chapter. By using such a classification system, it is possible to differentiate between overweight and obesity. Overweight is defined as weight exceeding the range of values established from general population norms for a given height and frame size. Obesity refers to the state of being overfat, i.e., the state in which the weight of fat expressed as a percentage of one's total body weight exceeds the value that is considered to represent the upper limit of normal. For men, values ranging from 20% to 25% body fat are considered borderline obesity, and values exceeding 25% are considered frank obesity. For women, the values are 30% to 35% for borderline obesity and 36% or above for frank obesity.

Most epidemiological studies have relied on standard or relative body weight, or weight in ratio to height, as an index of body composition. According to Grande and Keys (28), the standardized height-weight tables had their genesis in the 1912 publication entitled "Medico-Actuarial Mortality Investigation," published by the Association of Life Insurance Medical Directors and the Actuarial Society of America. The original tables gave the average values for men and women of specific ages who had obtained life insurance policies in the years of 1888–1905. The data were obtained mostly from urban centers on the eastern seaboard. Heights and weights were recorded with shoes and clothing, with a 1- and

2-inch allowance provided for heels for men and women, respectively, and a 4- to 9-pound allowance provided for clothing (29). Over the years, these same data have been reanalyzed and subsequently placed into a format that provides weight ranges for each of three frame sizes for each 1-inch increment in height (30). According to Grande and Keys, the placement of the original data into frame sizes was done in a rather arbitrary manner, and no instructions were given as to how to determine frame size (28). In 1983, the Metropolitan Life Insurance Company published a revised standard height-weight table that has generally increased weight allowances for a given height and frame size and has proven highly controversial (31).

The major problem associated with the use of the standardized height-weight tables is that the composition of the body is not taken into consideration. This was clearly demonstrated by Welham and Behnke (32) in their study of professional football players, reported in 1942. The players were found to be grossly overweight, yet they had high levels of lean body weight and low levels of body fat. This illustrates the important point that one can be overweight but of normal body composition. Likewise, it is possible to be excessively fat yet fall within the normal weight range for height and frame size.

A number of procedures have been devised for assessing body composition. In the laboratory, the densitometric technique has been regarded as the "gold standard," or the criterion technique. Other laboratory techniques include helium dilution for determining total body volume; isotopic dilution for determining total body water; ^{40}K for determining lean body weight; radiography of the extremities for determining proportions of bone, muscle, and fat; computer-assisted tomography to observe cross-sections of extremity segments; nuclear magnetic resonance; total body electrical conductivity; and electrical impedance. The densitometric technique was used in dogs in the early 1930's and in humans beginning with the early work of Behnke and his colleagues in the late 1930's and early 1940's (32, 33). Since that time, assessment of body composition by the densitometric technique has become widely used in both research and clinical practice.

The density of the human body can be accurately determined by measuring the mass or weight of the body and the volume of the body; i.e., density = mass/volume = weight/volume. Volume can be assessed by measuring the actual volume of water displaced by the body, or by the body's loss of weight when weighed under water; i.e., volume = $(\text{weight}_{\text{air}} - \text{weight}_{\text{water}})/\text{density}_{\text{water}}$. The volume of the body must be corrected for any trapped pockets of air within the body, such as gas in the gastrointestinal tract, or the volume of air remaining in the lungs at the time of the underwater weighing. Because the gastrointestinal gas is typically small in volume, highly variable, and difficult to accurately measure, most investigators will either ignore this volume or use a constant value of 100 ml (33). Thus, the equation for determining the density of the body is as follows:

$$\text{Density} = \text{weight}/\{[(\text{weight}_{\text{air}} - \text{weight}_{\text{water}})/\text{density}_{\text{water}}] - \text{residual volume}\}$$

Provided the measurements of weight in air and water and of residual volume are obtained properly, the resulting value for body density will be accurate. The major weakness of the densitometric technique is in the conversion of body density to relative body fat. The following equation is used in a two-component system for determining body composition:

$$\text{Density} = \frac{W_L + W_F}{W_L/d_L + W_F/d_F}$$

where W_L is the fraction of the total weight represented by the lean body weight. W_F is the fraction of the total weight represented by the fat weight, d_L is the density of the lean tissue, and d_F is the density of the fat tissue.

Solving this equation for W_F:

$$W_F = \frac{d_L d_F}{D(d_L - d_F)} - \frac{d_F}{(d_L - d_F)}$$

where D is the density of the body.

Although the underwater weighing technique is the most accurate laboratory test available to determine the total density of the body and its subsequent composition, it does have its limitations. For those individuals undergoing changes in bone mineral, i.e., increasing bone mineral in the youngster as he or she matures or decreasing bone mineral in the aging individual, the equations used to estimate relative fat from body density will be inaccurate, providing an overestimation of the actual fat percentage. For those individuals with a larger preponderance of bone or bone that is denser than the population average, the estimate of relative fat will also be inaccurate, with an underestimation of the actual fat percentage. These inaccuracies occur infrequently, but they do occur. The primary reason for these inaccuracies relates to the fact that, for the several equations available for translating body density to relative body fat, only a single constant is used to represent the density of all body fat, and a single constant is used to represent the density of all of the body's lean tissue.

There is little disagreement as to the appropriate value to use for the density of the fat tissue because there appears to be consistency in the density of fat among different sites on the same individual and consistency among individuals. The value typically selected to represent the density of fat is 0.9007 gm/cm^3 (34). There is, however, considerable variation in the density of the lean body mass among individuals. Although the value 1.100 gm/cm^3 has been used in most equations (34), this assumes a chemically mature individual who has not undergone substantial changes in lean tissue associated with inactivity and aging. Lohman (35) argues that, for certain groups of individuals, the density of the lean tissue should be less than the assumed value of 1.100 gm/cm^3. These groups would include children, females, certain athletic groups, and the aged. For other groups, such as black persons and lean athletes, it appears that the density of the lean tissue would be higher than that presently used (36). Schutte et al. (36) have recently presented convincing data that support the use of a higher value for the density of the lean tissue in a population of black persons, i.e., 1.113 versus 1.100 gm/cm^3. Although this may seem like an inconsequential difference, for a body density of 1.075 the relative body fat calculated from the previous formula would be 10.5% using a constant of 1.100 gm/cm^3 and 14.1% using a constant of 1.113 gm/cm^3. Lohman (37) has reported a similar finding for children, but in the opposite direction. He reported that a constant of 1.081 gm/cm^3 is much more appropriate for children 7–9 years, prepubescent children who have not achieved the density of the adult lean tissue.

This variability in the density of the lean tissue leads to a violation of the basic assumptions underlying the derivation of a specific density value for the conversion equation. To derive a single value for a specific equation, it must be assumed that each component of the lean body mass has a constant density that does not vary among individuals. In addition, the assumption must be made that each of these components makes a constant proportional contribution to the density of the lean tissue. It is now recognized that both of these assumptions are not based on a solid foundation, and that there are individuals and entire populations that violate these basic assumptions (38).

The underwater weighing technique, despite these limitations, is considered the criterion method for evaluating body composition. However, the procedure is time consuming, requires considerable space and equipment, and must be conducted by someone who is highly trained in body composition assessment. As a result, most body composition evaluations, particularly in nonresearch, clinical settings, are derived through anthropometric techniques. Using skinfold thickness measurements, girths, and diameters, either singly or in combination, it is possible to derive accurate estimates of body composition. For many years, data were reported that suggested that equations to predict body composition were highly specific to the population from which they were derived. Jackson and Pollock (39) and Jackson, Pollock, and Ward (40) derived a series of equations for men and women, respectively, that are generalized, i.e., applicable to all men and women of varying age and body composition. Sinning and his associates (41) recently confirmed the validity of these equations in groups of women athletes (41) and men athletes (42).

Relative body fat values vary by sex and level of physical activity. Although there is a substantial difference between college-aged males and females, i.e., mean values of from 13% to 16% for males and 20% to 25% for females, differences between the sexes are much less when comparing equally trained athletes. This is illustrated in appendix B, table B–2. The larger differences in the nonathletic population are probably the result of a more active male being compared with a relatively sedentary female (9). Thus, true sex-specific differences in relative body fat, independent of the level of physical activity, are more likely to range from 4 to 6 percentage units. Once full maturity is reached, relative body fat tends to increase with age. Again, this appears to be more related to reduced levels of physical activity and inappropriate dietary intake than to a "natural" response to aging.

The assessment of physical fitness

The theoretical basis for each of the four components of physical fitness having been provided, this section will be devoted to an indepth study of assessment procedures for each of the individual components. For each component, the criterion assessment procedure will be presented, followed by a discussion of each of the available laboratory and field tests and the associated administrative concerns. Administrative concerns will include an assessment of the quality of the resulting data, major sources of error associated with a specific test, the length of time needed for testing, number of support personnel needed, safety of the test procedures, and other considerations that may be unique to any one specific test or procedure.

Cardiorespiratory endurance capacity

As discussed in the introductory section of this paper, cardiorespiratory endurance capacity of an individual is best represented by his or her $\dot{V}O_2$max. $\dot{V}O_2$ max can be assessed directly, or it can be estimated from either laboratory or field tests. The direct assessment of $\dot{V}O_2$max is considered to be the criterion measurement of cardiorespiratory endurance capacity, and both laboratory and field tests can be used to estimate $\dot{V}O_2$max.

Criterion method

The direct assessment of $\dot{V}O_2$max involves the measurement of the inspired oxygen volume and the expired oxygen volume, with the difference representing the net oxygen consumed, or the oxygen uptake. Most laboratories have simplified this procedure by using the Haldane transformation, which allows the calculation of oxygen uptake without having to measure both inspired and expired ventilatory volumes (43). The Haldane transformation allows for the calculation of inspired volume from the expired volume, or vice versa, on the basis of the assumption that nitrogen is neither produced nor consumed by the body; i.e., the volume of nitrogen inspired equals the volume of nitrogen expired. Although this assumption was seriously challenged in the 1960's and early 1970's, subsequent research confirmed its validity (43). With the Haldane transformation, oxygen uptake is calculated from the following equation:

$$\dot{V}O_2 = \dot{V}E\{[1 - (FEO_2 + FECO_2)\ 0.265] - FEO_2\}$$

where $\dot{V}E$ is the expired ventilation corrected (STPD), FEO_2 is the fraction of oxygen in the expired air, and $FECO_2$ is the fraction of carbon dioxide in the expired air. It is also possible to calculate $\dot{V}O_2$ using the inspired ventilatory volume instead of VE. Some laboratories prefer to measure the inspired volume because of the characteristics of their gas volume measurement system.

The measurement of $\dot{V}O_2$ can be accomplished through the use of a manual system (44), a semiautomated system (45), or a totally automated system (46). Studies comparing the accuracy of all three systems indicate that, when properly calibrated, they provide results within an acceptable range; i.e., standard errors of measurement are relatively low. When selecting a specific system for use in a large population survey, it is important to consider both the advantages and disadvantages of each system. Manual systems are still considered the criterion, or gold standard, against which proposed systems are compared. Manual systems, however, require considerably more time and technical assistance than automated or semiautomated systems. Automated systems simplify considerably the measurement process, but backup components must be available in the event of a systems failure. Also, proper calibration of any automated system is critical for obtaining accurate results. There is a tendency to accept at face value the manufacturer's specifications without conducting periodic calibrations of each component comprising the total system. Semiautomated systems are possibly the best compromise, as they simplify considerably the data collection process, yet they are relatively easy to calibrate, trouble shoot in the event of a problem, and repair once the problem has been isolated.

The selection of an appropriate ergometer is also of considerable importance when designing a large population survey. Over the past 50 years, an impressive number of ergometric exercise testing devices have been developed and utilized in both clinical and research settings. Each of these devices has unique features that make it particularly appropriate for specific testing situations. Those devices that have been used most frequently in previous surveys will be discussed.

Step tests have been used both in the clinical evaluation of coronary artery disease (47) and in the general assessment of cardiorespiratory fitness (48, 49). Step tests are much more appropriate for submaximal testing, as the stepping rate or bench height needed for maximal levels of exertion would constitute a safety hazard. Step-test protocols have varied, with some protocols using a fixed bench height and a stepping rate that is increased at specified time intervals. Other protocols use a fixed stepping rate and increase the bench height at specified time intervals. Finally, there are protocols that increment the rate of work by increasing both stepping rate and bench height at specified time intervals. Step tests are the least expensive and most portable of the available ergometric modalities, but they have distinct disadvantages. First, the total amount of work performed in a standard period of time is totally weight and rate dependent. Second, it is difficult to monitor the test to assure that the individual executes the stepping procedure properly. As one approaches fatigue, there is the tendency to not complete a full step up on the bench, thus reducing the actual amount of work performed. Lastly, as a result of the considerable movement that occurs in both the vertical and horizontal dimensions, it is very difficult to monitor blood pressure, and artifacts in the electrocardiogram are common. Kamon (50) has described a laddermill that allows the physiological assessment of both negative and positive work. The laddermill is driven by an electric motor at speeds varying from 4 to 35 $m \cdot min^{-1}$. The ladder is inclined 30° from the vertical, and the distance between rungs is 35.6 cm.

Cycle ergometers have been the primary ergometric device for a number of years, being used almost exclusively in the early research studies in exercise physiology. Although they are still used extensively in both research and clinical settings, there has been a trend in the United States toward a more widespread use of treadmills. Cycle ergometers can be used in either the normal upright position or supine position.

Cycle ergometers generally operate on the basis of one of four principles: Mechanical friction, electrical resistance, air resistance, and hydraulic fluid resistance. With mechanical friction devices, a belt encompassing a flywheel is tightened or loosened, adjusting the resistance against which the individual pedals (51). Mechanical friction devices can be easily calibrated with known weights. However, these devices must be constantly monitored during the test to correct for subtle changes in resistance resulting from an altered coefficient of friction as the belt becomes warm. Resistance also can vary with an uneven application of force. This is particularly evident at high power outputs when the individual is struggling to maintain a constant pedal rate. Mechanically braked cycle ergometers are rate dependent; i.e., the power output is dependent on the subject's pedal rate. Therefore, care must be taken to monitor the subject's pedal rate, as any variation in pedal rate will alter the desired power output, thus affecting the standardization of the test protocol. Usually, testing is conducted at pedal rates of 50 or 60 rpm, with a metronome used to assist the subject in maintaining a constant pedal rate. It is necessary to use either a mechanical or electrical revolution counter, actuated by the crank shaft, to obtain the actual pedal revolutions. Actual pedal revolutions are then multiplied by the fixed distance the flywheel travels per revolution and by the resistance to obtain a precise calculation of power output.

Electrically braked cycle ergometers are rate independent; i.e., the power output is independent of the pedal rate. Viscous resistance is provided by a conductor that moves though a magnetic or electromagnetic field, with the strength of the field determining the resistance to pedaling. Rate independency is gained by a feedback loop that automatically increases the resistance as the rate decreases, and decreases the resistance as the pedal rate increases, maintaining a constant power output. The major difficulty with electrically braked cycle ergometers is in the calibration procedure. Calibration is typically performed at the factory prior to shipping the ergometer to the buyer. Further calibrations require sending the unit back to the manufacturer, purchasing a calibration device that is approximately equivalent to the price of the ergometer itself, or periodically renting a calibration unit. The electrically braked cycle ergometer is very expensive compared to mechanically braked ergometers, and the requirement of electrical power makes it less adaptable for field testing. Recent modifications have improved considerably on the original design of these electrically braked ergometers and have included microcomputers with programming capabilities (52).

Air resistance cycle ergometers are relatively new (53). The flywheel on the front of a standard mechanically braked ergometer is replaced by a wheel that contains a series of blades arranged in a manner similar to spokes on a wheel. The fan blades displace air as the wheel turns, with the resistance being directly proportional to the cube of the wheel's rpm; i.e., as the pedal speed increases, there is a corresponding increase in resistance (54). The concept is novel, and the ergometers appear to be sturdy and durable. However, it is presently difficult to achieve an accurate calibration of these devices; i.e., it requires a rather expensive calibration device (55). In addition, because the change in resistance is rate dependent, the effects of a changing mechanical efficiency with changes in pedal frequency (56) need to be evaluated with respect to the stated purpose of testing.

Cycle ergometers using hydraulic fluid to vary the resistance to pedaling have the capability of producing constant power outputs independent of pedal rate and have the added advantage of sustaining power outputs up to or in excess of approximately 3,000 kpm·min^{-1}. Unfortunately, these devices are factory calibrated, and there does not appear to be an acceptable method for properly calibrating these devices outside of the factory. The one device presently available on the market, the Fitron, manufactured by Lumex, is highly durable and is the only cycle ergometer presently available that can accommodate tall individuals; i.e., the seat height can be raised to accommodate those who are more than 6 feet tall.

There are many advantages to using cycle ergometers as opposed to other ergometric devices. The upper body is relatively immobile, which allows for greater accuracy in blood pressure determinations, rebreathing procedures, catheterizations, and blood-sampling procedures. The mechanically braked ergometers are relatively inexpensive and portable, the rate of work or power output can be accurately defined, and the task is weight independent; i.e., the rate of work can be set independently of the individual's body weight. This latter feature is important whenever investigating the submaximal responses to a standardized rate of work, or power output, when sequential tests are given and there is an intervening change in body weight. As an example, a 5-pound loss of weight resulting from an exercise training program will confound submaximal data derived from treadmill testing because the physiological responses to a standardized speed and grade on the treadmill will vary with the weight of the individual. With the cycle ergometer, the loss of weight will not alter the physiological responses to a standardized power output, a most important fact that is often overlooked by investigators. The ability to have precisely definable power outputs is also an advantage for certain types of investigations. Computing power output, or the rate of work on the treadmill, has presented a major problem that is yet to be resolved.

There are certain disadvantages in using the cycle ergometer. First, individuals tend to become more fatigued from local leg muscle fatigue than from general, overall cardiovascular, respiratory, and metabolic fatigue. In addition, and closely related to the problem of local fatigue, the peak or maximum physiological parameters obtained on the cycle ergometer are frequently lower than the same parameters obtained during treadmill exercise. This could be the result of localized fatigue; a reduction in venous return from the legs resulting from the longer period of contraction during cycling as opposed to running, thus reducing local arterial inflow through a reduction in cardiac output; or a lower active muscle mass when compared with treadmill exercise. Faulkner et al. (57) presented evidence that suggested that stroke volume is lower in cycle ergometer exercise, which they felt was the result of reduced venous return, i.e., trapping of blood in the active lower extremities, and/or a reduced diastolic filling time associated with a higher heart rate for the same rate of work. Miyamura et al. (58) reported similar findings and attributed the lower $\dot{V}O_2$max on the cycle ergometer to a lower cardiac output because of a lower maximum heart rate, a lower arteriovenous oxygen difference because of a smaller working muscle mass, and a lower calf blood flow.

The importance of proper calibration cannot be overemphasized when using any ergometer, and this is particularly important with cycle ergometers. Wilmore et al. (59) conducted a mechanical and a physiological calibration of four different cycle ergometers, each with a different system for applying resistance, i.e. friction-braked with a fabric belt, friction-braked with a disc brake, friction-braked with a hydraulic system with a Prony brake, and electrically braked. One ergometer was found to be consistently 10% below actual values at each power output. A simple adjustment of this ergometer allowed accuracy to within ±3%. A second ergometer demonstrated considerable variable error, from 4% to 80%. The remaining two ergometers were generally accurate to within ±3%. Although the errors associated with the last two ergometers were certainly within an acceptable range, the one ergometer would not be recommended for use in any research setting, and results for the other emphasized the importance of initial and periodic calibrations.

Treadmills have become the ergometer of choice for an increasing number of researchers and clinicians, particularly in the United States. There are a number of advantages in the use of treadmills. First, the treadmill is rate independent; i.e., the individual either maintains the belt speed or is carried off the back of the treadmill. Thus, unlike the case with most cycle ergometers, investigators do not have to concern themselves with monitoring resistance and cadence, as the rate of work remains constant for any given speed and elevation. Treadmill walking is a very natural activity, and individuals normally adjust relatively rapidly to the skill of treadmill walking after the first minute or two. However, even for the highly trained athlete, treadmill walking is a unique skill, and the first few minutes on the treadmill must be closely monitored.

Once the individual achieves a normal walking gait, with the arms swinging naturally, further difficulty is seldom encountered. Treadmills generally provide the highest physiological values for the respective variables measured, e.g., $\dot{V}O_2$max and maximal heart rate (HR max), when compared with the other ergometric devices. Although some athletes achieve higher values on ergometers that most closely approximate their mode of training or competing (60), the average individual will almost always achieve his or her highest values on the treadmill.

There are certain disadvantages in using treadmills. Treadmills are generally very expensive, particularly in comparison with other ergometric devices. They are also very bulky, occupying a great deal of space, and usually not very portable. For most testing, the standard power availability of 115V, 60 Hz, and 15 A is simply not adequate, and a power requirement of 208–220V, 3 phase, 60 Hz, and 7.8 A is necessary. With standard power, the treadmill will have cyclic decreases in speed with every foot strike, particularly with heavier individuals. Obtaining accurate blood pressure measurements is also a major problem when using a treadmill. This is the result of both the noise associated with normal treadmill operation and the difficulty of obtaining accurate measurements once the speed is such that the individual must switch from a brisk walk to a jog in order to maintain his or her position on the belt. Excessive upper body movement associated with brisk walking and jogging can result in considerable artifact in the electrocardiogram when the electrodes are not properly applied. This problem is further magnified by the heel strike.

An additional disadvantage in using treadmills relates to the safety factor. Although this seldom occurs, individuals can lose their balance or fail to keep up with the speed of the belt as they are getting tired, or the treadmill can be accidentally turned on while resting measurements are being obtained with the individual seated in a chair on the treadmill belt or while standing on the treadmill. To reduce the risk of loss of balance or falling, subjects or patients are sometimes allowed to either hold on to the handrails or to maintain a light fingertip contact with the handrails. This provides a partial solution to the problem of safety, but it creates another problem. Several studies have demonstrated that holding on to the handrails or even making light contact with the handrails reduces substantially the energy cost associated with any given rate of work (61, 62). Although this may not be a problem for a test that is strictly diagnostic, i.e., determining the normality of the response of the electrocardiogram to increased myocardial work associated with exercise, it is a major problem for a test that uses some end point of treadmill performance, e.g., length of treadmill test, to predict cardiorespiratory endurance capacity or $\dot{V}O_2$max. As an example, Haskell et al. (63) reported that holding onto the treadmill handrails significantly increased the estimated peak $\dot{V}O_2$ from 32.7 to 37.9 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$ in postmyocardial-infarction patients, yet directly measured peak $\dot{V}O_2$ was not affected, i.e., 32.1 versus 31.8 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$.

The relative merits of step testing, cycle ergometry, and treadmill testing are presented in table 2. This summary table evaluates many practical aspects of testing not normally considered. These should certainly be taken into consideration when selecting an appropriate ergometer for a large population survey.

When selecting a mode for testing in a large survey, it is important to recognize the concept of test specificity. As was mentioned earlier, most individuals will normally attain the highest $\dot{V}O_2$max when tested on a treadmill. If an accurate estimation of aerobic or endurance capacity is desired, this is an important point. Stromme et al. (60) tested 14 female and 10 male cross-country skiers, 8 elite male rowers, and 8 elite male cyclists while running to exhaustion on the treadmill and during maximal performance on their specific sport activity, i.e., uphill roll skiing, rowing in a single sculler, and uphill bicycling on a treadmill. The $\dot{V}O_2$max values were higher in almost every case when the athletes were tested on the sport-specific activity, with an average of 3.5, 4.6, and 5.8 percent higher $\dot{V}O_2$max while skiing, rowing, and cycling, respectively, compared with uphill running. These results underscore the specific local adaptations that occur in response to training. For the average untrained individual, this would not be a factor. This point is further illustrated by Magel et al. (64), who studied alterations in $\dot{V}O_2$max with swim training (1 hr/day, 3 days/wk for 10 weeks), with subjects performing maximal treadmill running and tethered swimming tests both before and after training. The initial $\dot{V}O_2$max while swimming was 15 percent lower than the $\dot{V}O_2$ during running. Following 10 weeks of swim training, the swimming $\dot{V}O_2$max increased by 11.2%, and the treadmill $\dot{V}O_2$max increased by only 1.5% (nonsignificant). Thus, had the treadmill been used as the exclusive mode of testing, it would have been concluded that swim training had no

Table 2. Relative merits of the step test, cycle ergometer, and treadmill

Criterion	Type of test Steps	 Upright bicycle	 Supine bicycle	 Treadmill
Ease of performance				
Familiarity with task required?	+ + +	+ +	—	+ + + [a]
Ease of obtaining high VO_2	+ +	+ +	±	+ + +
Subject's performance to $\dot{V}O_2max$	+	+ +	+	+ + +
Ease of instrument calibration	([b])	+ +[c], − −[d]	+ +[c], − −[d]	+ or ± [e]
Ease of measuring applied power	+ +[f]	+ + +	+ + +[f]	([g])
Ease of recording or obtaining the following during maximum test:				
ECG	±	+ +	+ +	±
Blood pressure	− −	+ +	+ + +	−
Blood samples	− − −	+ +	+ + +	±
Respiratory volume and oxygen	±	+ +	+ +	+
Need for providing for emergency care[h]	+	−	+ + +	− − −
Ease of breathing	+ + +	+ +	+	+ + +
Ease of obtaining a nearly continuous increase of effort[h]	±	+ +[c], + + +[d]	+ +[c], + + +[d]	+ + or ± [e]
Freedom from undesirable features				
Hazards	+ + + or ± [i]	+	+ + +	− −
Need for skill	+	+	−	+ +[i]
Occurrence of local muscle fatigue at high exercise levels	+	−	− −	+ +
Need for trained personnel	+ +	+ +	+ +	±
Cost of equipment	+ + +	+ +[c], − −[d]	+[c], − −[d]	− − −
Ease of maintenance (including need for constant calibration)	+ + +	+ +[c], ±[d]	+ +[c], ±[d]	±
Freedom from noise	+ + +	±	±	− −
Bulk of equipment[h]	+ + +	+	−	− − −
Ease of transporting equipment[h]	+ + +	+ +[c], ±[d]	±[c], − −[d]	− − −
Need for electricity[h]	([b])	+ +[c], − −[d]	+ +[c], − −[d]	− − −
Need for neuromuscular-skeletal coordination	−	−	+	− −
Ease of rate control[h]	− −	−[c], + +[d]	−[c], + +[d]	+ + +

Note: This table evaluates each of the 4 types of tests according to the criteria listed. A grading of + + + indicates easiest, greatest freedom from undesirable features, most advantageous, etc.; a grading of − − − indicates most difficult, least freedom from undesirable features, least advantageous, etc. The intermediate point is represented by a grading of ±. Throughout the table, therefore, the greater the number of plus signs (or the fewer the number of minus signs), the fewer the problems presented by the test concerned.

[a] More difficult when the rate and slope are high.

[b] Unnecessary.

[c] Friction type.

[d] Electric type.

[e] Calibration easy for angle, less easy for rate.

[f] Less easy at maximum power.

[g] Can be estimated only.

[h] Less important factor.

[i] Less at low stepping rate, greater at high rate.

Source: Exercise Tests in Relation to Cardiovascular Function. *World Health Organization Technical Report Series*, Report No. 388, 1968.

influence on $\dot{V}O_2max$; but in fact, major changes were realized. The results of this study were recently confirmed by Gergley et al. (65), who reported no change in $\dot{V}O_2max$ as determined on the treadmill following swim training, although $\dot{V}O_2max$ during tethered swimming improved by 18%. In a general population survey, there are certain groups of individuals for whom mode selection may be an important issue, e.g., athletes and those who perform predominantly upper body activity.

In addition to the selection of the appropriate mode of testing, selection of the appropriate test protocol is also of considerable importance. Exercise test protocols can basically be classified into one of two major categories, continuous and discontinuous. In the continuous type of test, the individual performs the test from its beginning to its conclusion without stopping. In the discontinuous type of test, the individual performs the test in fixed intervals, with a period of inactivity or light activity interspersed between exercise intervals. The rest intervals can extend from several seconds or minutes up to several days. Within each of these two major categories are several subcategories, including:

Continuous	*Discontinuous*
Steady state	Steady state
Steady-state incremental	Incremental
Nonsteady-state incremental	Physiologically controlled
Sinusoidal	
Ramp	
Branching	
Physiologically controlled	

In a steady-state exercise protocol, the objective is to maintain the power output constant for a period of time sufficient for the observed physiological parameters to reach a steady state, i.e., a plateau in observed values. Steady state is obtained for most physiological variables within 2–5 minutes. When power output is expressed as a percentage of the individual's maximum value, steady state will occur more rapidly at the lower work intensities and will occur at increasingly longer intervals as the relative intensity of work increases. Hagberg et al. (66) reported that the half time of the $\dot{V}O_2$ response increased as work intensity increased; i.e., it took longer to achieve a steady state. Once the relative intensity exceeds a certain critical point, possibly related to the onset of blood lactate accumulation (OBLA), there is a phase of cardiovascular drift, which is characterized by a gradual increase in heart rate and a gradual decrease in central venous pressure, stroke volume, and arterial pressure but a relatively constant cardiac output (67). This would imply that a true steady state, at least for these parameters, would be impossible to obtain once this critical intensity of exercise had been reached. Hagberg et al. (68) have observed the same upward drift in $\dot{V}O_2$ from the 5th to the 20th minute of exercise when the exercise intensity was maintained constant at 65% and 81% of $\dot{V}O_2$max. This drift in oxygen uptake was theorized to be the result of both the increased body temperature (Q_{10} effect) and the increased cost of ventilation.

With incremental power output protocols, the intent is to increase periodically the power output following either long periods of work (steady state) or short periods of work (nonsteady state). In the former, the power output is increased only after a steady state has been reached at the previous power output. This necessitates work intervals of at least 3-minutes duration and usually not longer than 10-minutes duration. In the nonsteady-state incremental power output protocol, the power output is increased either continuously from the start to the completion of the test, i.e., a continuous ramp function, or is increased following relatively short intervals at the previous power output, typically from 30-second to 2-minute intervals. The continuous ramp function protocol is a particularly attractive one as it is a relatively short protocol and is appropriate for determination of not only $\dot{V}O_2$max but anaerobic threshold, oxygen uptake kinetics, and work efficiency as well (69).

Sinusoidal power output protocols have been described by Wigertz (70) and Casaburi et al. (71). The sinusoidal protocol was developed to allow a better understanding of the kinetics of ventilatory and circulatory responses to exercise, i.e., physiological control and gas exchange dynamics. Variations in either amplitude and periodicity of the sinusoid or both allow the determination of controlling mechanisms of parameters such as ventilation and cardiac output. This would not be an appropriate protocol for a large population survey.

The branching protocol is a relatively new concept that is both versatile and efficient. With the branching protocol, all subjects follow the same work profile during the first stage of the test. At the end of the first stage, the heart rate response to work is evaluated, and if it is less than 70% of the HR max as estimated from the regression of HR max and age, the individual progresses to a second protocol that advances power output at a faster rate. If the heart rate is 70% of the predicted HR max or higher, the individual maintains the same protocol; i.e., the rate of change in power output does not change. At the end of the second stage, the same evaluation is performed. If the heart rate is still below 70% of HR max, the individual progresses to a third protocol that advances power output at an even faster rate. This continues until the individual reaches the point of volitional fatigue. A major advantage of this protocol is that all individuals reach the point of volitional fatigue within approximately the same time frame. This can be an important point when a large sample with large variations in endurance capacity is to be tested. With a fixed protocol, time to volitional fatigue could range from 4 minutes for the older, deconditioned individual to 30 minutes for the young, fit individual. The major disadvantage of the branching protocol is the lack of standardization from one subject to another. An example of a branching test protocol is illustrated in table 3.

The physiologically controlled power output protocol is used relatively infrequently, and its use is confined primarily to research applications. Any one of a number of physiological variables can be used as the controlling variable, although heart rate is typically the variable of choice. The ergometer power output is controlled by a feedback loop that adjusts power output to maintain a certain critical value. With heart rate as the controlling variable, power output could be controlled to provide exercise heart rates of 120, 140, 160, and 180 beats·min^{-1}. This type of protocol would not be very applicable for a large-scale population survey.

In selecting a protocol, it is important to understand that a protocol is simply a means to an end and not an end in itself. Thus, the purpose of the test must be clearly defined before the protocol can be selected. If the purpose of the test is to assess the cardiorespiratory endurance capacity of chronic obstructive lung disease patients, this will dictate an exercise test protocol substantially different from a protocol used to evaluate elite, world-class long distance runners. In fact, the protocol selected for the pulmonary patient would typically differ from that selected for the cardiac patient, and the protocol used in testing elite distance runners would

Table 3. An example of a branching protocol for treadmill testing

Stage	Time in test (min)	Treadmill speed and grade Branch 1 Speed	 Branch 1 Grade	 Branch 2 Speed	 Branch 2 Grade	 Branch 3 Speed	 Branch 3 Grade	 Branch 4 Speed	 Branch 4 Grade
I	0–1	3.1	0						
	1–2	3.1	2.5						
	2–3	3.1	5.0						
				↘ HR<70%					
II	3–4	3.1	7.5	3.5	5.0				
	4–5	3.1	10.0	3.5	7.5				
	5–6	3.1	12.5	3.5	10.0				
						↘ HR<70%			
III	6–7	3.1	15.0	3.5	12.5	3.75	10.0		
	7–8	3.1	17.5	3.5	15.0	3.75	12.5		
	8–9	3.1	20.0	3.5	17.5	3.75	15.0		
								↘ HR<70%	
IV	9–10	3.1	22.5	3.5	20.0	3.75	17.5	4.0	15.0
	10–11	3.1	25.0	3.5	22.5	3.75	20.0	4.0	17.5
	11–12	3.1	27.5	3.5	25.0	3.75	22.5	4.0	20.0

Source: Wilmore, J.H., and A.C. Norton. *The Heart and Lungs at Work.* Beckman Instruments, Inc., Schiller Park, IL, 1974.

differ from that used in testing elite sprinters. Too often, both clinicians and researchers make the mistake of selecting a common protocol for the testing of groups that vary widely in aerobic capacity.

When selecting a protocol for a specific group or population, certain factors must be considered. First, is it necessary to obtain steady-state values? If the purpose of the test is solely to determine the normality of the ECG response to exercise and $\dot{V}O_2$max, and steady-state values are not essential, then a steady-state protocol is not necessary. This would allow the use of a ramp protocol or a protocol with short work intervals, thus reducing the total time needed to administer the test. This is a particularly important consideration when planning a national survey in which thousands of individuals are to be tested.

The alternative to using shorter work intervals to reduce the time of the test is to use larger increments in power output from one step to the next. As an example, the Bruce protocol for treadmill testing uses 3-minute intervals but has rather major increases in speed-grade from one stage to the next (72). The metabolic equivalent of these increases in speed/grade from one stage to the next is almost 3.0 METS, a MET being the metabolic equivalent of the resting metabolic rate, i.e., approximately 3.5 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$. Although the 3-minute intervals per stage provide approximations of steady-state values, if steady-state values are not needed, it would be much easier for the individual performing the test to have 1-minute intervals and decrease the power output from one stage to the next by one-third, i.e., from 3.0 METS to 1.0 MET. Thus, the magnitude of the power output increment is the second major factor to be considered. The Bruce protocol, or a modification of the Bruce protocol, remains the protocol of choice for many clinical settings. Again, if steady-state data are required, then this is an acceptable protocol. However, there is the distinct possibility that this protocol will lead to premature termination of the test. As an example, a male patient is exercising on the treadmill using the Bruce protocol. He is in the third stage (3.4 mph, 14% grade) of the test and is approaching 90% of his $\dot{V}O_2$max by the end of the third and last minute of this stage. At the conclusion of this stage, the patient is advanced to the fourth stage (4.2 mph, 16% grade) and is unable to continue. Had the individual been given an increment of only 1.0 MET as opposed to 3.0 METS, he possibly could have continued for at least one more minute and achieved a more realistic $\dot{V}O_2$max value.

A third factor to be considered when selecting a test protocol, which is related to the first two factors, is the total length of time of the test. At one time, Boyer and Kasch (personal communication) proposed a protocol for determining $\dot{V}O_2$max on a cycle ergometer that would require a person of average fitness approximately 45 minutes to complete. At the other extreme, Åstrand and Saltin (73) found that $\dot{V}O_2$max could be achieved in athletes with a protocol having a duration of only 2 minutes. In a mass testing situation, where as many as 30–40 maximum tests are conducted in a single day, a short protocol of not more than 8–12 minutes in length will be required. In a clinical setting, where tests are scheduled one per hour, a much longer protocol can be selected. For diagnostic purposes, the physiological and clinical responses to exercise will dictate the length of the test. A brief test might be sufficient for identifying ST segment changes, but a longer protocol might be necessary to assess cardiac output changes with increasing power outputs or for the detection of dysrhythmias. From a physiological viewpoint, it is desirable to maintain the length of the test within a rather narrow range across the population to be tested. With wide variations in total time, uncertainty arises in the interpretation of the test results. Rather substantial shifts in plasma volume occur early in the test, and as the length of the protocol is extended, thermoregulation becomes of concern, and rather major adjustments in blood flow result (67). The drift from steady-state values, as mentioned earlier, is one consequence of these changes in plasma volume and blood flow. Thus, whether one is concerned only with maximal data or data collected at specific relative or absolute submaximal power outputs, *it is* extremely important to control for this time factor and the associated physiological alterations. If this is not controlled, it is necessary to recognize that it may not be appropriate to compare the data between two individuals or even to compare the data from one individual tested under two different conditions.

A fourth factor to be considered when selecting a protocol is the need for standardized normative data. As an example, many laboratories have selected the Bruce protocol primarily on the basis that there is a large data base available in the literature. This allows a comparison of the test results with those published in the literature, providing a more soundly based interpretation of those results. In addition, with the Bruce protocol, McDonough et al. (74) and Profant et al. (75) have published equations for the prediction of $\dot{V}O_2$max in men and women, respectively, using only the total length of time to volitional fatigue.

A fifth factor that must be considered when selecting a protocol is the influence that the protocol might have on the interpretation of longitudinal data obtained from studies in which there has been an intervention, e.g., training. As was discussed previously, there is a high degree of specificity involved with the choice of ergometer versus training mode when evaluating the physiological alterations resulting from training. Recently, Freund et al. (76) found that when subjects were trained by jogging or running on an inclined terrain, the increases in $\dot{V}O_2$max were greater when the subjects were evaluated on an inclined versus a horizontal treadmill protocol, i.e., 8.5% versus 5.3% improvement, respectively. Conversely, Allen et al. (77) reported that improvements in $\dot{V}O_2$max were not different between incremental and horizontal protocols when subjects were trained by jogging or running on a flat terrain.

From this discussion of protocols, it is clear that much care should be taken when selecting an exercise test protocol. An inappropriate protocol may provide erroneous data, lead to inaccurate diagnoses, and/or place the individual in a potentially hazardous situation. For most applications, the protocol selected should not exceed 15–20 minutes. The selected protocol should be of approximately the same length for all subjects to avoid the comparison of data collected under different physiological conditions. The selected protocol should provide increments in power output that do not exceed 1.0–2.0 METS, and the length of the work periods should be at least 3 minutes if steady-state data are desired and less than 3 minutes for a nonsteady-state test. Finally, it is recommended that protocols be designed that are customized to the problem under study. Existing protocols can be used as guidelines, but a customized protocol meets the needs of the specific study and should result in more representative data.

With the cycle ergometer, few standardized protocols have been developed. Most protocols start with a relatively low power output (150–300 kpm·min^{-1}, or 25–50 watts) and increase the power output by steps of 150 kpm·min^{-1}, or 25 watts, until the point of volitional fatigue. The duration of the work interval is usually 1 to 3 minutes, and pedal frequency is typically from 50 to 80 rpm.

A number of protocols have been developed for the treadmill. The Bruce test is by far the most widely used protocol (78). The protocol is as follows (72):

Stage	*Speed (mph)*	*Grade (%)*	*Time (min)*
I	1.7	10	3
II	2.5	12	3
III	3.4	14	3
IV	4.2	16	3
V	5.0	18	3
VI	5.5	20	3

Balke and Ware (79) proposed a treadmill protocol for use with U.S. Air Force aircrewmen, which has become known as the Balke protocol. This is a continuous protocol with a constant speed of 3.4 mph and 1-minute work intervals. The subject starts the protocol by walking for 2 minutes at 3.4 mph and at a 0% grade. The grade is then increased to 2% for the next minute and is increased by 1% each minute thereafter. This protocol has the advantage of having gradual increases in the rate of work, but it can be an excessively long test, e.g., longer than 20 minutes for individuals with $\dot{V}O_2$max values of 45 ml·kg^{-1}·min^{-1} or greater. Balke subsequently revised this protocol with a speed of 3.0 mph and grade increases of 2.5% every 2 minutes.

Froelicher (80) and Faris et al. (81) have both proposed modifications of the original Balke-Ware test in an attempt to reduce the length of time of the protocol. These protocols are as follows:

Stage	*Speed (mph)*	*Grade (%)*	*Time (min)*
I	2.0/3.3	3/0	3
II	3.3	6/5	3
III	3.3	9/10	3
IV	3.3	12/15	3
V	3.3	15/20	3
VI	3.3	18/25	3
VII	3.3	21/30	3
VIII	3.3	24/35	3

Sheffield and his associates at the University of Alabama, Birmingham, have conducted numerous studies with thousands of subjects. They have used predominantly the protocol of Doan et al. (82), which is nearly identical to the Bruce protocol, with the major difference being that the speed is expressed in km/hour.

Stage	*Speed (km/hr)*	*Grade (%)*	*Time (min)*
I	2.7	10	3
II	4.0	12	3
III	5.5	14	3
IV	6.8	16	3
V	8.0	18	3

Kattus et al. (83) proposed a protocol that has received rather widespread use. This test was developed for the patient population; thus the protocol begins at 1.0, 1.5, or 2.0 mph, depending on the severity of the patient's symptoms, if present. The grade is standardized at 10% and the speed is increased by 0.5 mph increments every 3 minutes. This protocol was later modified by Kattus et al. (84) as follows, the Kattus et al. modification being intended for a much healthier population:

Stage	*Speed (km/hr)*	*Grade (%)*	*Time (min)*
I	2.0	10	3
II	3.0	10	3
III	4.0	10	3
IV	4.0	14	3
V	4.0	18	3
VI	4.0	22	3

Ellestad (85) has described a protocol that he used for both symptomatic and asymptomatic populations. His protocol is as follows:

Stage	*Speed (mph)*	*Grade (%)*	*Time (min)*
I	1.7	10	3
II	3.0	10	2
III	4.0	10	2
IV	5.0	10	3
V	5.0	15	2
VI	6.0	15	3

Naughton (86) has proposed the following protocol, which has been used in a number of clinical settings:

Stage	*Speed (mph)*	*Grade (%)*	*Time (min)*
I	1.0	0	3
II	2.0	0	3
III	2.0	3.5	3
IV	2.0	7.0	3
V	2.0	10.5	3
VI	2.0	14.0	3
VII	2.0	17.5	3

Finally, several protocols have been developed for use primarily with healthy and/or athletic populations. The Taylor protocol is as follows (87):

Stage	*Speed (mph)*	*Grade (%)*	*Time (min)*
Warmup	3.5	10.0	15
I	7.0	0.0	3
II	7.0	2.5	3
III	7.0	5.0	3
IV	7.0	7.5	3
V	7.0	10.0	3
VI	7.0	12.5	3

The Åstrand protocol is as follows (88):

Stage	*Speed (mph)*	*Grade (%)*	*Time (min)*
I	(1)	0.0	3
II	(1)	2.5	2
III	(1)	5.0	2
IV	(1)	7.5	2
V	(1)	10.0	2
VI	(1)	12.5	2

According to Jopke (78), a national survey of 1,400 exercise testing facilities showed that the treadmill is the most widely used mode of testing (71%), followed by the cycle ergometer (17%) and step test (12%). Of the treadmill protocols, the Bruce (65.5%), Balke (9.7%), Naughton (6.0%), and Ellestad (3.1%) were the most widely used.

Several studies have compared the reproducibility of and physiologic responses to various protocols. Froelicher et al. (89) investigated the responses of 15 volunteers to maximal tests using the Bruce, Balke, and Taylor protocols. The Taylor protocol was modified from the original by including 5-minute rest periods between exercise stages, as opposed to conducting each stage on a different day. They found that the Taylor protocol resulted in significantly higher values for $\dot{V}O_2$max, and there was no significant difference between the Bruce and Balke protocols. All three protocols were found to be highly reproducible. Pollock et al. (88) tested a total of 51 subjects on each of four maximal treadmill protocols: Balke, Bruce, Ellestad, and Åstrand. Although there were no significant differences among the Bruce, Ellestad, and Åstrand protocols, the Balke protocol resulted in $\dot{V}O_2$max values that were slightly but significantly lower than those for the Åstrand protocol. The correlations among protocols for $\dot{V}O_2$max were all above $r = 0.90$.

One of the areas of contention in exercise testing is the use of maximal versus submaximal tests. Particularly in clinical medicine, there has been a trend toward submaximal testing. Yet, Bruce (90) and Froelicher (91), two leaders in clinical exercise testing, both advocate maximal or near-maximal exercise tests, i.e., sign and symptom limited. There are several major problems associated with submaximal testing. First, a submaximal test must establish some fixed end point for terminating the test. In most cases, a fixed percentage of the individual's predicted maximal heart rate is used as the end point. Typically, maximal heart rate is predicted on the basis of the linear decrease that has been noted in maximal heart rate with age in a number of cross-sectional studies. Fox et al. (92), in their summary of 10 of these cross-sectional studies, indicated that HR max could be estimated from the simple equation: HR max = 220 − age. This equation has become widely used for estimating HR max in submaximal testing, and a submaximal rate is

---

[1] Speed is maintained constant at from 5 to 8.5 mph, dependent on the subject's ability to run.

established from it. The specific end point used generally ranges from 80% to 90% of HR max, depending on the laboratory in which the test is administered. Unfortunately, although HR max does decrease with age in a linear manner, approximating 1.0 beat· min^{-1}, there is a very high standard deviation of approximately 10–15 beats· min^{-1}. For a 40-year-old man, the estimated HR max would be 180 beats· min^{-1} on the basis of the equation proposed by Fox et al. (92). Yet, the actual HR max could be anywhere from 150 to 210 beats·min^{-1}, using ± two standard deviations, which would still be considered a normal response. Using an end point of 80% of estimated HR max would result in a target heart rate of 144 beats·min^{-1}, and using 90% of estimated HR max would result in a target heart rate of 162 beats·min^{-1}. Thus, it is possible to conduct a submaximal test using a target heart rate that exceeds the actual HR max. Therefore, the major problem associated with the submaximal test is the fact that it is impossible to know what that test represents relative to the individual's actual capacity. For the same arbitrarily established end point, e.g., 85% of the estimated HR max, some individuals will be well below the predicted end point and others will be well above, i.e., from <70% to >100% of true HR max.

A second major problem associated with submaximal versus maximal tests concerns the lower sensitivity in diagnostic capability; i.e., the submaximal test will fail to correctly diagnose those with true coronary ischemia to a greater extent than will a maximal test. Cumming (93) reported that, out of 510 male volunteers, 27%, or 138, had a positive electrocardiogram indicative of myocardial ischemia. Had the test been terminated at 85% of the subject's predicted HR max, 50% of the positive tests would not have been detected. Although a certain percentage of those undetected positive tests would have been false positive tests, the incidence of false positive tests in this test population would not have been high.

Maximal tests provide superior screening results compared with submaximal tests (91). Further, Bruce states that maximal tests can provide much more information than submaximal tests, and the information can be obtained simply, reliably, and safely (90). One of the strongest arguments for maximal testing is the fact that a maximal test provides an objective, highly definable end point, where each individual is tested to the same extent. Part of the hesitation to use maximal testing concerns the fear of an associated increase in risk. This fear appears to be unfounded, as two studies have shown low risk in maximal or near-maximal testing. In fact, the risk of maximal testing appears to be no greater than that of submaximal testing (94). Stuart and Ellestad (94) reported complication rates of 3.58 myocardial infarctions, 4.78 serious arrhythmias, and 0.5 death per 10,000 tests in a national survey of exercise stress testing facilities, of which 70% were using symptom-limited maximum tests. Sheffield et al. (95), in a survey of 12 laboratories participating in the Lipid Research Clinics' Prevalence Survey involving 9,464 treadmill tests, reported no deaths, myocardial infarctions, or cardiac arrests during or within 24 hours of testing. The test protocol used in this survey was near maximal, not maximal.

In order for a test to be considered truly maximal, certain criteria should be met. Åstrand and Rodahl (96) have stated that the correct criterion for an attained maximal oxygen uptake should be a plateau, or leveling off of the oxygen uptake, despite an increase in the rate of work. Taylor et al. (87) have stated the $\dot{V}O_2$max has been achieved when there is less than a 0.15 liters·min^{-1}, or 2.1 ml·kg^{-1}·min^{-1}, increase in the oxygen uptake with an increase in the grade on a treadmill test of 2.5%. Because a plateau is difficult to achieve in all individuals tested, additional criteria have been proposed that include a respiratory exchange ratio in excess of 1.15 (97); a heart rate equivalent to or above the estimated HR max, predicted on the basis of the individual's age (98); and a blood lactate in excess of 100 mg/100 ml, or 8 mMol·liter^{-1} (99). Ratings of perceived exertion (RPE) are also being widely used in exercise testing. Because there is a linear relationship between RPE and oxygen uptake as the exercise test progresses, it is possible that an RPE of 18 to 20 would also be an acceptable criterion for the attainment of $\dot{V}O_2$max (100). Even with these supplementary criteria, it is not always apparent whether $\dot{V}O_2$max has been achieved. Niemela et al. (101) reported that, in 55 men and 44 women 35–62 years of age, the plateau in oxygen uptake was shown in only 17 subjects, the respiratory exchange ratio in excess of 1.15 was achieved by only 9 subjects, and HR max was often below the predicted value on the basis of the subject's age, even though most of the subjects achieved exhaustion by self-report on the cycle ergometer. For a large population survey, it is imperative that multiple criteria be established and that a true $\dot{V}O_2$max be accepted only if certain minimal criteria are achieved.

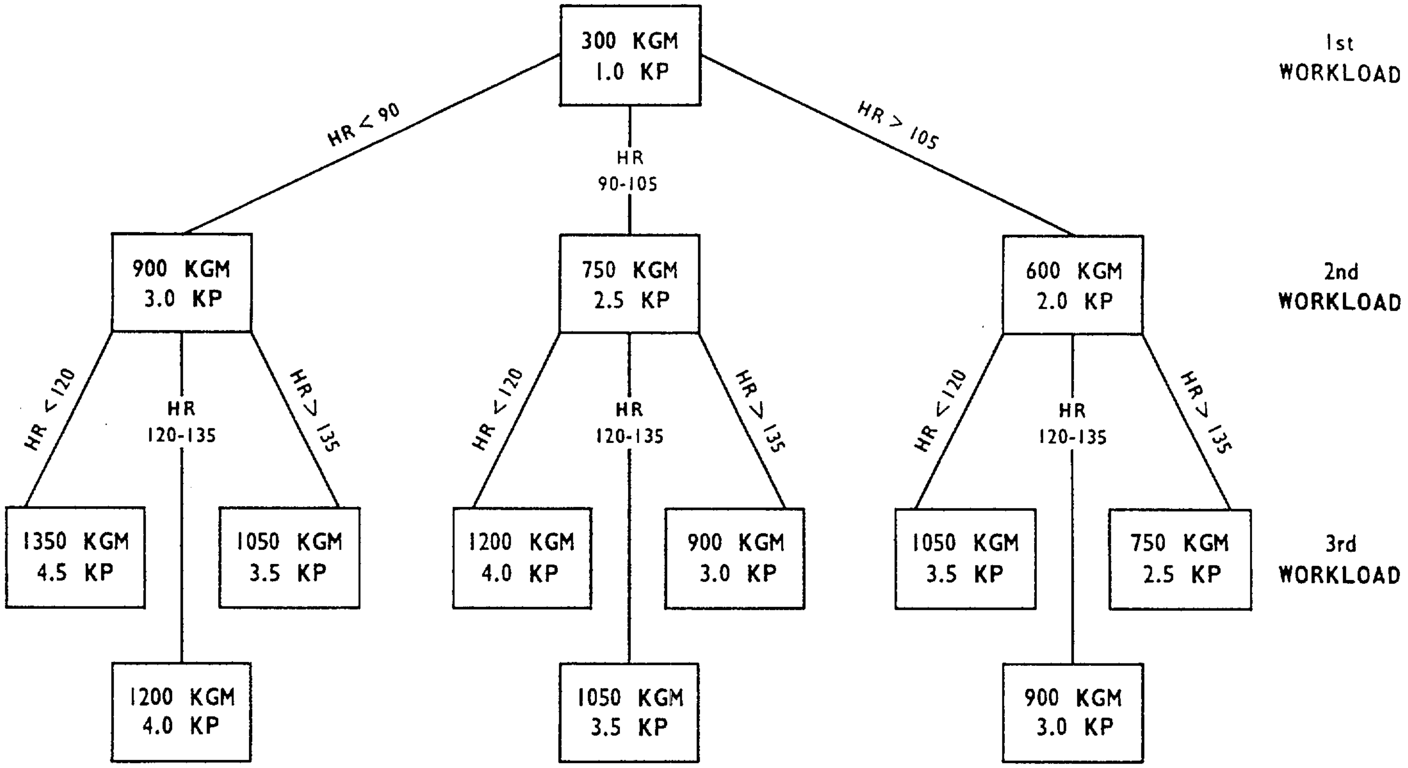
Additional methods

The direct assessment of $\dot{V}O_2$max requires expensive instrumentation, trained technicians, and a maximal effort on the part of the individual who is tested. Consequently, investigators have attempted to define simple submaximal tests that would provide accurate estimates of $\dot{V}O_2$max and could be administered to large numbers of individuals with a minimum of time and expense.

Techniques for the estimation of $\dot{V}O_2$max from submaximal tests developed in the earliest studies used either a step test (102–104) or cycle ergometry (102, 105, 106) as the mode of exercise. Because of difficulty in standardizing the step test, as discussed earlier, most studies estimating $\dot{V}O_2$max from submaximal exercise tests have used the cycle ergometer. The tests developed by Sjostrand (105) and Åstrand and Ryhming (102), or modifications of these test protocols, have been the most extensively used. Both tests are based on the fact that heart rate and oxygen uptake are linearly related over a range of work rates varying in intensity from light to heavy. The basic assumption underlying both tests is that the more fit individual will achieve a higher rate of work for the same heart rate and thus will be able to achieve a much higher rate of work before reaching HR max. In the Sjostrand (105) and the Wahlund (106) tests, working capacity is defined as the maximum work intensity consistent with steady state. Wahlund defined the concept of a functional maximal work capacity, and he referred to this as the physical working capacity at a heart rate of 170 beats·min^{-1}, i.e., the rate of work that elicits a heart rate response of 170 beats·min^{-1}. In the original Åstrand and Ryhming (102) test, $\dot{V}O_2$max is estimated from a single steady-state heart rate response to a fixed rate of work.

In 1973, the YMCA published a modification of the Sjostrand and Wahlund test procedures (107) that provides an estimate of $\dot{V}O_2$max by extrapolation of the HR/power output curve to an age-adjusted predicted maximal heart rate. Assuming a constant metabolic efficiency, the predicted maximal power output is converted to an oxygen consumption expressed in liter·min^{-1}, which, when divided by body weight, provides an estimate of $\dot{V}O_2$max in ml·kg^{-1}·min^{-1}. The test was revised in 1982 (108), and the guides to setting workloads are presented in figure 3 for men and figure 4 for women. The test procedure is outlined in figure 5 with an example of a 40-year-old male with a predicted maximal heart rate of 180 beats·min^{-1} (HR max = 220 − age). The estimated $\dot{V}O_2$max of 3.2 liters·min^{-1} translates into a value of 40 ml·kg^{-1}·min^{-1} when divided by the individual's body weight, i.e., 3,200 ml/80 kg.

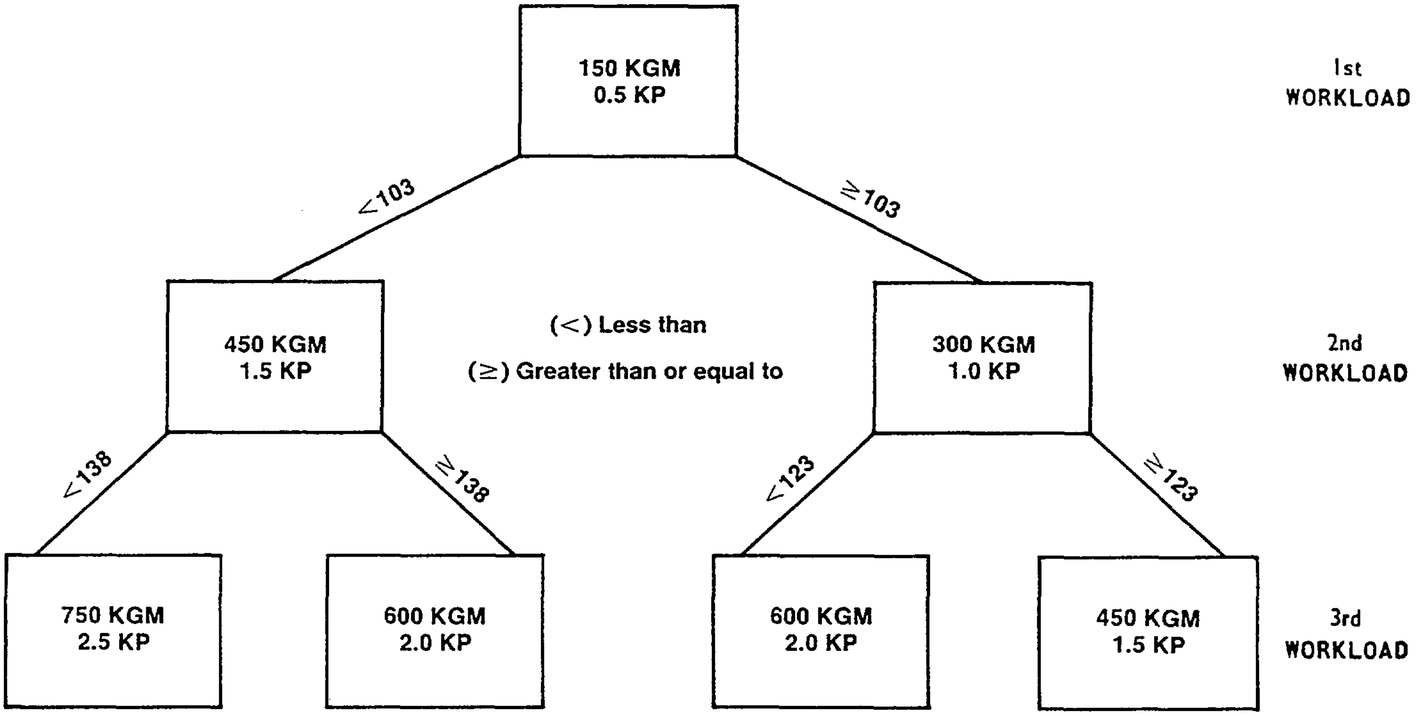
Siconolfi et al. (109) recently published a modification of the Åstrand-Ryhming test that appears to improve on the predictive accuracy of the original test. Starting at an initial power output of 24.5 watts (W) for men 35 years of age and older and for women of all ages and at 49 W



DIRECTIONS

1. Set the 1st workload at 300 kgm/min (1.0 KP)
2. If HR in 3rd min is: Less than (<) 90, set 2nd load at 900 kgm (3 KP)
   Between 90 and 105, set 2nd load at 750 kgm (2.5 KP)
   Greater than (>) 105, set 2nd load at 600 kgm (2.0 KP)
3. Follow the same pattern for setting 3rd and final load.
4. NOTE: If the 1st workload elicits a HR of 110 or more, it is used on the graph, and only ONE more workload will be necessary.

Figure 3. Guide for setting power outputs for men on submaximal cycle ergometer tests. Reprinted from Golding, L.A., C.R. Myers, and W.E. Sinning, editors. *The Y's Way to Physical Fitness,* Revised (1982) with permission of the YMCA of the USA, 101 N. Wacker Drive, Chicago, IL 60606.



DIRECTIONS

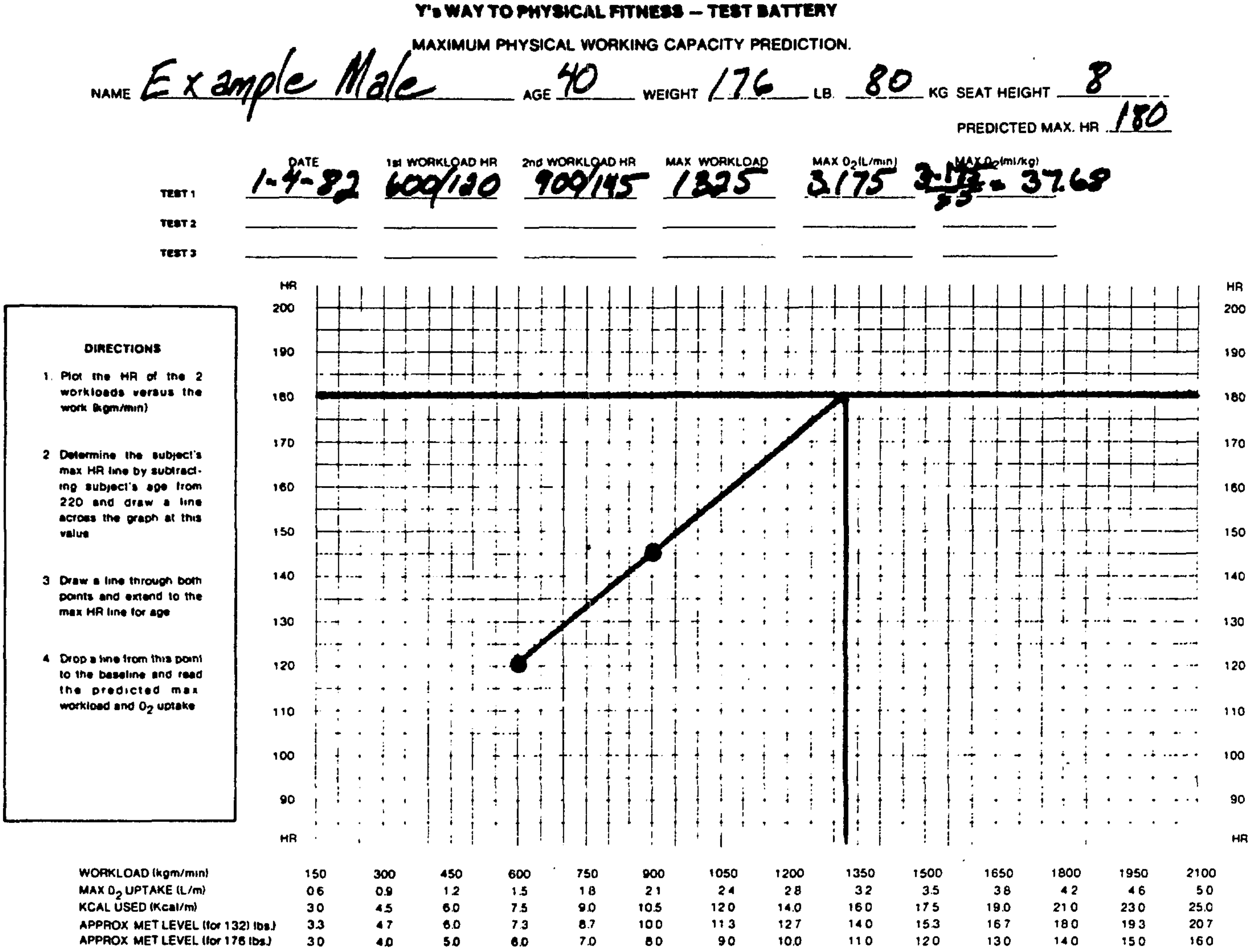
1. **Set the first workload to 150 kgm/min (.5 KP).**
2. **If steady-state heart rate is < 103, set 2nd load at 450 kgm/min (1.5 KP).**
   If steady-state heart rate is ≥ 103, set 2nd load at 300 kgm/min (1.0 KP).
3. **Follow this same pattern for setting the third and final load.**
4. **NOTE: If the 1st workload elicits a HR of 110 or more, it is used on the graph, and only ONE more workload will be necessary.**

Figure 4. Guide for setting power outputs for women on submaximal cycle ergometer tests. Reprinted from Golding, L.A., C.R. Myers, and W.E. Sinning, editors. *The Y's Way to Physical Fitness,* Revised (1982) with permission of the YMCA of the USA, 101 N. Wacker Drive, Chicago, IL 60606.

for men under age 35 years, the power output is increased by 24.5 and 49 W respectively every 2 minutes until the individual reaches a target heart rate of 70% of predicted HR max (220 – age). For men under age 35 years, the power output is increased by 24.5 W once they achieve 60% of predicted HR max, i.e., between 60% and 70% of HR max. Once the 70% of predicted HR max is achieved, the power output is maintained at this level until a steady-state heart rate is reached. $\dot{V}O_2$ max is then predicted (y) from the oxygen uptake using the Åstrand-Ryhming nomogram (X_1) and from age (X_2), as follows.

$$\text{Males: } y = 0.348\,(X_1) - 0.035\,(X_2) + 3.011$$
$$(R = 0.86,\ SEE = 0.359\ \text{liters}\cdot\text{min}^{-1})$$
$$\text{Females: } y = 0.302\,(X_1) - 0.019\,(X_2) + 1.593$$
$$(R = 0.97,\ SEE = 0.199\ \text{liters}\cdot\text{min}^{-1})$$

In addition to submaximal testing for estimating $\dot{V}O_2$max, it is also possible to use maximal tests. Several running tests have been proposed that have relatively high correlations with actual $\dot{V}O_2$max values. Balke (110) and Cooper (111) developed all-out walk-jog-run tests to determine the maximum distance that can be attained in a fixed time, i.e., 15 and 12 minutes, respectively. For administrative ease, these fixed-time tests were translated into fixed-distance tests of 1.0 and 1.5 miles, and the time taken to complete the distance was used as an estimate of $\dot{V}O_2$max. These tests are appropriate for children and young adults, but they are not advised for middle-aged and older adults, who may have underlying disease. Table 4 provides estimates of $\dot{V}O_2$max based on the time taken to complete 1.5 miles. Table 5 provides estimates of $\dot{V}O_2$max based on the



Y's WAY TO PHYSICAL FITNESS — TEST BATTERY

MAXIMUM PHYSICAL WORKING CAPACITY PREDICTION.

NAME Example Male AGE 40 WEIGHT 176 LB. 80 KG SEAT HEIGHT 8

PREDICTED MAX. HR 180

	DATE	1st WORKLOAD HR	2nd WORKLOAD HR	MAX WORKLOAD	MAX O2(L/min)	MAX O2(ml/kg)
TEST 1	1-4-82	600/120	900/145	1325	3.175	[illegible] = 37.68
TEST 2						
TEST 3						

DIRECTIONS

1. Plot the HR of the 2 workloads versus the work (kgm/min)
2. Determine the subject's max HR line by subtracting subject's age from 220 and draw a line across the graph at this value
3. Draw a line through both points and extend to the max HR line for age
4. Drop a line from this point to the baseline and read the predicted max workload and O_2 uptake

WORKLOAD (kgm/min)	150	300	450	600	750	900	1050	1200	1350	1500	1650	1800	1950	2100
MAX O_2 UPTAKE (L/m)	0.6	0.9	1.2	1.5	1.8	2.1	2.4	2.8	3.2	3.5	3.8	4.2	4.6	5.0
KCAL USED (Kcal/m)	3.0	4.5	6.0	7.5	9.0	10.5	12.0	14.0	16.0	17.5	19.0	21.0	23.0	25.0
APPROX MET LEVEL (for 132) lbs.)	3.3	4.7	6.0	7.3	8.7	10.0	11.3	12.7	14.0	15.3	16.7	18.0	19.3	20.7
APPROX MET LEVEL (for 176 lbs.)	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	15.0	16.0

Figure 5. Example of an estimation of $\dot{V}O_2$ max from heart rate and power output. Reprinted from Golding, L.A., C.R. Myers, and W.E. Sinning, editors. *The Y's Way to Physical Fitness,* Revised (1982) with permission of the YMCA of the USA, 101 N. Wacker Drive, Chicago, IL 60606.

distance completed in 12 minutes. $\dot{V}O_2$max can also be estimated from the length of time taken to complete a maximal treadmill test. Bruce et al. (112) reported a correlation of $r = 0.91$ between maximal treadmill time and $\dot{V}O_2$max. However, Froelicher et al. (89, 113, 114) have reported considerably lower correlations, generally below $r = 0.80$, and rather high standard errors of estimate, e.g., 2.6–4.7 $ml \cdot kg^{-1} \cdot min^{-1}$, when comparing time to exhaustion on the Bruce and Balke protocols with $\dot{V}O_2$max. Foster et al. (115) developed multiple-regression equations to estimate $\dot{V}O_2$max from treadmill time using the Bruce protocol. They reported correlations of $r = 0.98$ with standard errors of estimate of 3.1–3.3 $ml \cdot kg^{-1} \cdot min^{-1}$. Table 6 provides estimates of $\dot{V}O_2$max from maximal treadmill time on several different protocols.

Muscular strength and endurance

Strength is defined as the maximum force a muscle or a muscle group can produce voluntarily. The scientist, however, is likely to define strength very specifically to avoid ambiguity. This is illustrated by the definition of Atha (21), who states that strength is the ability to develop force against an unyielding resistance in a single contraction of unrestricted duration. Strength measured in this manner would be classified as isometric, or static, strength. Dynamic strength, although more difficult to control experimentally, is of considerable interest as it is that form of strength that governs daily activity. Muscular endurance refers to the ability of the muscle or muscle group to sustain a fixed or static contraction for an extended period of time. In the context of human movement, dynamic muscular endurance would relate to the ability of a muscle or muscle group to maintain a certain power output over time.

Criterion method

As defined previously, strength is measured as the maximum force the individual can apply statically under the conditions present at the time of testing. Thus, the assessment of static strength requires a device that will measure the force produced by the muscle or muscle group. In the past, several different types of devices have been used in the assessment of static strength, including spring-type devices, pressure devices, and electrical

Table 4. Estimation of $\dot{V}O_2$max from the 1.5-mile test

Time (min and sec)	Estimated VO_2max ml·kg × min
7:30 and under	75
7:31–8:00	72
8:01–8:30	67
8:31–9:00	62
9:01–9:30	58
9:31–10:00	55
10:01–10:30	52
10:31–11:00	49
11:01–11:30	46
11:31–12:00	44
12:01–12:30	41
12:31–13:00	39
13:01–13:30	37
13:31–14:00	36
14:01–14:30	34
14:31–15:00	33
15:01–15:30	31
15:31–16:00	30
16:01–16:30	28
16:31–17:00	27
17:01–17:30	26
17:31–18:00	25

Source: Wilmore, J.H. *Training for Sport and Activity: The Physiological Basis of the Conditioning Process.* 2nd Edition. Boston: Allyn and Bacon, Inc., 1982, p. 273.

devices. Stull and Clarke (116) published a comprehensive review of these three types of devices in 1978. The following discussion will borrow extensively from this excellent review and from the review of Hunsicker and Donnelly (117), which appeared in 1955.

Spring-type devices include dynamometers, spring balances and the cable tensiometer. According to the review of Hunsicker and Donnelly (117), the first spring-type dynamometer was developed by an English scientist by the name of Graham, and the first elliptical spring dynamometer of metal was developed in France by Regnier in 1807. Regnier's dynamometer is considered to be the prototype of the spring steel dynamometers used today. His dynamometer was used for measuring grip strength, back strength, and two-arm pulling strength (117). In the late 1800's, strength testing evolved as an important research and clinical procedure in the United States largely through the work of Dudley A. Sargent at Harvard University. At about this same time, Smedley, an American anthropologist, developed the adjustable hand-grip dynamometer, allowing the device to accommodate the size of the subject's hand (116).

A major advancement in strength testing occurred during the Second World War. Arising from a need to test the strength of individual muscle groups involved in orthopedic disabilities, Clarke (118) devised a series of strength tests utilizing a cable tensiometer. The cable tensiometer was originally designed to test the tension of aircraft control cables (118). Cable tension is determined by measuring the force applied to a riser that causes an offset in a cable stretched taut between two sectors. Refer to figure 6 for an illustration of the cable tensiometer. The original tensiometer was adapted for strength testing by special calibration for an "up-pull" on the cable rather than placement on a taut cable, by the addition of a pointer to facilitate reading of the subject's peak score, and by deletion of the brake lever rod (118). In addition to the tensiometer, equipment needs for cable tension testing include a special testing table, cables, a goniometer for adjusting joint angles, and special items for testing finger movements and supination and pronation of the forearm. After extensive testing with this device, 38 tests were eventually proposed to measure the strength of muscle groups over the following locations: Finger, thumb, wrist, forearm, elbow, shoulder, neck, trunk, hip, knee, and ankle. During the development of the test battery, it was necessary to determine the body position that permitted the greatest application of strength for each joint movement, select the joint angle that resulted in the strongest movement, and study such factors as the location of the pulling strap on the body part serving as the fulcrum and the effect of gravity on test scores (118). Figure 7 illustrates the use of the cable tensiometer in the assessment of hip extension.

Table 5. Estimation of $\dot{V}O_2$max from the 12-minute test

Distance (miles)	Laps (¼-mile track)	Maximal oxygen consumption (ml/kg/min)
<1.0	<4	[1]<25.0
1.000	4	[1]25.0
1.030	...	[1]26.0
1.065	4¼	27.0
1.090	...	28.2
1.125	4½	29.0
1.150	...	30.2
1.187	4¾	31.6
1.220	...	32.8
1.250	5	33.8
1.280	...	34.8
1.317	5¼	36.2
1.340	...	37.0
1.375	5½	38.2
1.400	...	39.2
1.437	5¾	40.4
1.470	...	41.6
1.500	6	42.6
1.530	...	43.8
1.565	6¼	45.0
1.590	...	46.0
1.625	6½	47.2
1.650	...	48.0
1.687	6¾	49.2
1.720	...	50.2
1.750	7	51.6
1.780	...	52.6
1.817	7¼	53.8
1.840	...	54.8
1.875	7½	56.0
1.900	...	57.0
1.937	7¾	58.2
1.970	...	59.2
2.000	8	60.2

[1] Insufficient data at these distances to make reliable comparisons.
Source: Cooper, K.H. A means of assessing maximal oxygen intake. *JAMA* 203:201–204, 1968.

Table 6. Estimation of $\dot{V}O_2$max from maximal treadmill time on several treadmill protocols

Fitness classification	Maximum O_2 uptake (ml/kg·min^{-1})	MET's [1]	Treadmill protocols						1.5-mile run (min and sec)
			Bruce [2]	Ellestad [2]	Balke [2, 3] (3.3 mph)	Balke [2, 4] (3.0 mph)	Naughton [2, 5]	Åstrand (mph)	
1	7	2	—	—	—	—	2:07	—	—
	10.5	3	—	—	1:00	3:00	4:17	—	—
	14	4	2:30	2:00	2:00	4:00	6:28	—	—
2	17.5	5	4:00	3:00	3:00	7:30	8:38	—	—
	21.0	6	6:00	4:45	6:00	10:30	10:49	—	—
	24.5	7	7:20	5:00	8:00	13:30	12:59	—	—
3	28.0	8	8:20	5:45	9:45	17:00	15:10	5:00	18:45
	31.5	9	9:15	6:40	12:00	19:30	17:20	5:25	16:30
	35.0	10	10:10	7:30	14:30	22:00	19:30	5:50	15:00
4	38.5	11	11:00	8:20	17:00	24:00	21:40	5.75	13:00
	42.0	12	12:00	9:10	19:00	27:00	23:51	6.25	12:00
5	45.5	13	12:45	10:15	21:30	30:00	26:01	6.50	11:00
	49.0	14	13:40	11:15	24:15	33:00	28:12	7.00	10:00
6	52.5	15	14:30	—	26:15	36:00	30:22	7.50	9:30
	56.0	16	15:15	—	27:45	—	32:33	8.00	9:00
7	59.5	17	16:10	—	29:00	—	—	8.50	8:15
	63.0	18	17:00	—	30:00	—	—	9.00	7:45
	66.5	19	18:00	—	31:15	—	—	9.25	7:15
8	70.0	20	19:20	—	32:00	—	—	9.75	6:52
	73.5	21	21:00	—	33:45	—	—	10.50	6:30
	77.0	22	22:30	—	35:45	—	—	11.00	6:10

[1] MET = metabolic equivalent above the resting metabolic level. Value at rest is approximately 3.5 ml/kg·min^{-1}.

[2] Data expressed in minutes and seconds of test protocol (duration) completed.

[3] Balke protocol, 3.3 mph, at 1-percent grade increase in work level per minute.

[4] Balke protocol, 3.0 mph, at 2.5-percent grade increase in work level every 3 minutes.

[5] Naughton protocol, modified to use 2-minute rather than 3-minute stages.

(Adapted with permission from Pollock, M.L., Wilmore, J.H., and Fox, S.M.: *Health and Fitness Through Physical Activity.* New York, John Wiley and Sons, 1978.)

Source: Pollock, M.L., J.H. Wilmore, and S.M. Fox, III, *Exercise in Health and Disease: Evaluation and Prescription for Prevention and Rehabilitation.* Philadelphia: W.B. Saunders, 1984, p. 163.

According to Stull and Clarke (116), spring-type devices have the advantage of being relatively inexpensive and simple to use. They do lack the precision of the more expensive transducers (to be discussed), and they require extensive calibration throughout the dynamic range of test scores encountered during testing. The stretching and bending qualities of metal may change with use, thus making frequent calibration essential. A further problem with spring-type devices is the inability

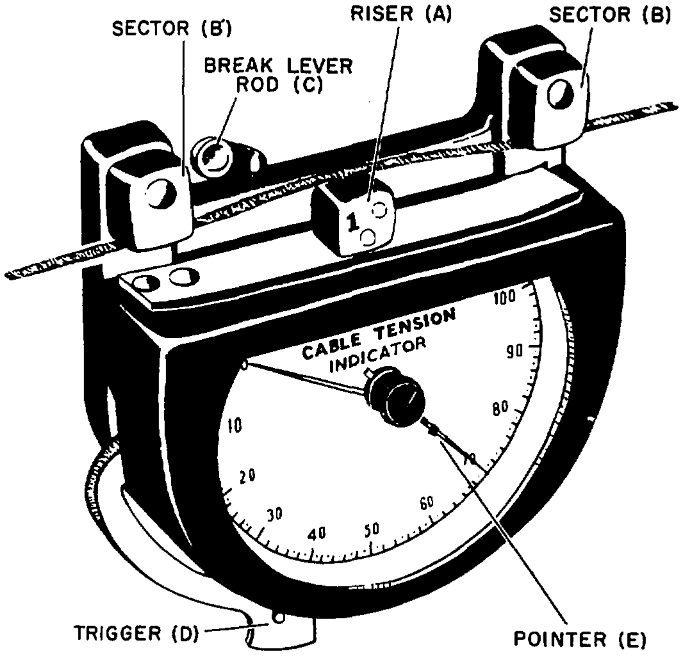


Figure 6. The cable tensiometer. From Clarke, H.H. *Muscular Strength and Endurance in Man.* Englewood Cliffs, NJ: Prentice Hall, Inc., 1966, p. 8.

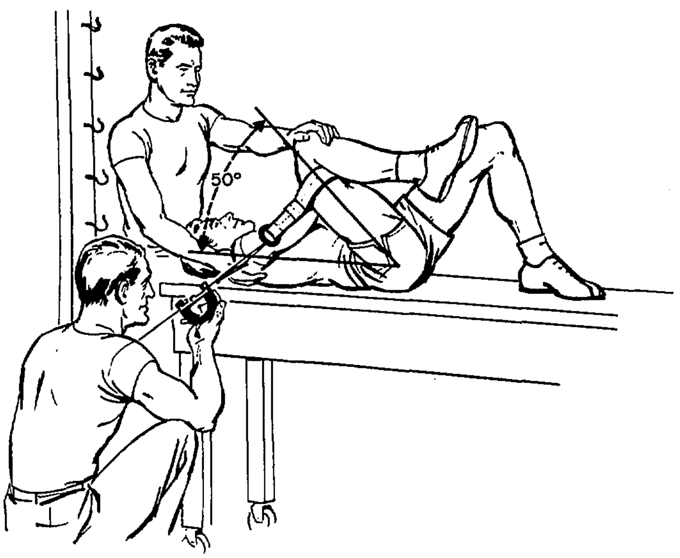


Figure 7. Use of the cable tensiometer to measure hip extension strength. From Clarke, H.H. *Muscular Strength and Endurance in Man.* Englewood Cliffs, N.J.: Prentice Hall, Inc., 1966, p. 27.

to alter their sensitivity, thus necessitating multiple units, each to measure small ranges of force development in the spectrum of forces measured from the weakest to the strongest individual and from the weakest to strongest muscle group.

Pressure devices have used mercury or air as the transducer of force when assessing strength (116). These devices are generally limited to the assessment of a specific movement; thus they do not have the versatility of other strength-testing devices. Consequently, they will not be discussed in further detail as they would not be appropriate for general fitness testing in a large population survey.

Electrical devices for assessing strength were first utilized by Ralston and his coworkers at the University of California, San Francisco, in 1948. They used wire strain gauges for measuring muscle tension in human muscle (116). With resistance wire strain gauges, as the wire is stretched in response to the applied force, the change in the wire's electrical resistance is directly proportional to the change in the length of the wire. For measuring applied forces, strain gauges are connected to a standard Wheatstone bridge circuit. According to Stull and Clarke (116), unmounted strain gauges are relatively inexpensive, they can yield highly accurate measures of strength, and calibration is highly linear providing the elastic limits of the supporting structure have not been violated. Sensitivity can be increased by altering the amount of electrical current applied to the bridge, greatly increasing the versatility of the device. Load cells are strain gauges that have been permanently mounted as a complete unit and are usually temperature compensated (116).

In addition to static strength testing, dynamic strength testing provides for a more functional assessment of strength. Dynamic testing can range from simple tests of the maximum amount of weight that can be lifted for any one movement (e.g., elbow flexion or two-arm biceps curl) just one time to the use of sophisticated laboratory devices that allow the assessment of muscle torque at each point in the range of motion. The former, referred to as the one-repetition maximum test (1–RM), will be discussed in greater detail under "Additional methods." The development of more sophisticated systems for assessing muscle function started in the mid-1960's. James Perrine, a bioengineer from New York City, developed the first of what are now considered to be the new generation of strength-testing equipment (119). Using the principle of isokinetics, the device developed by Perrine allowed the assessment of the torque produced by the muscle or muscle group throughout the range of movement. With this device, a lever arm is attached to a hydraulic system so that, as torque is applied, the lever arm moves at a constant velocity. The hydraulic system prevents any additional movement of the overall system, thus maintaining a constant velocity, and the distortion of the lever arm translates into a distortion or unbalancing of a Wheatstone bridge (116). The torque curve and the respective angle through which the movement has occurred is displayed simultaneously on a strip chart recorder or can be relayed directly to a computer. The Perrine invention is now manufactured by the Lumex Company in the State of New York and sells under the trade name Cybex.

With the isokinetic testing device, it is possible to measure strength at various angles in the range of motion and at various velocities from 0 (static) up to 300 degrees·sec^{-1}. Figure 8 illustrates the variation in strength relative to the angle of contraction, with 100 percent representing the angle at which strength is optimal. The isokinetic testing device allows the assessment of torque at each of a specific number of points throughout the range of motion with a single maximal contraction. Obviously, several trials would be important to determine the most representative trial. By assessing the muscle at several different speeds, it is possible to estimate the muscle fiber composition of the contracting muscles on the basis of their differential response to different speeds and to a series of repeated contractions (24). Thus, with such a device, it is possible to determine both static and dynamic strength at multiple angles in the range of motion and at multiple speeds.

The assessment of muscular endurance can also be accomplished with an isokinetic testing device. Although a number of protocols have been reported in the research literature, each protocol specifies a time period and rate of contraction, and then defines muscular endurance by any one of a number of measurements that can be obtained from the resulting data. As an example, a popular test protocol for assessing knee extensor

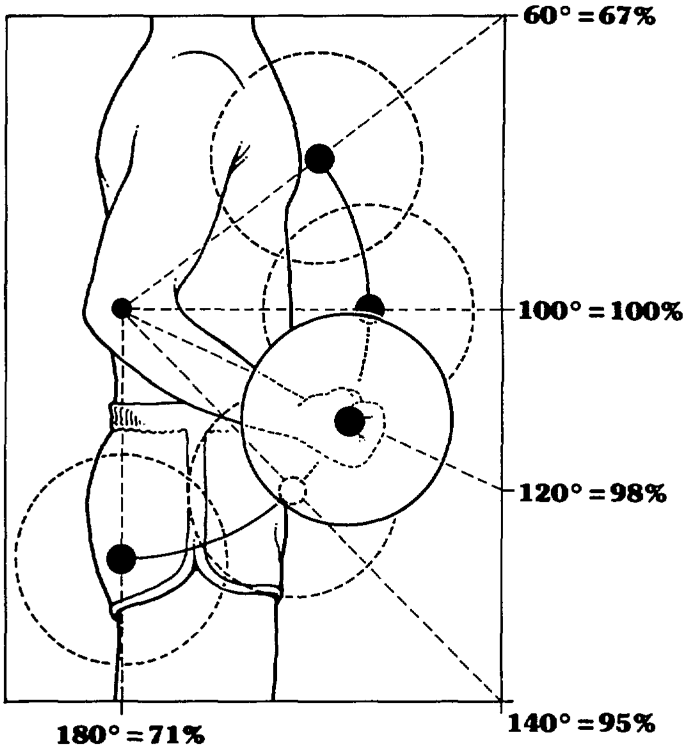


Figure 8. Variation in strength relative to the angle of contraction, with 100 percent representing the angle at which strength is optimal. From Wilmore, J.H. *Training for Sport and Activity: The Physiological Basis of the Conditioning Process.* 2nd Edition. Boston: Allyn and Bacon, 1982, p. 74.

endurance has the individual perform a series of maximal knee extensor contractions at a rate of one repetition every 2 seconds for a period of 60 seconds. Endurance can then be assessed by the total work performed over the 60-second time interval or by an index of work decrement from the first few repetitions to the last few repetitions. The latter measurement would be a more specific measure of endurance, because total work would be related to the individual's initial level of strength, and interpretation of endurance could then be made only by comparisons with individuals of similar initial strength. The use of a work decrement index provides an excellent index of fatigue, which is considered to be the reciprocal of endurance.

When selecting a test battery, it is important to recognize that strength is highly specific to the muscle group tested. It is possible to have considerable lower body strength but poor upper body strength; equally possible to have good upper body strength but poor lower body strength; and even possible to have excellent upper and lower body strength but poor abdominal strength. Thus, as stated in the introduction, strength assessment in at least three areas, i.e., upper body, trunk, and lower body, would seem essential to approximate an estimate of total body strength for any national population survey. Similar research has not been conducted regarding the degree of specificity of muscular endurance. It is quite possible that muscle endurance is also highly specific to the individual muscle group tested.

When administering tests of strength and muscular endurance, it is critical to standardize testing procedures and to control, as well as possible, the psychological and the physiological environment. The position of the subject must be carefully standardized, for a slight shift in the subject's posture could radically change the mechanical advantage in the performance of a given task. Positioning of the hands, arms, and legs is very important. As an example, when performing a knee extension test with the Cybex isokinetic testing device, there is a marked difference in the results if the subject is allowed to grab the side rails and pull with his or her arms during the performance of that test, as opposed to performing the test with the arms folded across the chest. Careful attention must be given to every detail of posture control if the results are to have any meaning.

With respect to the environment, the test must be conducted in a room controlled for light, temperature, and humidity. Further, the psychological state of the individual must be controlled. Ikai and Steinhaus (120) reported that test results can vary considerably according to the psychological state of the individual. Having the subject shout at the time of performing the test, shooting a blank pistol unexpectedly several seconds before the strength test, and hypnosis all substantially increased voluntary strength scores. Even factors such as the personality and sex of the technician administering the test can influence the obtained values.

Although these criterion measures can be administered reliably in a standardized manner, it is important to recognize that, at the present time, there are no accepted standards with which to compare obtained values. This will be a major limitation for the interpretation of any resulting data from a national survey. Yet, it points to the importance of including such a test battery in a national survey in order to obtain objective national standards.

Additional methods

Isometric strength assessment is basically a laboratory test that requires equipment and testing procedures similar to those previously outlined. Thus, additional methods of isometric strength assessment designed for field testing are not presently available. There are, however, additional field methods for assessing dynamic strength and muscular endurance. These will be briefly discussed.

Strength can easily be assessed by the one-repetition maximum test (1–RM). For any given lift, e.g., bench press, leg press, or two-arm curl, the individual is given a series of trials to determine the greatest weight that he or she can lift just once. If the individual is inexperienced in lifting weights, this test is conducted largely by trial and error. Start with a weight that the individual can lift comfortably, and then add weight for each subsequent lift until the individual can lift the weight correctly just one time. If this weight can be lifted more than once, more weight should be added until a true 1–RM is established. The 1–RM assessment can be obtained for any one of a number of basic weight training exercises. Test batteries, however, usually select three or four exercises that represent the body's major muscle groups (9). Table 7 provides what have been considered to be optimal strength values on the basis of body weight for young adults for four standard lifts: Bench press, standing press, curl, and leg press. These values were derived to serve as guidelines, not absolute standards, because the data necessary for deriving absolute standards are not available.

Muscular endurance has been measured in a number of different ways, including the greatest number of situps that can be performed in a fixed period of time, e.g., 30 seconds or 1 minute; the maximum number of pushups, pullups, or bar-dips that can be performed; or the length of time that one can sustain a flexed arm hang. Most of these tests are a part of the AAHPERD Youth Fitness Test (6) and the President's Council on Physical Fitness and Sports' Youth Physical Fitness Test (7). Many of these tests penalize the participant who has long legs, short arms, or a heavy body weight. As a result, Wilmore has suggested a simple test of endurance that attempts to standardize for these factors (9). It is suggested that the individual lift a fixed percentage of his or her maximum strength for any given lift, and that this fixed percentage be constant for all lifts. The percentage of 70% of the individual's 1–RM was suggested. The individual then lifts this weight as many times as possible until reaching the point of fatigue or exhaustion. Because this is a relatively new concept, norms or standards have not been established.

Table 7. Optimal strength values for various body weights based on the 1-repetition maximum test

Body weight (lb)	Bench press		Standing press		Curl		Leg press	
	Male	Female	Male	Female	Male	Female	Male	Female
80	80	56	53	37	40	28	160	112
100	100	70	67	47	50	35	200	140
120	120	84	80	56	60	42	240	168
140	140	98	93	65	70	49	280	196
160	160	112	107	75	80	56	320	224
180	180	126	120	84	90	63	360	252
200	200	140	133	93	100	70	400	280
220	220	154	147	103	110	77	440	308
240	240	168	160	112	120	84	480	336

Source: Wilmore, J.H. *Training for Sport and Activity: The Physiological Basis of the Conditioning Process.* 2nd Edition. Boston: Allyn and Bacon, Inc., 1982, p. 270.

Flexibility

Flexibility, as was stated in the introduction, refers to the range of motion, or to the looseness or suppleness of the body or specific joints, and reflects the interrelationships among muscles, tendons, ligaments, skin, and the joint itself (121). DeVries (122) cites a number of factors in addition to the anatomical factors listed previously that influence the flexibility of joints. First, the more active the individual, generally the more flexible he or she is. Activity that is performed through a limited range of motion, however, can actually decrease flexibility. The long distance runner who does little or no stretching usually has limited or poor flexibility because of the fact that running does not utilize the full range of movement. Thus, activity is beneficial only as long as it emphasizes full range of motion. Second, females tend to be more flexible than males. Third, flexibility is age related, decreasing from birth to the age of 10–12 years and then increasing into young adulthood. With aging, however, flexibility begins a gradual decline. Finally, flexibility is greatly influenced by temperature. Local warming of a joint will increase the flexibility of that joint, whereas cooling will bring about the opposite effect.

Kraus and Raab (123) reported that muscle weakness and poor flexibility were largely responsible for chronic back pain in nearly 80% of those patients afflicted with this disability. Thus, the assessment of flexibility in a large population survey would appear to be an important part of a general fitness assessment battery.

Criterion method

Flexibility of individual joints is measured directly by goniometry. In its simplest form, a protractorlike device can be used to measure the angle of the joint at both extremes in the range of motion, with the difference between these two joint angles representing the extent of joint movement. The amount of soft tissue around the joint greatly influences the accuracy of measurement because the axis of the bones forming any one joint are not readily visible. This introduces a certain amount of subjectivity into the measurement technique.

More accurate measurements of flexibility can be made using two instruments devised specifically for flexibility assessment. The Leighton flexometer, illustrated in figure 9, consists of a weighted 360-degree dial and a weighted pointer mounted in a case. The dial and pointer move independently and are both controlled by gravity. Both can be locked in position independently of each other. The segment is usually positioned at one extreme in the range of motion, the dial locked in position, and then the segment moved through the full range of motion. The pointer follows the movement of the segment, thus indicating the extent of joint movement in degrees (figure 10). Because the length of limbs or segments is not a factor in this assessment, the device provides a more accurate estimate of flexibility than most other instruments or field tests (9).

A second device developed for measuring flexibility is the electrogoniometer, or the ELGON. The ELGON is a protractorlike device in which the protractor has been replaced by a potentiometer. The potentiometer provides an electrical signal that is directly proportional to the angle of the joint. This device can give continuous recordings during a variety of activities. The versatility of this unit allows a much more accurate and realistic as-

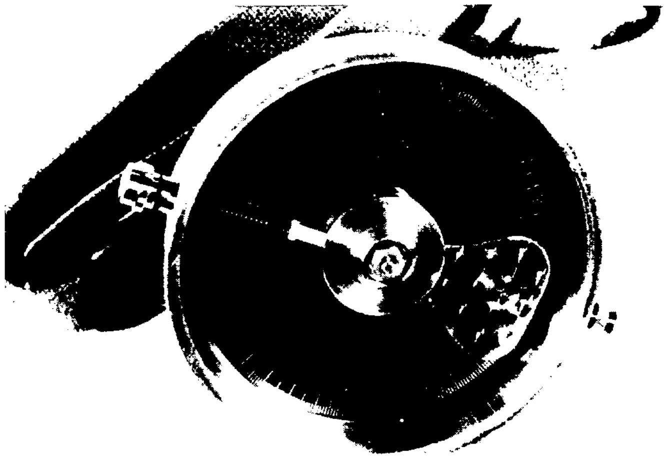


Figure 9. Illustration of the Leighton flexometer. From Montoye, H.J. *An Introduction to Measurement in Physical Education.* Volume 4. Indianapolis, IN: Phi Epsilon Kappa Fraternity, 1970, p. 94.

sessment of functional flexibility, i.e., the degree of flexibility exhibited during an actual physical activity (9).

Flexibility, similar to strength, is highly specific to the joint measured, as was reviewed in the introduction. Thus, for a national survey, it will be necessary to define those joints or composite movements that are considered vital to health for inclusion in a test battery.

Additional methods

A number of field tests have been proposed for assessing flexibility. Probably the most popular test is the sit-and-reach test. With the sit-and-reach test, the individual sits on the floor with the feet extended forward in front of him or her, the feet pressed flat against a box that supports the measuring device. With the back of the knees pressed flat against the floor, the individual leans forward and extends the fingertips as far forward as possible. The distance reached is recorded and serves as an approximation of one's flexibility at the hips. Although tests such as the sit-and-reach test are adequate for providing a crude index of flexibility for mass screening, these tests do not allow for differences in limb length or proportional differences between the legs and arms. The individual with long arms and short legs will attain a good score on the sit-and-reach test even with limited or poor flexibility. Likewise, the individual with short arms and long legs will be penalized and find it nearly impossible to attain a good score. The sit-and-reach test just described has been modified to correct at least partially the problem of limb length disproportionality. The test is started with the subject seated with the back against the wall, and the legs extended forward. An initial measure is obtained with the back pressed firmly against the wall and the arms extended forward. The individual is then tested in the standard forward stretched position, and the difference between the two scores is an index of flexibility at the hips, i.e., lower back and hamstrings.

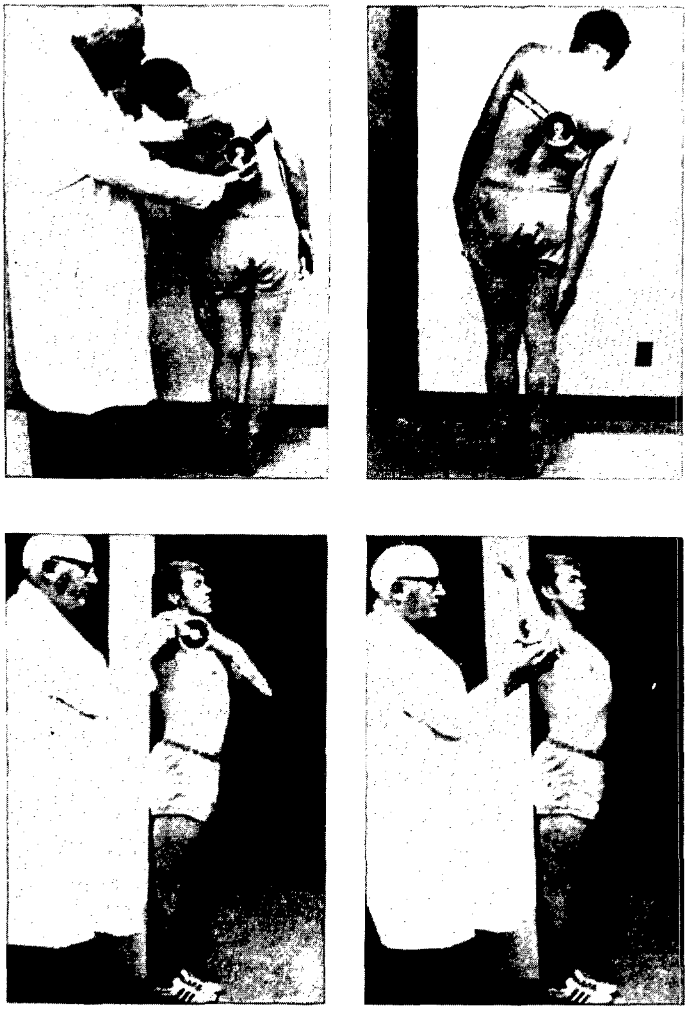


Figure 10. Illustration of the Leighton flexometer being used to test trunk lateral flexion (upper figures) and arm rotation (lower figures). From Montoye, H.J. *An Introduction to Measurement in Physical Education.* Volume 4. Indianapolis, IN: Phi Epsilon Kappa Fraternity, 1970, p. 101.

Body composition

The assessment of body composition presents somewhat of a dilemma because there has been considerable controversy recently regarding the accuracy of what has been accepted as the criterion method, i.e., hydrostatic weighing. From the 1930's until the mid-1970's, there was general acceptance of the hydrostatic weighing technique as the most accurate technique available. Although others had questioned certain aspects of this technique earlier, the papers of Bakker and Struikenkamp (124) in 1977 and Lohman (35) in 1986 focused attention on the actual magnitude of error associated with the hydrostatic weighing technique. In fact, the measurement error associated with this criterion technique is nearly of the same magnitude as the standard error of estimate (SEE) of several techniques that have been evaluated against the hydrostatic weighing technique. Siri (125) has estimated the SEE associated with variability in the density and proportional composition of the lean body mass to be approximately 3.5%. Bakker and Struikenkamp (124) and Lohman (35) estimate that this error may be lower, i.e., 2.7% fat in a chemically mature adult population. The SEE associated with anthropometric prediction of relative body fat under the best of conditions is approximately 3.5% to 3.9% (126).

Criterion method

Despite its limitations, the hydrostatic weighing technique is still considered the best of available methods for determining body composition. The subject normally arrives for testing in a postabsorptive state, with the bowel and bladder freshly evacuated. Body weight is first obtained by weighing the subject nude or while wearing a light bathing suit. The subject is then immersed in a body of water up to the level of the chin or chest and is instructed to wipe his or her body free of air bubbles, including bubbles that might be trapped in the bathing suit. The subject then immerses his or her head to wet the hair, and the hair is rubbed to remove all air bubbles. The subject is next asked to sit in a chairlike structure that is supported by a scale hanging directly above the subject. This chair must hang freely from the scale and must not touch the sides or bottom of the water container when the subject is submerged at the time of the underwater weighing. The subject is then asked to bend forward at the waist until the head is totally immersed under water while at the same time blowing all air out of the lungs, i.e., exhale to the point of the residual lung volume. The weight of the subject while immersed is taken at that point when he or she reaches full expiration. This procedure is repeated a minimum of 9 trials for a total of 10 trials. The underwater weight is obtained to within ± 10–20 grams using either an autopsy scale with 9 kilogram capacity and 10-gram divisions (33) or an electronic load cell, as described by Fahey (127). Although the load cell is expensive, it acts to dampen the oscillations that occur while the subject is under water, and it provides both a digital readout and a strip chart recording of the underwater weight. The underwater weighing can be conducted in a specially designed tank, e.g., Plexiglas; a standard redwood hot tub with a vinyl liner; or a swimming pool.

The final underwater weight used in the calculation of body density can be determined by one of two established criteria. Katch (128) proposed averaging the last few trials of a total of 10 trials in order to approximate the most dependable underwater weight, although he found that 40% of his sample of 135 men and women failed to achieve a leveling off in their values by the 10th trial. The first few trials involve learning a novel task, i.e., expelling all of the air from the lungs while totally submerged under water. It is only after the first few trials that the subject starts to understand what is expected and is able to achieve a maximal expiration to the end point of residual volume. Thus, the underwater weight increases progressively until the subject is able to achieve maximal expiration. Behnke and Wilmore (33) proposed a different method for selecting the most representative underwater weight. They proposed the selection of (a) the highest obtained weight if it is observed more than twice, i.e., three times or more; (b) the second highest weight if it is observed more than once and if the first criterion is not attained; and (c) the third highest weight if neither of the first two criteria are met. This procedure was used to reduce the possibility of underestimating the actual underwater weight for subjects who achieve their highest values during the first five to seven trials.

The obtained underwater weight must be corrected for two trapped air volumes, i.e., the residual volume and the volume of gas in the gastrointestinal tract. The latter volume is nearly impossible to determine and is usually a relatively small volume. Consequently, investigators either ignore this volume, or they use an assumed value of 100 ml, as originally proposed by Buskirk (129). The residual volume can be measured either in the water at the time of the underwater weighing or out of water either before or after the underwater weighing. Assessment in water is desirable because it measures the actual volume of air remaining in the lungs at the time of the underwater weighing. However, it is much easier to measure subjects while out of the water, and there is no electrical hazard (such as subject interface with electronic gas analyzer) when measurements are conducted with the subject out of water. Studies have been conducted comparing residual volume determined in the water versus out of the water. The results of these studies are equivocal, with some showing no difference, some showing higher residual volumes out of water, and some showing higher residual volumes in water. (See Girandola et al. (130) for specific references.) When differences have been demonstrated, they are typically less than 300 ml, which would have a relatively small influence on the calculated body density and relative body fat, i.e., <3% relative fat. It is also possible to conduct the underwater weighing at the functional residual volume (131), at 50% of the vital capacity (132), and at total lung capacity (133, 134). For some, holding the breath for 10–15 seconds under water at the point of residual lung volume is uncomfortable; thus, these alternative procedures may find greater subject acceptability.

Additional methods

A number of methods have been proposed for assessing body composition, and these can be classified as either laboratory or field methods. The field methods use anthropometric measurements, i.e., girths, diameters, and skinfolds, in the estimation of body density, lean body weight, and absolute and relative body fat. These will be discussed later in this section. The laboratory methods are categorized as follows (135):

Fat and fat-free body	*Muscle*	*Bone*
Densitometry	Spectrometry (^{40}K)	Photon absorptiometry
Hydrometry	Ultrasonics	Radiographics
Spectrometry (^{40}K)	Radiographics	Neutron activation
Ultrasonics	Neutron activation	Nuclear magnetic resonance
Radiographics	Nuclear magnetic resonance	Computer tomography
Electrical conductivity	Computer tomography	
Neutron activation	Creatinine excretion	
Nuclear magnetic resonance	Serum creatinine	
Computer tomography	Urinary 3-methylhistidine	

Hydrometry, or the assessment of total body water, involves isotopic dilution using tracers such as deuterium oxide or tritium oxide. Ethanol and antipyrine have also been used as tracers. The biological constant for water in the lean body mass (fat-free body) has been established as 73.2% of the lean tissue. Thus, lean body mass = (total body water)/0.732. The technique is based on the assumption that there is a uniform distribution of the isotope or tracer throughout the body fluids. The tracer can be either ingested or injected, and samples can be obtained from the blood, urine, or both. Hydrometry in combination with densitometry has been shown to reduce the error associated with densitometry alone (125).

Spectrometry requires the use of a whole body counter to measure the gamma radiation from ^{40}K, the radioactive isotope that is present in small quantities in the human body. The body's ^{40}K content is directly proportional to the total body potassium, which, in turn, constitutes a relatively stable proportion of the lean body mass. The liquid scintillation counter is used most frequently to assess ^{40}K. The subject is placed horizontally in the well of the counter and is surrounded by a layer of liquid scintillator solution. This solution converts the photon energies of the gamma rays from ^{40}K into light impulses. These light impulses, or scintillations, are detected by photomultiplier tubes mounted around the outer wall of the chamber, and the scintillations are amplified and converted into proportional voltages (33).

Ultrasonics utilize high frequency sound waves to differentiate between tissue types. Sound waves, generated by a special transducer, pass into the tissue, and when a change in density is encountered, a portion of these waves is reflected, sensed, and converted to electrical impulses that are then amplified and recorded. The thickness and density of the tissue through which the ultrasonic waves pass determine the characteristics of the reflected sound wave (33).

Radiographic analysis of body composition uses soft tissue x ray to differentiate among the various layers of skin, fat, bone, and muscle. This technique has largely been confined to the upper arm, but the measurements for this local area have been reported to be highly correlated to total body composition, i.e., $r = 0.89$ and SEE = 2.54% (136). Using the upper arm in a horizontal position, the x ray is taken at a KV of 76; an exposure time of 1/30th s, 300 MA; and a focal length of 72 inches. The total radiation is 10 mR.

Measurement of the electrical conductivity of the body has been proposed as a method for estimating lean body mass because of the difference in electrical conductivity between lean and fat tissue. Two instruments were introduced during the early 1980's that assess the body's electrical conductivity, i.e., TOBEC (total body electrical conductivity) and a bioelectrical impedance device. With TOBEC, the individual is placed inside a large poleroidal coil, and a small radio-frequency current is passed through the body. This procedure takes less than 1 minute per assessment, and there is minimal risk and no subject discomfort. Initial studies have indicated excellent agreement with densitometry and spectrometry (137, 138). With the bioelectric impedance device, four electrodes, two on the arm and two on the foot, are attached to the skin, and a radio frequency signal is introduced into the deep tissues of the body (800 uA at 50 kHz). Whole body impedance is supposedly closely related to total body water. Considerable research has been conducted to determine its validity in the assessment of total body composition (135). The most recent reports suggest that accuracy varies considerably, with SEE of from 2.7% (139) to 6.1% (138).

Neutron activation analysis is a relatively new technique that is based on the observation that fast neutrons are captured by elements in the body and unstable isotopes are produced that revert to a stable condition following emission of one or more gamma rays. The body becomes temporarily radioactive, and the emissions are recorded in a whole body counter. Total body nitrogen and total body potassium are estimated. At this time, neutron activation analysis would have to be considered an experimental technique but one that has considerable promise for the future.

Nuclear magnetic resonance (NMR) imagery is clearly a technique of the future for determining regional and total body composition. In this method, the body is exposed to a static magnetic field that affects the rotation of the nucleus of atoms with an odd number of protons or neutrons. The body is then exposed to an alternating magnetic field of the same frequency. Measurement of one or more parameters of these nuclei, after removing the attenuating magnetic field, enables the formation of body images (135). This method, although very expensive, has considerable promise for detailed body composition analyses.

Computed tomography is an x-ray scanning technique that provides cross-sectional images. With respect to body composition assessment, computed tomography has been used to assess the composition of the limbs at the level of the midthigh and mid-upper arm and for the entire forearm (140). The relationship between local scans and total body composition has yet to be determined.

Photon absorptiometry has been used to determine the mineral content of a cross-section of bone. Using a ^{131}I source, a beam of photons is passed over the forearm, and the amount that passes through is inversely proportional to the bone mineral content of the underlying bone (135). Mazess et al. (141) have recently reported data indicating that dual-photon absorptiometry (^{153}Gd) provides an excellent estimate of total body composition. Although these results are provocative and encouraging, only 18 subjects were evaluated, 14 females and 4 males with an age range from 23 to 61 years. Considerably more research will be necessary to determine the accuracy of this technique.

Creatinine excretion and total plasma creatinine have been proposed as indexes of muscle mass. Likewise, 3-methylhistidine has been proposed as an index of total muscle mass. Although these markers appear to be accurate indexes of total body muscle, it is less clear how applicable these techniques would be in the determination of total body composition.

The field methods for estimating body density, lean body weight, and absolute and relative body fat use girths, bony diameters, skinfolds, or a combination of these measures. Brozek and Keys (142) were the first to estimate body composition from anthropometric variables, i.e., skinfolds, in 1951. In the 1960's and 1970's, a number of investigators developed equations using multiple-regression techniques to estimate body composition parameters from a number of anthropometric variables (143). Typically, investigators measured a number of anthropometric variables on a selected population (e.g., college-age males, college-age females, wrestlers) and then, using the hydrostatic weighing technique as the criterion, entered the data into the computer and derived multiple-regression equations using stepwise, multiple-regression procedures. These studies resulted in the generation of a large number of equations that were later found to be population specific, i.e., they maintained prediction accuracy only when used on populations similar to the population from which the equation was derived (38). This created considerable confusion among clinicians and practitioners who were faced with the necessity of selecting an equation that best fit the characteristics of the population to be evaluated. In addition, the finding that many of the previously established equations were population specific led to a further proliferation of new equations for specific populations.

In 1974, Durnin and Womersley (144) published what is now considered to be the first generalized equation for predicting body density from anthropometric measures. The equations developed were applicable to a broader segment of the population, and, by maintaining a constant slope but varying the intercept of the equation, they were able to account for variations in the age of the population under study. The original work of Durnin and Womersley was later expanded by the studies of Jackson and Pollock (39) and Jackson et al. (40), providing generalized equations for men and women, respectively. It is now recognized that the relationship between the sum of skinfolds and total body density is curvilinear, and not linear, as was assumed in the original equations derived by linear multiple-regression techniques. Although linear equations demonstrate correlations with hydrostatically determined body density similar to those demonstrated by generalized quadratic equations, the SEE for the quadratic equations is generally much lower (145). This minimizes large prediction errors that occur at the extremes of the body density distribution when using linear regression equations.

In a recent review, Pollock and Jackson (145) stated that the sum of several skinfolds provides the most representative sample of subcutaneous body fat and is more highly correlated with body density than are individual sites. The correlations among skinfold sites are generally ≥0.90, and the correlations among various combinations of sums of three skinfold sites typically are >0.97 (145). Consequently, Pollock and Jackson (145) recommend that the sum of three or more skinfold sites be used in the estimation of body density from anthropometric measures. The specific equations developed by Jackson and Pollock (39) and Jackson et al. (40) are presented in table 8. From this table, it is apparent that it is not necessary to use a total of seven skinfold sites, as the correlations and SEE associated with the three skinfold site equations are nearly identical.

Table 8. Generalized regression equations for predicting body density for adult females (40) and males (39)

Variable	Regression equation	*r*	SE(BD)	SE(%F)
	Adult females			
∈ 7, Age	$\text{BD} = 1.0970 - 0.00046971(X_1) + 0.00000056(X_1)^2 - 0.00012828(X_6)$	0.85	0.008	3.8
∈ 3, Age	$\text{BD} = 1.0994921 - 0.0009929(X_2) + 0.0000023(X_2)^2 - 0.0001392(X_6)$	0.84	0.009	3.9
∈ 3, Age	$\text{BD} = 1.0902369 - 0.0009379(X_5) + 0.0000026(X_5)^2 - 0.0001087(X_6)$	0.84	0.009	3.9
	Adult males			
∈ 7, Age	$\text{BD} = 1.11200000 - 0.00043499(X_1) + 0.00000055(X_1)^2 - 0.00028826(X_6)$	0.90	0.008	3.5
∈ 3, Age	$\text{BD} = 1.1093800 - 0.0008267(X_3) + 0.0000016(X_3)^2 - 0.0002574(X_6)$	0.91	0.008	3.4
∈ 3, Age	$\text{BD} = 1.1125025 - 0.0013125(X_4) + 0.0000055(X_4)^2 - 0.0002440(X_6)$	0.89	0.008	3.6

Note: X_1 = sum of seven skinfolds (mm); X_2 = sum of triceps, suprailium, and thigh skinfolds (mm); X_3 = sum of chest, abdomen, and thigh skinfolds (mm); X_4 = sum of chest, triceps, and subscapular skinfolds (mm); X_5 = sum of triceps, suprailium, and abdomen skinfolds (mm); X_6 = age in years.
Source: Pollock, M.L., D.H. Schmidt, and A.S. Jackson. Measurement of cardiorespiratory fitness and body composition in the clinical setting. *Comprehensive Therapy* 6:12–27, 1980.

The estimation of relative body fat from body density provides a clear distinction between overweight and obesity. This distinction is important, as obesity is clearly the factor more closely related to health risk. Recently, results from several studies have indicated that the pattern of adipose tissue distribution is highly related to increased risk for cardiovascular disease, hypertriglyceridemia, hyperinsulinemia, and glucose intolerance (146, 147, 148). Those with a more masculine type of adipose tissue distribution, i.e., a high ratio of waist-to-hip circumference, are at higher risk. Thus, it would be important to consider the inclusion of simple circumferential measurements, i.e., hip and waist, in any large national survey.

Certain methodological considerations must be addressed when taking anthropometric measurements. These were clearly identified in a recent study by Lohman et al. (149), who observed the effect of skinfold caliper, investigator technique, and various skinfold prediction equations on estimates of body fat in female college basketball players. The study demonstrated that, even with experienced investigators, there could be differences in the actual location of any one specific site, even though the investigators were using the same site description. Further, there were considerable differences among four different brands of skinfold calipers for the skinfold thickness obtained at any one given site, with the Lange caliper providing values ranging from 14% to 38% higher than the caliper providing the next closest values.

Finally, attempts have been made over the past 2 years to improve the accuracy of anthropometric measurements by establishing standardized measurement sites for the most commonly obtained skinfolds, girths, and widths. The National Institutes of Health sponsored a standardization conference that was held in the summer of 1985. A book is to be published in the winter of 1988 that summarizes the consensus of those attending this conference with respect to each individual measurement site. A thorough description, a line drawing, and a photograph will be provided for each site.

Summary and recommendations

The body of this paper has provided a broad overview of design issues and alternatives in assessing physical fitness among apparently healthy individuals in a health examination survey of the general population. This paper reviewed previous definitions of physical fitness, reviewed those components that constitute the entity health-related physical fitness, and then presented an overview of the state of the art relative to the assessment of each of these components of physical fitness.

For the purpose of designing a physical fitness test battery for a national examination survey of the general adult population, it is recommended that the concept of health-related physical fitness be accepted. Developed and adopted by the American Alliance for Health, Physical Education, Recreation, and Dance, health-related physical fitness is defined as: (a) The ability to perform strenuous physical activity with vigor and without excessive fatigue and (b) a demonstration of physical activity traits and capacities that are consistent with minimal risk of developing hypokinetic diseases, i.e., diseases associated with physical inactivity. Health-related physical fitness includes the components of cardiorespiratory endurance, muscular strength and endurance, body composition, and flexibility. Previous tests of physical fitness included the additional components of agility, power, speed, and balance, which are collectively referred to as components of motor fitness. Although motor fitness is an important concept for the growing child, the aging adult, and the athlete at all ages, it should not be considered as a part of a health-related physical fitness test battery for the general population. Therefore, it is recommended that the National Health and Nutrition Examination Survey program within NCHS limit fitness assessment to those components considered a vital part of health-related physical fitness.

In accordance with this, the physical fitness test battery should be designed to assess all four components comprising health-related physical fitness. The following recommendations are made relative to each of these four specific components.

Cardiorespiratory endurance

The criterion measure of cardiorespiratory endurance capacity in the research laboratory is the maximal oxygen uptake ($\dot{V}O_2$max). Although indirect estimates of $\dot{V}O_2$max are possible with certain field tests, e.g., 1.5-mile and 1.0-mile runs, with reasonable correlations and standard errors of estimate, an all-out maximal field test would not be appropriate for a large population survey including individuals of all ages and with varying states of health. The estimation of $\dot{V}O_2$max from submaximal tests is a second alternative, but the accuracy of submaximal tests is greatly limited by the uncertainty of the accuracy of the assumed maximal values to which the submaximal values are extrapolated, e.g., extrapolation of submaximal heart rate values to age-predicted maximal heart rate. The correlations of $\dot{V}O_2$max estimated from submaximal tests versus actual $\dot{V}O_2$max are within an acceptable range (i.e., r = 0.80 to 0.93), but the standard errors of estimate are relatively large. Therefore, it is recommended that a symptom- and sign-limited test to volitional fatigue be used, during which a 12-lead electrocardiogram is obtained and oxygen uptake is measured. The test protocol should be designed to allow each subject to exercise for not less than 6 minutes but not more than 15 minutes. Preparation time, test administration time, and recovery time should not exceed a total of 25 to 30 minutes per person. Either a semiautomated or fully automated system should be employed for obtaining metabolic measurements. Further, backup components should be on site in case of equipment failure. If the direct measurement of metabolic function is not considered feasible, it is recommended that a standardized protocol be utilized allowing the estimation of $\dot{V}O_2max$ from the time on the treadmill test.

Strength and muscular endurance

Strength and muscular endurance can be assessed directly, either by simple field tests such as the one-repetition maximum or by using sophisticated equipment such as the Cybex II isokinetic testing device. It is recommended that sophisticated equipment be used to assess static, or isometric, and dynamic strength of the upper and lower extremities. Further, it is recommended that the bench press movement be adopted to represent upper body strength and the leg press (extension) be used to represent lower body strength. Three measurements of static strength could be obtained initially, followed by a 60-second all-out dynamic exercise at 30 contractions over the 60 seconds (1 contraction/2 seconds). Averaging the first five contractions could provide an estimate of dynamic strength, and the decrement in strength between the average of the first five and the average of the last five contractions could provide an index of muscle endurance. This test battery, conducted for both upper and lower extremities, should not take more than 15 minutes total time.

Flexibility

The measurement of flexibility is considered essential to a health-related physical fitness test battery. Unfortunately, there is no universal agreement on what constitutes an adequate flexibility test battery. Because flexibility is highly specific to the joint being evaluated, it is important to select carefully those joints to be evaluated. It is recommended that this area be assigned a high priority for future research, to establish both sites of measurement and methods of measurement. Although instruments such as the Leighton flexometer and the ELGON appear to have promise, the author of this position paper is not sufficiently knowledgeable in this area to make an educated recommendation.

Body composition

The assessment of body composition is also critical to a health-related physical fitness test battery. Although the hydrostatic, or underwater, weighing technique is considered to be the criterion method, the recent criticism of the accuracy of this method of certain populations and the problems associated with obtaining accurate measurements on such a diverse population with this technique would strongly support its not being included in a national examination survey of the general adult population. It is recommended that the current skinfold test battery be expanded to include a minimum of six sites and that one of the generalized equations be selected for estimating body density from the sum of three to six skinfold sites. Further, it is highly recommended that both waist and hip circumferences be obtained to determine fat patterning. Several recent studies have presented strong evidence supporting the concept that the male pattern (i.e., high ratio of waist to hip circumference) presents a much higher risk for certain disease entities. The addition of these skinfold and circumference measurements should not add more than 2–3 minutes to the present anthropometric test battery.

References

1. *Promoting Health/Preventing Disease: Objectives for the Nation.* Department of Health and Human Services, Public Health Service, U.S. Government Printing Office, Washington, D.C., 1980.
2. *The 1990 Health Objectives for the Nation: A Midcourse Review.* U.S. Department of Health and Human Services, Public Health Service, U.S. Government Printing Office, Washington, D.C., November 1986.
3. Skinner, J.S. (editor) *Exercise Testing and Exercise Prescription for Special Cases: Theoretical Basis and Clinical Application.* Philadelphia: Lea & Febiger, 1987.
4. Hunsicker, P. *Physical Fitness.* What Research Says to the Teacher Series, #26. National Education Association, Washington, D.C., 1963.
5. Getchell, B. *Physical Fitness A Way of Life.* 3rd edition. New York: John Wiley & Sons, 1983.
6. *AAHPERD Youth Fitness Test Manual.* Reston, VA: American Alliance for Health, Physical Education, Recreation, and Dance, 1976.
7. *Youth Physical Fitness.* President's Council on Physical Fitness. U.S. Government Printing Office, Washington, D.C., 1967.
8. Pate, R.R. A new definition of youth fitness. *Physician Sportsmed.* 11 (#4):77–83, 1983.
9. Wilmore, J.H. and D.L. Costill. *Training for Sport and Activity: The Physiological Basis of the Conditioning Process.* 3rd edition. Boston: Allyn and Bacon, 1987.
10. Mitchell, J.H. and G. Blomqvist. Maximal oxygen uptake. *New England J. Med.* 284:1018–1022, 1971.
11. Pollock, M.L. The quantification of endurance training programs. *Exerc. Sport Sci. Rev.* 1:155–188, 1973.
12. Greenleaf, J.E. and S. Kozlowski. Physiological consequences of reduced physical activity during bed rest. *Exerc. Sport Sci. Rev.* 10:84–119, 1982.
13. Klissouras, V. Heritability of adaptive variation. *J. Appl. Physiol.* 31:338–344, 1971.
14. Bouchard, C. and R.M. Malina. Genetics of physiological fitness and motor performance. *Exerc. Sport Sci. Rev.* 11:306–339, 1983.
15. Wilmore, J.H. and J.J. McNamara. Prevalence of coronary heart disease risk factors in boys, 8 to 12 years of age. *J. Pediatr.* 84:527–533, 1974.
16. Gilliam, T.B., V.L. Katch, W. Thorland, and A. Weltman. Prevalence of coronary heart disease risk factors in active children, 7 to 12 years of age. *Med. Sci. Sports Exerc.* 9:21–25, 1977.
17. Wilmore, J.H., S.H. Constable, P.R. Stanforth, W.Y. Tsao, T.C. Rotkis, R.M. Paicius, C.M. Mattern, and G.A. Ewy. Prevalence of coronary heart disease risk factors in 13- to 15-year-old boys. *J. Cardiac Rehab.* 2:223–233, 1982.
18. Saris, W.H.M., R.A. Binkhorst, A.B. Cramwinckel, F. van Waesberghe, and A.M. van der Veen-Hezemans. The relationship between working performance, daily physical activity, fatness, blood lipids, and nutrition in schoolchildren. In *Children and Exercise IX.* Edited by K. Berg and B.O. Eriksson. Baltimore: University Park Press, 1980.

19. Noonan, J.M. Physical Activity Patterns in 13- to 15-Year-Old Boys, M.S. Thesis, Department of Physical Education, University of Arizona, 1983.
20. Ashton, N.J. Relationship of chronic physical activity levels to physiological and anthropometric variables in 9–10 year old girls. *Med. Sci. Sports Exerc.* 15:143 (abstr), 1983.
21. Atha, J. Strengthening muscle. *Exerc. Sport Sci. Rev.* 9:1–73, 1982.
22. Clarke, H.H. Toward a better understanding of muscular strength. *Phys. Fitness Res. Digest* 3:1–20, January 1973.
23. Jackson, A., M. Watkins, and R.W. Patton. A factor analysis of twelve selected maximal isotonic strength performances on the Universal Gym. *Med. Sci. Sports Exerc.* 12:274–277, 1980.
24. Clarkson, P.M., W. Kroll, and T.C. McBride. Plantar flexion fatigue and muscle fiber type in power and endurance athletes. *Med. Sci. Sports Exerc.* 12:262–267, 1980.
25. Komi, P.V., J.H.T. Viitasalo, M. Havu, A. Thortensson, B. Sjodin, and J. Karlsson. Skeletal muscle fibres and muscle enzyme activities in monozygous and dizygous twins of both sexes. *Acta Physiol. Scand.* 100:385–392, 1977.
26. Hubley, C. Testing flexibility. In *Physiological Testing of the Elite Athlete.* Edited by J.D. MacDougall, H.A. Wenger, and H.J. Green. Canadian Association of Sports Sciences: Mutual Press Limited, 1982.
27. Harris, M.L. A factor analytic study of flexibility. *Res. Quart.* 40:62–70, 1969.
28. Grande, F., and A. Keys. Body weight, body composition and calorie status. In *Modern Nutrition in Health and Disease.* 6th edition, edited by R.S. Goodhart and M.E. Shils. Philadelphia: Lea & Febiger, 1980.
29. U.S. Department of Health, Education and Welfare, Public Health Service: *Obesity and Health.* PHS Pub. No. 1485. Washington, D.C., U.S. Public Health Service, 1966.
30. Metropolitan Life Insurance Company: New weight standards for men and women. *Stat. Bull. Metropol. Life Insur. Co.* 40:3, 1959.
31. Metropolitan Life Insurance Company. *1983 Metropolitan Height and Weight Tables.*
32. Welham, W.C., and A.R. Behnke. The specific gravity of healthy men; body weight divided by volume and other physical characteristics of exceptional athletes and of naval personnel. *JAMA* 118:498–501, 1942.
33. Behnke, A.R., and J.H. Wilmore. *Evaluation and Regulation of Body Build and Composition.* Englewood-Cliffs, NJ: Prentice-Hall, 1974.
34. Brozek, J., F. Grande, J.T. Anderson, and A. Keys. Densitometric analysis of body composition: Revision of some quantitative assumptions. *NY Acad. Sci.* 110:113–140, 1963.
35. Lohman, T.G. Applicability of body composition techniques and constants for children and youths. *Exerc. Sport Sci. Rev.* 14:325–357, 1986.
36. Schutte, J.E., E.J. Townsend, J. Hugg, R.F. Shoup, R.M. Malina, and C.G. Blomqvist. Density of lean body mass is greater in Blacks than in Whites. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 56:1647–1649, 1984.
37. Lohman, T.G., R.A. Boileau, and M.H. Slaughter. Body composition in children and youth. In *Advances in Pediatric Sport Sciences,* R.A. Boileau, ed. Champaign, IL: Human Kinetics Publishers, Inc., 1984.
38. Wilmore, J.H. Body composition in sport and exercise: Directions for future research. *Med. Sci. Sports Exerc.* 15:21–31, 1983.
39. Jackson, A.S., and M.L. Pollock. Generalized equations for predicting body density of men. *Br. J. Nutr.* 40:497–504, 1978.
40. Jackson, A.S., M.L. Pollock, and A. Ward. Generalized equations for predicting body density of women. *Med. Sci. Sports Exerc.* 12:175–182, 1980.
41. Sinning, W.E., and J.R. Wilson. Validity of "generalized equations" for body composition analysis in women athletes. *Res. Quart. Exerc. Sports.* 55:153–160, 1984.
42. Sinning, W.E., D.G. Dolny, K.D. Little, L.N. Cunningham, A. Racaniello, S.F. Siconolfi, and J.L. Sholes. Validity of "generalized equations" for body composition analysis in male athletes. *Med. Sci. Sports Exerc.* 17:124–130, 1985.
43. Wilmore, J.H. and D.L. Costill. Adequacy of the Haldane transformation in the computation of exercise VO_2 in man. *J. Appl. Physiol.* 35:85–89, 1973.
44. Consolazio, C.F., R.E. Johnson, and L. Pecora. *Physiological Measurements of Metabolic Functions in Man.* New York: McGraw-Hill Book Company, 1963.
45. Wilmore, J.H., and D.L. Costill. Semiautomated systems approach to the assessment of oxygen uptake during exercise. *J. Appl. Physiol.* 36:618–620, 1974.
46. Wilmore, J.H., J.A. Davis, and A.C. Norton. An automated system for assessing metabolic and respiratory function during exercise. *J. Appl. Physiol.* 40:619–624, 1976.
47. Master, A.M. Two-step test of myocardial function. *Am. Heart J.* 10:495, 1934.
48. Christensen, E.H. Beitrage zur physiologie schwerer korperlicher arbeit. *Arbeitsphysiol.* 4:154–174, 1931.
49. Nagle, F.J., B. Balke, and J.P. Naughton. Gradational step tests for assessing work capacity. *J. Appl. Physiol.* 20:745–748, 1965.
50. Kamon, E. Negative and positive work in climbing a laddermill. *J. Appl. Physiol.* 29:1–5, 1970.
51. von Dobeln, W. A simple bicycle ergometer. *J. Appl. Physiol.* 7:222–224, 1954.
52. Giezendanner, D., P.E. DiPrampero, and P. Cerretelli. A programmable electrically braked ergometer. *J. Appl. Physiol: Respirat. Environ. Exercise Physiol.* 55:578–582, 1983.
53. Briggs, C.A., and R.D. Telford. Comparison of submaximal and maximal cardiorespiratory responses utilizing air-braked ergometers and a treadmill. *Aust. J. Sports Med.* 11:5–8, 1979.
54. Nagle, F.J., J.P. Richie, and M.D. Giese. VO_2 max responses in separate and combined arm and leg air-braked ergometer exercise. *Med. Sci. Sports Exercise* 16:563–566, 1984.
55. Telford, R.D., L.A. Hooper, and M.H.D. Chennells. Calibration and comparison of air-braked and mechanically-braked bicycle ergometers. *Aust. J. Sports Med.* 12:40–46, 1980.
56. Hagberg, J.M., J.P. Mullin, M.D. Giese, and E. Spitznagel. Effect of pedaling rate on submaximal exercise responses of competitive cyclists. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 51:447–451, 1981.
57. Faulkner, J.A., D.E. Roberts, R.L. Elk, and J. Conway. Cardiovascular responses to submaximum and maximum effort cycling and running. *J. Appl. Physiol.* 30:457–461, 1971.
58. Miyamura, M., K. Kitamura, A. Yamada, and H. Matsui. Cardiorespiratory responses to maximal treadmill and bicycle exercise in trained and untrained subjects. *J. Sports Med. Physical Fitness* 18:25–32, 1978.
59. Wilmore, J.H., S.H. Constable, P.R. Stanforth, M.J. Buono, Y.W. Tsao, F.B. Roby, Jr., B.J. Lowdon, and R.A. Ratliff. Mechanical and physiological calibration of four cycle ergometers. *Med. Sci. Sports Exercise* 14:322–325, 1982.
60. Stromme, S.B., F. Ingjer, and H.D. Meen. Assessment of maximal aerobic power in specifically trained athletes. *J. Appl. Physiol: Respirat. Environ. Exercise Physiol.* 42:833–837, 1977.
61. Zeimetz, G.A., R.F. Moss, N. Butts, P. Wilson, and R. Obma. Support versus nonsupport treadmill walking. *Med. Sci. Sports* 11:112 (abstract), 1979.
62. Ragg, K.E., T.F. Murray, L.M. Karbonit, and D.A. Jump. Errors in predicting functional capacity from a treadmill exercise stress test. *Amer. Heart J.* 100:581–583, 1980.
63. Haskell, W.L., W. Savin, N. Oldridge, and R. DeBusk. Factors influencing estimated oxygen uptake during exercise testing soon after myocardial infarction. *Amer. J. Cardiol.* 50:299–304, 1982.
64. Magel, J.R., G.F. Foglia, W.D. McArdle, B. Gutin, G.S. Pechar, and F.I. Katch. Specificity of swim training on maximum oxygen uptake. *J. Appl. Physiol.* 38:151–155, 1975.
65. Gergley, T.J., W.D. McArdle, P. DeJesus, M.M. Toner, S. Jacobowitz, and R.J. Spina. Specificity of arm training of aerobic power

during swimming and running. *Med. Sci. Sports Exerc.* 16:349–354, 1984.

66. Hagberg, J.M., F.J. Nagle, and J.L. Carlson. Transient O_2 uptake response at the onset of exercise. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 44:90–92, 1978.
67. Rowell, L.B. *Human Circulation Regulation During Physical Stress.* New York: Oxford University Press, 1986, pp. 363–406.
68. Hagberg, J.M., J.P. Mullin, and F.J. Nagle. Oxygen consumption during constant-load exercise. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 45:381–384, 1978.
69. Davis, J.A., B.J. Whipp, N. Lamarra, D.J. Huntsman, M.H. Frank, and K. Wasserman. Effect of ramp slope on determination of aerobic parameters from the ramp exercise test. *Med. Sci. Sports Exerc.* 14:339–343, 1982.
70. Wigertz, O. Dynamics of ventilation and heart rate in response to sinusoidal work load in man. *J. Appl. Physiol.* 29:208–218, 1970.
71. Casaburi, R., B.J. Whipp, K. Wasserman, W.L. Beaver, and S.N. Koyal. Ventilatory and gas exchange dynamics in response to sinusoidal work. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 42:300–311, 1977.
72. Bruce, R.A., J.R. Blackmon, J.W. Jones, and G. Strait. Exercise testing in adult normal subjects and cardiac patients. *Pediatr.* 32:742–756, 1963.
73. Åstrand, P.-O., and B. Saltin. Oxygen uptake during the first minutes of heavy muscular exercise. *J. Appl. Physiol.* 16:971–976, 1961.
74. McDonough, J.R., F. Kusumi, and R.A. Bruce. Variations in maximal oxygen intake with physical activity in middle-aged men. *Circulation* 41:743–751, 1970.
75. Profant, G.R., R. Early, K.L. Nilson, F. Kusumi, V. Hofer, and R.A. Bruce. Responses to maximal exercise in healthy middle-aged women. *J. Appl. Physiol.* 33:595–599, 1972.
76. Freund. B.J., D. Allen, and J.H. Wilmore. Interaction of test protocol and inclined run training on maximal oxygen uptake. *Med. Sci. Sports Exerc.* 18:588–592, 1986.
77. Allen D., B.J. Freund, and J.H. Wilmore. Interaction of test protocol and horizontal run training on maximal oxygen uptake. *Med. Sci. Sports Exerc.* 18:581–587, 1986.
78. Jopke, T. Choosing an exercise testing protocol. *Physician Sportsmed.* 9:141–146, 1981.
79. Balke, B., and R.W. Ware. An experimental study of physical fitness of Air Force personnel. *U.S. Armed Forces Med. J.* 10:675–688, 1959.
80. Froelicher, V.F. Use of the exercise electrocardiogram to identify latent coronary atherosclerotic heart disease. In *Exercise in Cardiovascular Health and Disease.* Edited by E.A. Amsterdam, J.H. Wilmore, and A.N. DeMaria. New York: Yorke Medical Books, 1977, p. 189.
81. Faris, J.V., P.L. McHenry, and S.N. Morris. Concepts and applications of treadmill exercise testing and the exercise electrocardiogram. *Amer. Heart J.* 95:102–114, 1978.
82. Doan, A.E., D.R. Peterson, J.R. Blackmon, and R.A. Bruce. Myocardial ischemia after maximal exercise in healthy men. *Amer. Heart J.* 69:11–21, 1965.
83. Kattus, A.A., A. Alvaro, and R.N. MacAlpin. Treadmill exercise tests for capacity and adaptation in angina pectoris. *J. Occup. Med.* 10:627–637, 1968.
84. Kattus, A.A., C.R. Jorgensen, R.E. Worden, and A.B. Alvaro. S-T-segment depression with near-maximal exercise in detection of preclinical coronary heart disease. *Circulation* 41:585–595, 1971.
85. Ellestad, M.H. *Stress Testing Principles and Practice.* Philadelphia: F.A. Davis, 1975.
86. Naughton, J., and R. Haider. Methods of exercise testing. In *Exercise Testing and Exercise Training in Coronary Heart Disease.* Edited by J.P. Naughton and H.K. Hellerstein. New York: Academic Press, Inc., 1973, p. 89.
87. Taylor, H.L., E. Buskirk, and A. Henschel. Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J. Appl. Physiol.* 8:73–80, 1955.
88. Pollock, M.L., R.L. Bohannon, K.H. Cooper, J.J. Ayres, A. Ward, S.R. White, and A.C. Linnerud. A comparative analysis of four protocols for maximal treadmill stress testing. *Amer. Heart J.* 92:39–46, 1976.
89. Froelicher, V.F., H. Brammel, G. Davis, I. Noguera, A. Stewart, and M.C. Lancaster. A comparison of the reproducibility and physiologic response to three maximal treadmill exercise protocols. *Chest* 65:512–517, 1974.
90. Bruce, R.A. Principles of exercise testing. In *Exercise Testing and Exercise Training in Coronary Heart Disease.* Edited by J.P. Naughton and H.K. Hellerstein. New York: Academic Press, Inc., 1973, pp. 45–59.
91. Froelicher, V.F. *Exercise Testing and Training.* New York: Le Jacq Publishing, Inc., 1983, pp. 12–14 and 95.
92. Fox, S.M., III, J.P. Naughton, and W.L. Haskell. Physical activity and the prevention of coronary heart disease. *Annals Clin. Res.* 3:404–432, 1971.
93. Cumming, G.R. Yield of ischemic exercise electrocardiograms in relation to exercise intensity in a normal population. *Brit. Heart J.* 34:919–923, 1972.
94. Stuart, R.J. and M.H. Ellestad. National survey of exercise stress testing facilities. *Chest* 77:94–97, 1980.
95. Sheffield, L.T., W. Haskell, G. Heiss, M. Kioschos, A. Leon, D. Roitman, and H. Schrott. Safety of exercise testing volunteer subjects: The Lipid Research Clinics' Prevalence Study experience. *J. Cardiac Rehab.* 2:395–400, 1982.
96. Åstrand, P.-O., and K. Rodahl. *Textbook of Work Physiology.* 3rd edition. New York: McGraw-Hill Book Company Inc., 1986.
97. Issekutz, B., N.C. Birkhead, and K. Rodahl. The use of respiratory quotients in assessment of aerobic work capacity. *J. Appl. Physiol.* 17:47–50, 1962.
98. Maritz, J.S., J.F. Morrison, J. Peter, N.B. Strydom, and C.H. Wyndham. A practical method of estimating an individual's maximum oxygen intake. *Ergonomics* 4:97–122, 1961.
99. Åstrand, P.-O. *Experimental Studies of Physical Working Capacity in Relation to Sex and Age.* Copenhagen: Munksgaard, 1952, pp. 15–37.
100. Borg, G.A.V., and B.J. Noble. Perceived exertion. *Exerc. Sport Sci. Rev.* 2:131–153, 1974.
101. Niemela, K., I. Palatsi, M. Linnaluoto, and J. Takkunen. Criteria for maximum oxygen uptake in progressive bicycle tests. *Europ. J. Appl. Physiol.* 44:51–59, 1980.
102. Åstrand, P.-O., and I. Ryhming. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *J. Appl. Physiol.* 7:218–221, 1954.
103. Kurucz, R.L., E.L. Fox, and D.K. Mathews. Construction of a submaximal cardiovascular step test. *Res. Quart.* 40:115–122, 1969.
104. Margaria, R., P. Aghemo, and E. Rovelli. Indirect determination of maximal O_2 consumption in man. *J. Appl. Physiol.* 20:1070–1073, 1965.
105. Sjostrand, T. Changes in the respiratory organs of workmen at an ore smelting works. *Acta Med. Scand.* 196:687–699, 1947.
106. Wahlund, H. Determination of the physical working capacity. *Acta Med. Scand.* Suppl. 215:1–78, 1948.
107. Myers, C.R., L.A. Golding, and W.E. Sinning. *The Y's Way to Physical Fitness.* Chicago: National Council of the YMCA, 1973.
108. Golding, L.A., C.R. Myers, and W.E. Sinning, editors. *The Y's Way to Physical Fitness.* Revised. Chicago: The YMCA of the USA, 1982.
109. Siconolfi, S.F., E.M. Cullinane, R.A. Carleton, and P.D. Thompson. Assessing VO_2max in epidemiologic studies: Modification of the Astrand-Ryhming test. *Med. Sci. Sports Exerc.* 14:335–338, 1982.
110. Balke, B. A simple field test for the assessment of physical fitness. *Federal Aviation Agency, Aviation Medical Service.* April 1963.
111. Cooper, K.H. A means of assessing maximal oxygen intake. *JAMA* 203:201–204, 1968.

112. Bruce, R.A., F. Kusumi, and D. Hosmer. Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. *Amer. Heart J.* 85:546–562, 1973.
113. Froelicher, V.F., and M.C. Lancaster. The prediction of maximal oxygen consumption from a continuous exercise treadmill protocol. *Amer. Heart J.* 87:445–450, 1974.
114. Froelicher, V.F., A.J. Thompson, I. Noguera, G. Davis, A.J. Stewart, and J.H. Triebwasser. Prediction of maximal oxygen consumption: Comparison of the Bruce and Balke treadmill protocols. *Chest* 68:331–336, 1975.
115. Foster, C., A.S. Jackson, M.L. Pollock, M.M. Taylor, J. Hare, S.M. Sennett, J.L. Rod, M. Sarwar, and D.H. Schmidt. Generalized equations for predicting functional capacity from treadmill performance. *Amer. Heart J.* 107:1229–1234, 1984.
116. Stull, G.A., and D.H. Clarke. Muscular strength testing and training. *Proceedings of the First International Conference on Lifestyle and Health: Optimal Health and Fitness for People with Physical Disabilities.* Minneapolis, MN: University of Minnesota, 1979, pp. 195–214.
117. Hunsicker, P.A., and R.J. Donnelly. Instruments to measure strength. *Res. Quart.* 26:408–420, 1955.
118. Clarke, H.H. *Muscular Strength and Endurance in Man.* Englewood Cliffs, NJ: Prentice-Hall, Inc., 1966, pp. 2–51.
119. Perrine, J.J. Isokinetic exercise. *J. Health Phys. Ed. Rec.* 39:40–44, 1968.
120. Ikai, M., and A.H. Steinhaus. Some factors modifying the expression of human strength. *J. Appl. Physiol.* 16:157–163, 1961.
121. Pollock, M.L., J.H. Wilmore, and S.M. Fox, III. *Exercise in Health and Disease: Evaluation and Prescription for Prevention and Rehabilitation.* Philadelphia: W.B. Saunders Company, 1984.
122. deVries, H.A. *Physiology of Exercise for Physical Education and Athletics.* 3rd edition. Dubuque, IA: Wm. C. Brown, 1980, pp. 462–466.
123. Kraus, H., and W. Raab. *Hypokinetic Disease.* Springfield, IL: Charles C. Thomas, 1961, p. 11.
124. Bakker, H.K., and R.S. Struikenkamp. Biological variability and lean body mass estimates. *Human Biol.* 49:187–202, 1977.
125. Siri, W.E. Body composition from fluid spaces and density. In *Techniques for Measuring Body Composition.* J. Brozek and A. Henschel, eds. Washington, D.C.: National Academy of Sciences, 1961, pp. 223–244.
126. Pollock, M.L., and A.S. Jackson. Research progress in validation of clinical methods of assessing body composition. *Med. Sci. Sports Exerc.* 16:606–613, 1984.
127. Fahey, T.D., and R. Schroeder. A load cell for hydrostatic weighing. *Res. Quart.* 49:85–87, 1978.
128. Katch, F.I. Practice curves and errors of measurement in estimating underwater weight by hydrostatic weighing. *Med. Sci. Sports* 1:212–216, 1969.
129. Buskirk, E.R. Underwater weighing and body density: A review of procedures. In *Techniques for Measuring Body Composition.* J. Brozek and A. Henschel, eds. Washington, D.C.: National Academy of Sciences, 1961, pp. 90–106.
130. Girandola, R.N., R.A. Wiswell, J.G. Mohler, G.T. Romero, and W.S. Barnes. Effects of water immersion on lung volumes: Implications for body composition analysis. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 43:276–279, 1977.
131. Thomas, T.R., and G.L. Etheridge. Hydrostatic weighing at residual volume and functional residual capacity. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 49:157–159, 1980.
132. Welch, B.E., and C.E. Crisp. Effect of the level of expiration on body density measurement. *J. Appl. Physiol.* 12:399–402, 1958.
133. Weltman, A., and V. Katch. Comparison of hydrostatic weighing at residual volume and total lung capacity. *Med. Sci. Sports Exerc.* 13:210–213, 1981.
134. Timson, B.F., and J.L. Coffman. Body composition by hydrostatic weighing at total lung capacity and residual volume. *Med. Sci. Sports Exerc.* 16:411–414, 1984.
135. Lohman, T.G. Research progress in validation of laboratory methods of assessing body composition. *Med. Sci. Sports Exerc.* 16:596–603, 1984.
136. Katch, F.I., and A.R. Behnke. Arm x-ray assessment of percent body fat in men and women. *Med. Sci. Sports Exerc.* 16:316–321, 1984.
137. Presta, E., K.R. Segal, B. Gutin, G.G. Harrison, and T.B. Van Itallie. Comparison in man of total body electrical conductivity and lean body mass derived from body density: Validation of a new body composition method. *Metabolism* 32:524–527, 1983.
138. Segal, K.R., B. Gutin, E. Presta, J. Wang, and T.B. Van Itallie. Estimation of human body composition by electrical impedance methods: A comparative study. *J. Appl. Physiol.* 58:1565–1571, 1985.
139. Lukaski, H.C., W.W. Bolonchuk, C.B. Hall, and W.A. Siders. Validation of tetrapolar bioelectric impedance method to assess human body composition. *J. Appl. Physiol.* 60:1327–1332, 1986.
140. Maughan, R.J., J.S. Watson, and J. Weir. The relative proportions of fat, muscle and bone in the normal human forearm as determined by computed tomography. *Clin. Sci.* 66:191–197, 1984.
141. Mazess, R.B., W.W. Peppler, and M. Gibbons. Total body composition by dualphoton (^{153}Gd) absorptiometry. *Amer. J. Clin. Nutr.* 40:834–839, 1984.
142. Brozek, J.F., and A. Keys. The evaluation of leanness-fatness in man. *Br. J. Nutr.* 5:194–206, 1951.
143. Jackson, A.S. Research design and analysis of data procedures for predicting body density. *Med. Sci. Sports Exerc.* 16:616–620, 1984.
144. Durnin, J.V.G.A., and J. Womersley. Body fat assessed from total body density and its estimation from skinfold thickness: Measurements on 481 men and women aged from 16 to 72 years. *Br. J. Nutr.* 32:77–97, 1974.
145. Pollock, M.L., and A.S. Jackson. Research progress in validation of clinical methods of assessing body composition. *Med. Sci. Sports Exerc.* 16:606–613, 1984.
146. Larsson, B., K. Svardsudd, L. Welin, L. Wilhelmsen, P. Bjorntorp, and G. Tibblin. Abdominal adipose tissue distribution, obesity and risk of cardiovascular disease and death: A 13 year follow up of participants in the study of men born in 1913. *Br. Med. J.,* 288:1401–1404, 1984.
147. Lapidus, L., C. Bengtsson, B. Larsson, K. Pennert, E. Rybo, and L. Sjostrom. Distribution of adipose tissue and risk of cardiovascular disease and death: A 12 year follow up of participants in the population study of women in Gothenburg, Sweden. *Br. Med. J.* 289:1261–1263, 1984.
148. Bjorntorp, P. Regional patterns of fat distribution. *Annals. Int. Med.* 103:994–995, 1985.
149. Lohman, T.G., M.L. Pollock, M.H. Slaughter, L.J. Brandon, and R.A. Boileau. Methodological factors and the prediction of body fat in female athletes. *Med. Sci. Sports Exerc.* 16:92–96, 1984.

Appendix A: Objectives for the Nation

PHYSICAL FITNESS AND EXERCISE

1. Nature and Extent of the Problem

The health benefits associated with regular physical fitness and exercise have not yet been fully defined. Based on what is now known it appears that substantial physical and emotional benefits, direct and indirect, are possible. Yet most Americans do not engage in appropriate physical activity, either during recreation or in the course of their work. For the purposes of this discussion, "appropriate physical activity" refers to exercise which involves large muscle groups in dynamic movement for periods of 20 minutes or longer, three or more days per week, and which is performed at an intensity requiring 60 percent or greater of an individual's cardiorespiratory capacity. Exercise to improve flexibility and muscular strength may reduce the frequency of musculoskeletal problems and is an important supplement to cardiovascular conditioning activities.

a. Health implications

- Most people feel better when they exercise.
- Physical inactivity can result in decreased physical working capacity at all ages, with concomitant decreases in physiologic function and health status.
- Physical inactivity is associated with an increased risk of developing obesity and its disease correlates.
- Physical inactivity is associated with increased risk of coronary heart disease.
- Appropriate physical activity may be a valuable tool in therapeutic regimens for control and amelioration (rehabilitation) of obesity, coronary heart disease, hypertension, diabetes, musculoskeletal problems, respiratory diseases, stress and depression/anxiety. Such physical activity, however, is still not routinely prescribed for the treatment of these conditions.

b. Status and trends

- Though physical fitness and exercise activities have increased in recent years—and over 50 percent of adults reported regular exercise in popular opinion polls—generous estimates place the proportions of regularly exercising adults ages 18 to 65 at something over 35 percent.
- Regular runners include approximately 5 percent of all Americans over age 20, and 10 percent of men aged 20 to 44.
- About 36 percent of adults ages 65 and older were estimated in 1975 to take regular walks.
- Only about a third of children and adolescents ages 10 to 17 are estimated to participate in daily school physical education programs, and the share is declining.
- Many high school programs focus on competitive sports that involve a relatively small proportion of students.
- Though growing, the awareness of the health benefits of regular exercise is limited.
- Only a small proportion (about 2.5 percent) of companies and institutions with greater than 500 employees offer fitness programs for their workers.
- Certain groups demonstrate disproportionately low rates of participation in appropriate physical activity, including girls and women, older people, physically and mentally handicapped people of all ages, inner city and rural residents, people of low socioeconomic status and residents of institutions.

2. Prevention/Promotion Measures

a. Potential measures

- Education and information measures include:
  - — using television and radio public service announcements to provide information on appropriate physical activity and its benefits;
  - — providing information in school and college-based programs;
  - — providing information in health care delivery systems, including incorporation of queries about exercise habits into the routine clinical history;
  - — encouraging health care providers, especially in HMOs, community health centers and other organized settings, to prescribe appropriate exercise in weight loss regimens as a complementary treatment modality in the management of several chronic diseases, and to give patients 65 years and older and the handicapped more detailed information on appropriate physical activity together with warnings about starting up exercise too fast;
  - — adopting an exercise component by community service agencies (such as the American Red Cross, the American Heart Association);

  - assuring that all programs and materials related to diet and weight loss have an active exercise component;
  - tailoring education programs to the needs and characteristics of specific populations.

- Service measures include:
  - providing physical fitness and exercise programs to school children, and ensuring that those programs emphasize activities for all children rather than just competitive sports for relatively few;
  - providing physical fitness and exercise programs in colleges;
  - providing worksite-based fitness programs which are linked to other health enhancement components (e.g., smoking cessation, nutrition improvement) and which have an active outreach effort;
  - incorporating exercise and fitness protocols as regular clinical tools of health providers.

- Technologic measures include:
  - increasing the availability of existing facilities and promoting the development of new facilities by public, private and corporate entities (e.g., fitness trails, bike paths, parks, pools);
  - upgrading existing facilities, especially in inner city neighborhoods, and involving the population to be served at all levels of planning.

- Legislative and regulatory measures include:
  - city council support for bicycle and walking paths for use in trips to work and school;
  - developing and operating local, State and National park facilities which can be used for physical fitness activities in urban areas;
  - increasing the number of school-mandated physical education programs that focus on health-related physical fitness;
  - establishing State and local councils on health promotion and physical fitness;
  - allowing expenditure of funds for fitness-related activities under Federally funded programs guided by Federal regulations.

- Economic measures include:
  - tax incentives for the private sector to offer physical fitness programs for employees;
  - encouraging employers to permit employees to exercise on company time and/or giving employees flexible time for use of facilities;
  - offering health and life insurance policies with reduced premiums for those who participate in regular vigorous physical activity.

b. Relative strength of the measures

- Programs which are most likely to be successful in recruiting new participants to appropriate physical activity include those which offer services and facilities to individuals, and economic incentives to groups and individuals.
- On the other hand, programs which can more easily be implemented include those related to the provision of public information and education and improving the linkages with other health promotion efforts.
- The effectiveness of all measures is handicapped by the limitation in knowledge with respect to:
  - the relation between exercise and physical and emotional health;
  - the optimum types of exercises for various groups of people with special needs;
  - the appropriate way to measure levels of physical fitness for various age groups.

3. Specific Objectives for 1990

- Improved health status
  - Increased levels of physical fitness may contribute to reduced heart and lung disease rates, possibly reduced injuries among the elderly, and, more broadly, an enhanced sense of well-being which may reinforce positive health behaviors in other areas. Currently, however, few quantifiable health status objectives for physical fitness and exercise can be developed.

- Reduced risk factors
  a. By 1990, the proportion of children and adolescents ages 10 to 17 participating regularly in appropriate physical activities, particularly cardiorespiratory fitness programs which can be carried into adulthood, should be greater than 90 percent. (Baseline data unavailable.)
  b. By 1990, the proportion of children and adolescents ages 10 to 17 participating in daily school physical education programs should be greater than 60 percent. (In 1974–75, the share was 33 percent.)
  c. By 1990, the proportion of adults 18 to 65 participating regularly in vigorous physical exercise should be greater than 60 percent. (In 1978, the proportion who regularly exercise was estimated at over 35 percent.)
  d. By 1990, 50 percent of adults 65 years and older should be engaging in appropriate physical activity, e.g., regular walking, swimming or other aerobic activity. (In 1975, about 36 percent took regular walks.)

- Increased public/professional awareness
  e. By 1990, the proportion of adults who can accurately identify the variety and duration of exercise thought to promote most effectively cardiovascular fitness should be greater than 70 percent. (Baseline data unavailable.)
  f. By 1990, the proportion of primary care physicians who include a careful exercise

history as part of their initial examination of new patients should be greater than 50 percent. (Baseline data unavailable.)

- Improved services/protection
  - g. By 1990, the proportion of employees of companies and institutions with more than 500 employees offering employer-sponsored fitness programs should be greater than 25 percent. (In 1979, about 2.5 percent of companies had formally organized fitness programs.)
- Improved surveillance/evaluation systems
  - h. By 1990, a methodology for systematically assessing the physical fitness of children should be established, with at least 70 percent of children and adolescents ages 10 to 17 participating in such an assessment.
  - i. By 1990, data should be available with which to evaluate the short and long-term health effects of participation in programs of appropriate physical activity.
  - j. By 1990, data should be available to evaluate the effects of participation in programs of physical fitness on job performance and health care costs.
  - k. By 1990, data should be available for regular monitoring of National trends and patterns of participation in physical activity, including participation in public recreation programs in community facilities.

4. Principal Assumptions

- Increased physical activity by the American public will result in overall improvements in health.
- Personal commitment to enhance health will become a prominent factor promoting increased participation in exercise activities in the United States.
- Voluntary agencies, private corporations and government will expand their commitment to physical fitness programs.
- Private industry and retailers will support activities promoting physical fitness, which will also promote increased sales of their products.
- Environmental, cultural and behavioral differences influence attitudes toward, and participation in, regular exercise.
- Inner city residents will continue to have fewer adequate facilities and appropriate activity programs.
- Special attention will be required to make gains in participation among lower socioeconomic groups.
- There will be a reversal of the trend in reductions of school-based programs aimed at promoting physical fitness. However, these programs will not necessarily be founded in the traditional physical education mold.
- New school-based programs will embrace activities which expand beyond competitive sports.
- The increasing costs associated with health care will compel public policy to emphasize measures such as physical fitness to enhance health.
- Reduced levels of physical fitness in the work force may result in increased absenteeism from acute illness and, accordingly, decreased productivity. Thus, employers have incentives for offering physical fitness programs to their employees.

5. Data Sources

a. To National level only

- Health Interview Survey (HIS). Extent of regular exercise; job related physical activity; regular participation in exercise DHHS-National Center for Health Statistics (NCHS). NCHS *Vital and Health Statistics,* Series 10, selected reports, and *Advance Data from Vital and Health Statistics,* No. 78–1250. Continuing survey; National probability sample.
- Extent of regular exercise. (Non-work related only.) Regular participation in exercise reported in household survey, and self-reported change over previous year. Survey for General Mills, conducted by Yankelovich, Skelly and White. *Family Health in an Era of Stress.* General Mills, Inc., 9200 Wayzata Boulevard, Minneapolis, Minnesota, 1979. One time survey; National probability sample.
- Extent of regular exercise. (Non-work related only.) Survey for Pacific Mutual Life Insurance Company, conducted by Louis Harris and Associates, Inc. *Health Maintenance,* 1978. Pacific Mutual Life Insurance, Newport Beach, California.
- Public attitudes regarding physical fitness. Attitudes, knowledge and behavior regarding physical fitness and exercise. Survey for Great Waters of France, conducted by Louis Harris and Associates, Inc. *The Perrier Study; Fitness in America,* 1979. One time survey; representative sample and special sample of runners.

b. To State and/or local level

- Exercise programs in schools. Student enrollment in physical fitness activities; program content and scheduling. Councils on Physical Fitness, selected States only.
- Student physical fitness levels. Councils on Physical Fitness, selected States only.

Appendix B: Estimates of maximal oxygen uptake and relative body fat

Table B–1. Maximal oxygen uptake of male and female athletes

Athletic group	Sex	Age (yr)	Height (cm)	Weight (kg)	$\dot{V}O_2$max (ml·kg^{-1}·min^{-1})	Reference[1]
Baseball/softball	Male	21	182.7	83.3	52.3	Novak (55)
	Male	28	183.6	88.1	52.0	Wilmore (95)
	Female	19–23	—	—	55.3	Rubal (70)
Basketball	Female	19	167.0	63.9	42.3	Conger (14)
	Female	19	169.1	62.6	42.9	Sinning (79)
	Female	19	173.0	68.3	49.6	Vaccaro (90)
Centers	Male	28	214.0	109.2	41.9	Parr (57)
Forwards	Male	25	200.6	96.9	45.9	Parr (57)
Guards	Male	25	188.0	83.6	50.0	Parr (57)
Bicycling (competitive)	Male	24	182.0	74.5	68.2	Gollnick (32)
	Male	24	180.4	79.2	70.3	Hermansen (38)
	Male	25	180.0	72.8	67.1	Burke (7)
	Male	—	180.3	67.1	74.0	Burke (7)
	Male	—	—	—	74.0	Saltin (73)
	Male	—	—	—	69.1	Stromme (85)
	Female	20	165.0	55.0	50.2	Burke (7)
	Female	—	167.7	61.3	57.4	Burke (7)
Canoeing/paddling	Male	19	173.0	64.0	60.0	Sidney (77)
	Male	22	190.5	80.7	67.7	Hermansen (38)
	Male	24	182.0	79.6	66.1	Rusko (72)
	Male	26	181.0	74.0	56.8	Gollnick (32)
	Female	18	166.0	57.3	49.2	Sidney (77)
Dancing:						
Ballet	Male	24	177.5	68.0	48.2	Cohen (13)
	Female	24	165.6	49.5	43.7	Cohen (13)
General	Female	21	162.7	51.2	41.5	Novak (56)
Football	Male	19	186.8	93.1	56.5	Smith (80)
	Male	20	184.9	96.4	51.3	Novak (55)
Defensive backs	Male	25	182.5	84.8	53.1	Wilmore (98)
Offensive backs	Male	25	183.8	90.7	52.2	Wilmore (98)
Linebackers	Male	24	188.6	102.2	52.1	Wilmore (98)
Offensive linemen	Male	25	193.0	112.6	49.9	Wilmore (98)
Defensive linemen	Male	26	192.4	117.1	44.9	Wilmore (98)
Quarterbacks/kickers	Male	24	185.0	90.1	49.0	Wilmore (98)
Gymnastics	Male	20	178.5	69.2	55.5	Novak (55)
	Female	15	159.7	48.8	49.8	Hermansen (38)
	Female	19	163.0	57.9	36.3	Conger (14)
Ice hockey	Male	11	140.5	35.5	56.6	Cunningham (23)
	Male	22	179.0	77.3	61.5	Rusko (72)
	Male	24	179.3	81.8	54.6	Seliger (75)
	Male	26	180.1	86.4	53.6	Wilmore (95)
Jockeys	Male	31	158.2	50.3	53.8	Wilmore (95)
Orienteering	Male	25	179.7	70.3	71.1	Hermansen (38)
	Male	31	—	72.2	61.6	Knowlton (42)
	Male	52	176.0	72.7	50.7	Gollnick (32)
	Female	23	165.8	60.0	60.7	Hermansen (38)
	Female	29	—	58.1	46.1	Knowlton (42)
Pentathlon	Female	21	175.4	65.4	45.9	Krahenbuhl (44)
Racquetball/handball	Male	24	183.7	81.3	60.0	Hermansen (38)
	Male	25	181.7	80.3	58.3	Pipes (61)
Rowing	Male	—	—	—	65.7	Stromme (85)
	Male	23	192.7	89.9	62.6	Mickelson (50)
	Male	25	189.9	86.9	66.9	Hermansen (38)
Heavyweight	Male	23	192.0	88.0	68.9	Hagerman (34)
Lightweight	Male	21	186.0	71.0	71.1	Hagerman (34)
	Female	23	173.0	68.0	60.3	Hagerman (34)
Skating:						
Speed	Male	20	175.5	73.9	56.1	Maksud (47)
	Male	21	181.0	76.5	72.9	Rusko (72)
	Male	25	183.1	82.4	64.6	Hermansen (38)
	Female	20	168.1	65.4	52.0	Hermansen (38)
	Female	21	164.5	60.8	46.1	Maksud (47)

Table B-1. Maximal oxygen uptake of male and female athletes—Continued

Athletic group	Sex	Age (yr)	Height (cm)	Weight (kg)	$\dot{V}O_2$max (ml·kg^{-1}·min^{-1})	Reference[1]
Skating—Con.						
Figure	Male	21	166.9	59.6	58.5	Niinimaa (53)
	Female	17	158.8	48.6	48.9	Niinimaa (53)
Skiing:						
Alpine	Male	16	173.1	65.5	65.6	Song (81)
	Male	21	176.0	70.1	63.8	Rusko (72)
	Male	22	177.8	75.5	66.6	Haymes (37)
	Male	26	176.6	74.8	62.3	Sprynarova (83)
	Female	19	165.1	58.8	52.7	Haymes (37)
Cross-country	Male	21	176.0	66.6	63.9	Niinimaa (54)
	Male	25	180.4	73.2	73.9	Hermansen (38)
	Male	26	174.0	69.3	78.3	Rusko (72)
	Male	23	176.2	73.2	73.0	Haymes (37)
	Male	—	—	—	72.8	Stromme (85)
	Female	20	163.4	55.9	61.5	Haymes (37)
	Female	24	163.0	59.1	68.2	Rusko (72)
	Female	25	165.7	60.5	56.9	Hermansen (38)
	Female	—	—	—	58.1	Stromme (85)
Nordic	Male	23	176.0	70.4	72.8	Rusko (72)
	Male	22	181.7	70.4	67.4	Haymes (37)
Ski jumping	Male	22	174.0	69.9	61.3	Rusko (72)
Soccer	Male	26	176.0	75.5	58.4	Raven (67)
Swimming	Male	12	150.4	41.2	52.5	Cunningham (21)
	Male	13	164.8	52.1	52.9	Cunningham (21)
	Male	15	169.6	59.8	56.6	Cunningham (21)
	Male	15	166.8	59.1	56.8	Vaccaro (89)
	Male	20	181.4	76.7	55.7	Magel (45)
	Male	20	181.0	73.0	50.4	Charbonnier (11)
	Male	21	182.9	78.9	62.1	Novak (55)
	Male	21	181.0	78.3	69.9	Gollnick (32)
	Male	22	182.3	79.1	56.9	Sprynarova (83)
	Male	22	182.3	79.7	55.9	Cunningham (21)
	Female	12	154.8	43.3	46.2	Cunningham (21)
	Female	13	160.0	52.1	43.4	Cunningham (21)
	Female	15	164.8	53.7	40.5	Cunningham (21)
Sprint	Male	19	181.1	75.0	58.3	Shephard (76)
Middle-distance	Male	22	178.0	74.6	55.4	Shephard (76)
Long-distance	Male	21	179.0	74.9	65.4	Shephard (76)
	Female	19	168.0	63.8	37.6	Conger (14)
Tennis	Male	42	179.6	77.1	50.2	Vodak (92)
	Female	39	163.3	55.7	44.2	Vodak (92)
Track and field	Male	21	180.6	71.6	66.1	Novak (55)
Run	Male	22	177.4	64.5	64.0	Sprynarova (83)
	Male	23	177.0	69.5	72.4	Gollnick (32)
Sprint	Male	17–22	—	—	51.0	Thomas (88)
	Male	46	177.0	74.1	47.2	Barnard (3)
Middle distance	Male	25	180.1	67.8	70.1	Costill (18)
	Male	25	179.0	72.3	69.8	Rusko (72)
Distance	Male	10	144.3	31.9	56.6	Mayers (49)
	Male	17–22	—	—	65.5	Thomas (88)
	Male	26	176.1	64.5	72.2	Hermansen (38)
	Male	26	178.9	63.9	77.4	Costill (18)
	Male	26	177.0	66.2	78.1	Rusko (72)
	Male	27	178.7	64.9	73.2	Costill (17)
	Male	32	177.3	64.3	70.3	Costill (19)
	Male	35	174.0	63.1	66.6	Costill (18)
	Male	36	177.3	69.6	65.1	Hagan (33)
	Male	40–49	180.7	71.6	57.5	Pollock (64)
	Male	55	174.5	63.4	54.4	Barnard (3)
	Male	50–59	174.7	67.2	54.4	Pollock (64)
	Male	60–69	175.7	67.1	51.4	Pollock (64)
	Male	70–75	175.6	66.8	40.0	Pollock (64)
	Male	—	—	—	72.5	Davies (24)
	Female	16	162.2	48.6	63.2	Burke (8)
	Female	16	163.3	50.9	50.8	Butts (9)

Table B-1. Maximal oxygen uptake of male and female athletes—Continued

Athletic group	Sex	Age (yr)	Height (cm)	Weight (kg)	$\dot{V}O_2$max (ml·kg^{-1}·min^{-1})	Reference[1]
Track and Field—Con.						
Distance-Con.						
	Female	21	170.2	58.6	57.5	Hermansen (38)
	Female	32	169.4	57.2	59.1	Wilmore (97)
	Female	44	161.5	53.8	43.4	Vaccaro (91)
	Female	—	—	—	58.2	Davies (24)
Race walking	Male	27	178.7	68.5	62.9	Franklin (28)
Jumping	Male	17–22	—	—	55.0	Thomas (88)
Shot/discus	Male	17–22	—	—	49.5	Thomas (88)
	Male	26	190.8	110.5	42.8	Wilmore (95)
	Male	27	188.2	112.5	42.6	Fahey (27)
	Male	28	186.1	104.7	47.5	Fahey (27)
Volleyball	Male	25	187.0	84.5	56.4	Conlee (15)
	Male	26	192.7	85.5	56.1	Puhl (66)
	Female	19	166.0	59.8	43.5	Conger (14)
	Female	20	172.2	64.1	56.0	Kovaleski (43)
	Female	22	183.7	73.4	41.7	Spence (82)
	Female	22	178.3	70.5	50.6	Puhl (66)
Weight lifting	Male	25	171.0	81.3	40.1	Gollnick (32)
	Male	25	166.4	77.2	42.6	Sprynarova (83)
Power	Male	26	176.1	92.0	49.5	Fahey (27)
Olympic	Male	25	177.1	88.2	50.7	Fahey (27)
Body building	Male	27	178.8	88.1	46.3	Pipes (60)
	Male	29	172.4	83.1	41.5	Fahey (27)
Wrestling	Male	21	174.8	67.3	58.3	Stine (84)
	Male	23	—	79.2	50.4	Taylor (86)
	Male	24	175.6	77.7	60.9	Nagle (52)
	Male	26	177.0	81.8	64.0	Fahey (27)
	Male	27	176.0	75.7	54.3	Gale (30)

[1] First author only. Complete references are presented in the following list.

References for table B-1

Citations in the reference list may not appear in the table.

1. Astrand, P.-O. and Englesson, S. A swimming flume. *J Appl Physiol* 33:514, 1972.
2. Astrand, P.-O. and Rodahl, K. *Textbook of Work Physiology.* 2nd ed. New York: McGraw-Hill, 1977, p. 321.
3. Barnard, R.J., Grimditch, G.K., Wilmore, J.H. Physiological characteristics of sprint and endurance masters runners. *Med Sci Sports* 11:167–171, 1979.
4. Bar-Or, O. Predicting athletic performance. *Physician Sportsmed* 6:80–85, 1975.
5. Bonen, A., Wilson, B.A., Yarkony, M., et al. Maximal oxygen uptake during free, tethered, and flume swimming. *J Appl Physiol* 48:232–235, 1980.
6. Burke, E.R. Physiological characteristics of competitive cyclists. *Physician Sportsmed* 8, No. 7:78–84, 1980.
7. Burke, E.R., Cerny, F., Costill, D.L., et al. Characteristics of skeletal muscle in competitive cyclists. *Med Sci Sports* 9:109–112, 1977.
8. Burke, E.R., Brush, F.C. Physiological and anthropometric assessment of successful teenage female distance runners. *Res Quart* 50:180–187, 1979.
9. Butts, N.K. Physiological profiles of high school female cross country runners. *Res Quart* 53:8–14, 1982.
10. Carey, P., Stensland, M., Hartley, L.H. Comparison of oxygen uptake during maximal work on the treadmill and the rowing ergometer. *Med Sci Sports* 6:101–103, 1974.
11. Charbonnier, J.P., Lacour, J.R., Riffat, J., et al. Experimental study of the performance of competition swimmers. *Europ J Appl Physiol* 34:157–167, 1975.
12. Christensen, E.H. Beitrage zur physiologie schwerer korperlicher arbeit. *Arbeitphysiol* 4:154–174, 1931.
13. Cohen, J.L., Segal, K.R., Witriol, I., et al. Cardiorespiratory responses to ballet exercise and the $\dot{V}O_2$max of elite ballet dancers. *Med Sci Sports Exercise* 14:212–217, 1982.
14. Conger, P.R., Macnab, R.B.J. Strength, body composition and work capacity of participants and non-participants in women's intercollegiate sports. *Res Quart* 38:184–192, 1967.
15. Conlee, R.K., McGown, C.M., Fisher, A.G., et al. Physiological effects of power volleyball. *Physician Sportsmed* 10, No. 2:93–97, 1982.
16. Costill, D.L. Use of a swimming ergometer in physiological research. *Res Quart* 37:564–567, 1966.
17. Costill, D.L. Metabolic responses during distance running. *J Appl Physiol* 28:251–255, 1970.
18. Costill, D.L., Fink, W.J., Pollock, M.L. Muscle fiber composition and enzyme activities of elite distance runners. *Med Sci Sports* 8:96–100, 1976.
19. Costill, D.L., Winrow, E. Maximal oxygen consumption among marathon runners. *Arch Phys Med* 51:317–320, 1970.
20. Crews, D., Wells, C.L., Burkett, L., et al. A physiological profile of four wheelchair marathon racers. *Physician Sportsmed* 10, No. 6:134–143, 1982.
21. Cunningham, D.A., Eynon, R.B. The working capacity of young competitive swimmers, 10–16 years of age. *Med Sci Sports* 5:227–231, 1973.
22. Cunningham, D.A., Goode, P.B., Critz, J.B. Cardiorespiratory response to exercise on a rowing and bicycle ergometer. *Med Sci Sports* 7:37–43, 1975.
23. Cunningham, D.A., Telford, P., Swart, G.T. The cardiopulmonary capacities of young hockey players; age 10. *Med Sci Sports* 8:23–25, 1976.

24. Davies, C.T.M., Thompson, M.W. Aerobic performance of female marathon and male ultramarathon athletes. *Eur J Appl Physiol* 41:233–245, 1979.
25. diPrampero, P.E., Pendergast, D.R., Wilson, D.W., et al. Energetics of swimming in man. *J Appl Physiol* 37:1–5, 1974.
26. Ellestad, M.H. *Stress Testing: Principles and Practice.* 2nd ed. Philadelphia: F.A. Davis Company, 1980.
27. Fahey, T.D., Akka, L., Rolph, R. Body composition and $\dot{V}O_2$max of exceptional weight-trained athletes. *J Appl Physiol* 39:559–561, 1975.
28. Franklin, B.A., Kaimal, K.P., Moir, T.W., et al. Characteristics of national-class race walkers. *Physician Sportsmed* 9, No. 9:101–108, 1981.
29. Freund, B.J., Allen, R.D., Wilmore, J.H. A comparison of maximal oxygen uptake on horizontal vs. inclined treadmill protocol before and after an inclined terrain running program. *Inter J Sports Med.* In press.
30. Gale, J.B., Flynn, K.W. Maximal oxygen consumption and relative body fat of high-ability wrestlers. *Med Sci Sports* 6:232–234, 1974.
31. Glaser, R.M., Sawka, M.N., Laubach, L.L., et al. Metabolic and cardiopulmonary responses to wheelchair and bicycle ergometry. *J Appl Physiol* 46:1066–1070, 1979.
32. Gollnick, P.D., Armstrong, R.B., Saubert, C., IV, et al. Enzyme activity and fiber composition in skeletal muscle of untrained and trained men. *J Appl Physiol* 33:312–319, 1972.
33. Hagan, R.D., Smith, M.G., Gettman, L.R. Marathon performance in relation to maximal aerobic power and training indices. *Med Sci Sports Exercise* 13:185–189, 1981.
34. Hagerman, F.C., Hagerman, G.R., Mickelson, T.C. Physiological profiles of elite rowers. *Physician Sportsmed* 7:74–83, 1979.
35. Hagerman, F.C., Lee, W.D. Measurement of oxygen consumption, heart rate, and work output during rowing. *Med Sci Sports* 3:155–160, 1971.
36. Hartung, G.H. Specificity of training as indicated by heart-rate response to exercise. *Perceptual Motor Skills* 36:639–645, 1973.
37. Haymes, E.M., Dickinson, A.L. Characteristics of elite male and female ski racers. *Med Sci Sports Exercise* 12:153–158, 1980.
38. Hermansen, L. Oxygen transport during exercise in human subjects. *Acta Physiol Scan Suppl* 399:1–104, 1973.
39. Hildebrandt, G., Voight, E.-D., Bahn, D., et al. Energy cost of propelling wheelchair at various speeds: Cardiac response and effect on steering accuracy. *Arch Phys Med Rehabil* 51:131–136, 1970.
40. Holmer, I. Oxygen uptake during swimming in man. *J Appl Physiol* 33:502–509, 1972.
41. Kamon, E. Negative and positive work in climbing a laddermill. *J Appl Physiol* 29:1–5, 1970.
42. Knowlton, R.G., Ackerman, K.J., Fitzgerald, P.I., et al. Physiological and performance characteristics of United States Championship class orienteers. *Med Sci Sports Exercise* 12:164–169, 1980.
43. Kovaleski, J.E., Parr, R.B., Hornak, J.E., et al. Athletic profile of women college volleyball players. *Physician Sportsmed* 8:112–118, 1980.
44. Krahenbuhl, G.S., Wells, C.L., Brown, C.H., et al. Characteristics of national and world class female pentathletes. *Med Sci Sports* 11:20–23, 1979.
45. Magel, J.R., Faulkner, J.A. Maximum oxygen uptake of college swimmers. *J Appl Physiol* 22:929–933, 1967.
46. Magel, J.R., Foglia, G.F., McArdle, W.D., et al. Specificity of swim training on maximum oxygen uptake. *J Appl Physiol* 38:151–155, 1975.
47. Maksud, M.G., Wiley, R.L., Hamilton, L.H., et al. Maximal $\dot{V}O_2$, ventilation, and heart rate of Olympic speed skating candidates. *J Appl Physiol* 29:186–190, 1970.
48. Master, A.M. Two-step test of myocardial function. *Am Heart J* 10:495, 1934.
49. Mayers, N., Gutin, B. Physiological characteristics of elite prepubertal cross-country runners. *Med Sci Sports* 11:172–176, 1979.
50. Mickelson, T.C., Hagerman, F.C. Anaerobic threshold measurements of elite oarsmen. *Med Sci Sports Exercise* 14:440–444, 1982.
51. Nagle, F.J. Physiological assessment of maximal performance. In, *Exercise and Sport Sciences Reviews,* Vol. 1, edited by J.H. Wilmore, New York: Academic Press, 1973.
52. Nagle, F.J., Morgan, W.P., Hellickson, R.O., et al. Spotting success traits in Olympic contenders. *Physician Sportsmed* 3:31–36, 1975.
53. Niinimaa, V. Figure skating: What do we know about it? *Physician Sportsmed* 10, No. 1:51–56, 1982.
54. Niinimaa, V., Dyon, M., Shephard, R.J. Performance and efficiency of intercollegiate cross-country skiers. *Med Sci Sports* 10:91–93, 1978.
55. Novak, L.P., Hyatt, R.E., Alexander, J.F. Body composition and physiologic function of athletes. *J Amer Med Assoc* 205:764–770, 1968.
56. Novak, L.P., Magill, L.A., Schutte, J.E. Maximal oxygen intake and body composition of female dancers. *Eur J Appl Physiol* 39:277–282, 1978.
57. Parr, R.B., Wilmore, J.H., Hoover, R., et al. Professional basketball players: Athletic profiles. *Physician Sportsmed* 6:77–84, 1978.
58. Pechar, G.S., McArdle, W.D., Katch, F.I., et al. Specificity of cardiorespiratory adaptation to bicycle and treadmill training. *J Appl Physiol* 36:753–756, 1974.
59. Perrier Study: Fitness in America. New York: Perrier, Great Waters of France, Inc., 1979.
60. Pipes, T.V. Physiological characteristics of elite body builders. *Physician Sportsmed* 7:116–122, 1979.
61. Pipes, T.V. The racquetball pro: A physiological profile. *Physician Sportsmed* 7:91–94, 1979.
62. Pollock, M.L. The quantification of endurance training programs. In, *Exercise and Sport Sciences Reviews,* Vol. 1, edited by J. H. Wilmore, New York: Academic Press, 1973.
63. Pollock, M.L., et al. Arm pedaling as an endurance training regimen for the disabled. *Arch Phys Med Rehabil* 55:418–424, 1974.
64. Pollock, M.L., Miller, H.S., Wilmore, J.H. Physiological characteristics of champion American track athletes 40 to 75 years of age. *J Gerontology* 29:645–649, 1974.
65. Pollock, M.L., Schmidt, D.H. (editors). *Heart Disease and Rehabilitation.* Boston: Houghton Mifflin, 1979.
66. Puhl, J., Case, S., Fleck, S., et al. Physical and physiological characteristics of elite volleyball players. *Res Quart* 53:257–262, 1982.
67. Raven, P.B., Gettman, L.R., Pollock, M.L., et al. A physiological evaluation of professional soccer players. *Brit J Sports Med* 10:209–216, 1976.
68. Ridge, B.R., Pyke, F.S., Roberts, A.D. Responses to kayak ergometer performance after kayak and bicycle ergometer training. *Med Sci Sports* 8:18–22, 1976.
69. Roberts, J.A., Alspaugh, J.W. Specificity of training effects resulting from programs of treadmill running and bicycle ergometer riding. *Med Sci Sports* 4:6–10, 1972.
70. Rubal, B.J., Rosentswieg, J., Hamerly, B. Echocardiographic examination of women collegiate softball champions. *Med Sci Sports Exercise* 13:176–179, 1981.
71. Rummel, J.A., Michael, E.L., Berry, C.A. Physiological response to exercise after space flight—Apollo 7 to Apollo 11. *Aerospace Med* 44:235–238, 1973.
72. Rusko, H., Hara, M., Karvinen, E. Aerobic performance capacity in athletes. *Eur J Appl Physiol* 38:151–159, 1978.
73. Saltin, B., Astrand, P.-O. Maximal oxygen uptake in athletes. *J. Appl Physiol* 23:353–358, 1967.
74. Saltin, B., Rowell, L.B. Functional adaptations to physical activity and inactivity. *Federation Proc* 39:1506–1513, 1980.
75. Seliger, V., Kostaka, V., Grusova, D., et al. Energy expenditure and physical fitness of ice hockey players. *Int Z Angew Physiol* 30:283–291, 1972.

76. Shephard, R.J., Godin, G., Campbell, R. Characteristics of sprint, medium and long-distance swimmers. *Eur J Appl Physiol* 32:99–103, 1974.

77. Sidney, K., Shephard, R.J. Physiological characteristics and performance of the whitewater paddler. *Eur J Appl Physiol* 32:55–70, 1973.

78. Simri, U. Assessment procedures for human performance. In, *Fitness, Health, and Work Capacity: International Standards for Assessment.* Edited by L.A. Larson. New York: Macmillan, 1974.

79. Sinning, W.E. Body composition, cardiovascular function, and rule changes in women's basketball. *Res Quart* 44:313–321, 1973.

80. Smith, D.P., Byrd, R.J. Body composition, pulmonary function and maximal oxygen consumption of college football players. *J Sports Med* 16:301–308, 1976.

81. Song, T.M.K. Relationship of physiological characteristics to skiing performance. *Physician Sportsmed* 10, No. 12:96–102, 1982.

82. Spence, D.W., Disch, J.G., Fred, H.L., et al. Descriptive profiles of highly skilled women volleyball players. *Med Sci Sports Exercise* 12:299–302, 1980.

83. Sprynarova, S., Parizkova, J. Functional capacity and body composition in top weightlifters, swimmers, runners and skiers. *Int Z Angew Physiol* 29:181–194, 1971.

84. Stine, G., Ratliff, R., Shierman, G., et al. Physical profile of the wrestlers at the 1977 NCAA championships. *Physician Sportsmed* 7:98–105, 1979.

85. Stromme, S.B., Ingjer, F., Meen, H.D. Assessment of maximal aerobic power in specifically trained athletes. *J Appl Physiol* 42:833–837, 1977.

86. Taylor, A.W., Brassard, L., Proteau, L., et al. A physiological profile of Canadian Greco-Roman wrestlers. *Can J Appl Sport Sci* 4:131–134, 1979.

87. Taylor, H.L., Buskirk, E., Henschel, A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J Appl Physiol* 8:73–80, 1955.

88. Thomas, T.R., Etheridge, G.L. The effect of track-and-field training on cardiovascular fitness. *Physician Sportsmed* 9, No. 2:48–60, 1981.

89. Vaccaro, P., Clarke, D.H., Morris, A.F. Physiological characteristics of young well-trained swimmers. *Eur J Appl Physiol* 44:61–66, 1980.

90. Vaccaro, P., Clarke, D.H., Wrenn, J.P. Physiological profiles of elite women basketball players. *J Sports Med* 19:45–54, 1979.

91. Vaccaro, P., Morris, A.F., Clarke, D.H. Physiological characteristics of masters female distance runners. *Physician Sportsmed* 9, No. 7:105–108, 1981.

92. Vodak, P.A., Savin, W.M., Haskell, W.L., et al. Physiological profile of middle-aged male and female tennis players. *Med Sci Sports Exercise* 12:159–163, 1980.

93. Vrijens, J., et al. Effects of training on maximal working capacity and haemodynamic response during arm and leg-exercise in a group of paddlers. *Eur J Appl Physiol* 34:113–119, 1975.

94. World Health Organization. Exercise Tests in Relation to Cardiovascular Function. Geneva: WHO, 1968 (Technical Reports Series No. 388).

95. Wilmore, J.H. *Training for Sport and Activity: The Physiological Basis of the Conditioning Process.* 2nd ed. Boston: Allyn & Bacon, 1982.

96. Wilmore, J.H. Inferiority of female athletes: Myth or reality? *Am J Sports Med* 3:1–6, 1975.

97. Wilmore, J.H., Brown, C.H. Physiological profiles of women distance runners. *Med Sci Sports* 6:178–181, 1974.

98. Wilmore, J.H., Parr, R.B., Haskell, W.L., et al. Athletic profile of professional football players. *Physician Sportsmed* 4:45–54, 1976.

Table B–2. Relative body fat in male and female athletes

Athletic group	Sex	Age (yr)	Height (cm)	Weight (kg)	Relative fat (percent)	Reference[1]
Baseball	Male	20.8	182.7	83.3	14.2	Novak et al., 1968
	Male	—	—	—	11.8	Forsyth and Sinning, 1973
	Male	27.4	183.1	88.0	12.6	Wilmore, unpublished
Basketball	Female	19.1	169.1	62.6	20.8	Sinning, 1973
	Female	19.4	167.0	63.9	26.9	Conger and Macnab, 1967
Centers	Male	27.7	214.0	109.2	7.1	Parr et al., 1978
Forwards	Male	25.3	200.6	96.9	9.0	Parr et al., 1978
Guards	Male	25.2	188.0	83.6	10.6	Parr et al., 1978
Canoeing	Male	23.7	182.0	79.6	12.4	Rusko et al., 1978
Football	Male	20.3	184.9	96.4	13.8	Novak et al., 1968
	Male	—	—	—	13.9	Forsyth and Sinning, 1973
Defensive backs	Male	17–23	178.3	77.3	11.5	Wickkiser and Kelly, 1975
	Male	24.5	182.5	84.8	9.6	Wilmore et al., 1976
Offensive backs	Male	17–23	179.7	79.8	12.4	Wickkiser and Kelly, 1975
	Male	24.7	183.8	90.7	9.4	Wilmore et al., 1976
Linebackers	Male	17–23	180.1	87.2	13.4	Wickkiser and Kelly, 1975
	Male	24.2	188.6	102.2	14.0	Wilmore et al., 1976
Offensive linemen	Male	17–23	186.0	99.2	19.1	Wickkiser and Kelly, 1975
	Male	24.7	193.0	112.6	15.6	Wilmore et al., 1976
Defensive linemen	Male	17–23	186.6	97.8	18.5	Wickkiser and Kelly, 1975
	Male	25.7	192.4	117.1	18.2	Wilmore et al., 1976
Quarterbacks, kickers	Male	24.1	185.0	90.1	14.4	Wilmore et al., 1976
Gymnastics	Male	20.3	178.5	69.2	4.6	Novak et al., 1968
	Female	19.4	163.0	57.9	23.8	Conger and MacNab, 1967
	Female	20.0	158.5	51.5	15.5	Sinning and Lindberg, 1972
	Female	14.0	—	—	17.0	Parizkova, 1972
	Female	23.0	—	—	11.0	Parizkova, 1972
	Female	23.0	—	—	9.6	Parizkova and Poupa, 1963
Ice hockey	Male	26.3	180.3	86.7	15.1	Wilmore, unpublished
	Male	22.5	179.0	77.3	13.0	Rusko et al., 1978
Jockeys	Male	30.9	158.2	50.3	14.1	Wilmore, unpublished
Orienteering	Male	31.2	—	72.2	16.3	Knowlton et al., 1980
	Female	29.0	—	58.1	18.7	Knowlton et al., 1980
Pentathlon	Female	21.5	175.4	65.4	11.0	Krahenbuhl et al., 1979
Racquetball	Male	25.0	181.7	80.3	8.1	Pipes, 1979
Rowing:						
Heavyweight	Male	23.0	192.0	88.0	11.0	Hagerman et al., 1979
Lightweight	Male	21.0	186.0	71.0	8.5	Hagerman et al., 1979
	Female	23.0	173.0	68.0	14.0	Hagerman et al., 1979
Skiing	Male	25.9	176.6	74.8	7.4	Sprynarova and Parizkova, 1971
Alpine	Male	21.2	176.0	70.1	14.1	Rusko et al., 1978
	Male	21.8	177.8	75.5	10.2	Haymes and Dickinson, 1980
	Female	19.5	165.1	58.8	20.6	Haymes and Dickinson, 1980
Cross-country	Male	21.2	176.0	66.6	12.5	Niinimaa et al., 1978
	Male	25.6	174.0	69.3	10.2	Rusko et al., 1978
	Male	22.7	176.2	73.2	7.9	Haymes and Dickinson, 1980
	Female	24.3	163.0	59.1	21.8	Rusko et al., 1978
	Female	20.2	163.4	55.9	15.7	Haymes and Dickinson, 1980
Nordic combination	Male	22.9	176.0	70.4	11.2	Rusko et al., 1978
	Male	21.7	181.7	70.4	8.9	Haymes and Dickinson, 1980
Ski jumping	Male	22.2	174.0	69.9	14.3	Rusko et al., 1978
Soccer	Male	26.0	176.0	75.5	9.6	Raven et al., 1976
Speed skating	Male	21.0	181.0	76.5	11.4	Rusko et al., 1978
Swimming	Male	21.8	182.3	79.1	8.5	Sprynarova and Parizkova, 1971
	Male	20.6	182.9	78.9	5.0	Novak et al., 1968
	Female	19.4	168.0	63.8	26.3	Conger and Macnab, 1967
Sprint	Female	—	165.1	57.1	14.6	Wilmore et al., 1977
Middle distance	Female	—	166.6	66.8	24.1	Wilmore et al., 1977
Distance	Female	—	166.3	60.9	17.1	Wilmore et al., 1977
Tennis	Male	—	—	—	15.2	Forsyth and Sinning, 1973
	Male	42.0	179.6	77.1	16.3	Vodak et al., 1980
	Female	39.0	163.3	55.7	20.3	Vodak et al., 1980
Track and field	Male	21.3	180.6	71.6	3.7	Novak et al., 1968
	Male	—	—	—	8.8	Forsyth and Sinning, 1973
Run	Male	22.5	177.4	64.5	6.3	Sprynarova and Parizkova, 1971

Table B-2. Relative body fat in male and female athletes—Continued

Athletic group	Sex	Age (yr)	Height (cm)	Weight (kg)	Relative fat (percent)	Reference[1]
Track and field—Con.						
Distance	Male	26.1	175.7	64.2	7.5	Costill et al., 1970
	Male	26.2	177.0	66.2	8.4	Rusko et al., 1978
	Male	40–49	180.7	71.6	11.2	Pollock et al., 1974
	Male	55.3	174.5	63.4	18.0	Barnard et al., 1979
	Male	50–59	174.7	67.2	10.9	Pollock et al., 1974
	Male	60–69	175.7	67.1	11.3	Pollock et al., 1974
	Male	70–75	175.6	66.8	13.6	Pollock et al., 1974
	Male	47.2	176.5	70.7	13.2	Lewis et al., 1975
	Female	19.9	161.3	52.9	19.2	Malina et al., 1971
	Female	32.4	169.4	57.2	15.2	Wilmore and Brown, 1974
Middle distance	Male	24.6	179.0	72.3	12.4	Rusko et al., 1978
Sprint	Female	20.1	164.9	56.7	19.3	Malina et al., 1971
	Male	46.5	177.0	74.1	16.5	Barnard et al., 1979
Discus	Male	28.3	186.1	104.7	16.4	Fahey et al., 1975
	Male	26.4	190.8	110.5	16.3	Wilmore, unpublished
	Female	21.1	168.1	71.0	25.0	Malina et al., 1971
Jumping and hurdling	Female	20.3	165.9	59.0	20.7	Malina et al., 1971
Shot-put	Male	27.0	188.2	112.5	16.5	Fahey et al., 1975
	Male	22.0	191.6	126.2	19.6	Behnke and Wilmore, 1974
	Female	21.5	167.6	78.1	28.0	Malina et al., 1971
Volleyball	Female	19.4	166.0	59.8	25.3	Conger and Macnab, 1967
	Female	19.9	172.2	64.1	21.3	Kovaleski et al., 1980
Weight lifting	Male	24.9	166.4	77.2	9.8	Sprynarova and Parizkova, 1971
Power	Male	26.3	176.1	92.0	15.6	Fahey et al., 1975
Olympic	Male	25.3	177.1	88.2	12.2	Fahey et al., 1975
Body building	Male	29.0	172.4	83.1	8.4	Fahey et al., 1975
	Male	27.6	178.8	88.1	8.3	Pipes, 1979
Wrestling	Male	26.0	177.8	81.8	9.8	Fahey et al., 1975
	Male	27.0	176.0	75.7	10.7	Gale and Flynn, 1974
	Male	22.0	—	—	5.0	Parizkova, 1972
	Male	23.0	—	79.3	14.3	Taylor et al., 1979
	Male	19.6	174.6	74.8	8.8	Sinning, 1974
	Male	15–18	172.3	66.3	6.9	Katch and Michael, 1971
	Male	20.6	174.8	67.3	4.0	Stine et al., 1979

[1] Complete references are presented in the following list.

References for table B–2

Citations in the reference list may not appear in the table.

1. Adams, J., M. Mottola, K. M. Bagnall, and K. D. McFadden. Total body fat content in a group of professional football players. *Can. J. Appl. Spt. Sci.* 7:36–40, 1982.
2. Askew, E. W., H. Barakat, G. L. Kuhl, and G. L. Dohm. Response of lipogenesis and fatty acid synthetase to physical training and exhaustive exercise in rats. *Lipids* 10:491–496, 1975.
3. Barnard, R. J., G. K. Grimditch, and J. H. Wilmore. Physiological characteristics of sprint and endurance masters runners. *Med. Sci. Sports* 11:167–171, 1979.
4. Behnke, A. R., and J. H. Wilmore. *Evaluation and Regulation of Body Build and Composition.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1974.
5. Belbeck, L. W., and J. B. Critz. Effect of exercise on the plasma concentration of anorexigenic substance in man. *Proc. Soc. Exp. Biol. Med.* 142:19–21, 1973.
6. Boileau, R. A., E. R. Buskirk, D. H. Horstman, J. Mendez, and W. C. Nicholas. Body composition changes in obese and lean men during physical conditioning. *Med. Sci. Sports* 3:183–189, 1971.
7. Brobeck, J.R. Food intake as a mechanism of temperature regulation. *Yale J. Biol. Med.* 20:545, 1948.
8. Brown, C.H., and J.H. Wilmore. The effects of maximal resistance training on the strength and body composition of women athletes. *Med. Sci. Sports* 6:174–177, 1974.
9. Brozek, J., and A. Keys. The evaluation of leanness-fatness in man: Norms and interrelationships. *Br. J. Nutr.* 5:194–206, 1951.
10. Brozek, J., F. Grande, J. T. Anderson, and A. Keys. Densitometric analysis of body composition: Revision of some quantitative assumptions. *Ann. N.Y. Acad. Sci.* 110:113–140, 1963.
11. Carter, J. E. L., and W. H. Phillips. Structural changes in exercising middle-aged males during a 2-year period. *J. Appl. Physiol.* 27:787–794, 1969.
12. Conger, P. R., and R. B. J. Macnab. Strength, body composition and work capacity of participants and nonparticipants in women's intercollegiate sports. *Res. Quart.* 38:184–192, 1967.
13. Costill, D. L., R. Bowers, and W. F. Kammer. Skinfold estimates of body fat among marathon runners. *Med. Sci. Sports* 2:93–95, 1970.
14. Deb, S., and R. J. Martin. Effects of exercise and of food restriction on the development of spontaneous obesity in rats. *J. Nutr.* 105:543–549, 1975.
15. Dempsey, J. A. Anthropometrical observations on obese and nonobese young men undergoing a program of vigorous physical exercise. *Res. Quart.* 35:275–287, 1964.
16. Fahey, T. D., and C. H. Brown. The effects of anabolic steroid on the strength, body composition, and endurance of college males

when accompanied by a weight training program. *Med. Sci. Sports* 5:272–296, 1973.

17. Fahey, T. D., L. Akka, and R. Rolph. Body composition and $\dot{V}O_2$ max of exceptional weight-trained athletes. *J. Appl. Physiol.* 39:559–561, 1975.
18. Forsyth, H. L., and W. E. Sinning. The anthropometric estimation of body density and lean body weight of male athletes. *Med. Sci. Sports* 5:174–180, 1973.
19. Franklin, B. A., P. C. Mackeen, and E. R. Buskirk. Body composition effects of a 12-week physical conditioning program for normal and obese middle-aged women, and status at 18-month follow-up. *Int. J. Obesity* 2:394, 1978.
20. Gale, J. B., and K. W. Flynn. Maximal oxygen consumption and relative body fat of high-ability wrestlers. *Med. Sci. Sports* 6:232–234, 1974.
21. Garfield, D. S., P. Ward, R. Cobb, J. Disch, and D. Southwick. Circuit weight training. Houston: Dynamics Health Equipment, 1979.
22. Garrow, J.S. *Energy Balance and Obesity in Man.* 2nd edition. New York: Elsevier/North Holland Biomedical Press, 1978.
23. Getchell, L. H., and J. C. Moore. Physical training: Comparative responses of middle-aged adults. *Arch. Phys. Med. Rehab.* 56:250–254, 1975.
24. Gettman, L. R., J. J. Ayres, M. L. Pollock, and A. Jackson. The effect of circuit weight training on strength, cardiorespiratory function, and body composition of adult men. *Med. Sci. Sports* 10:171–176, 1978.
25. Gettman, L. R., J. J. Ayres, M. L. Pollock, J. L. Durstine, and W. Grethman. Physiological effects on adult men of circuit strength training and jogging. *Arch. Phys. Med. Rehab.* 60:115–120, 1979.
26. Girandola, R. N. Body composition changes in women: Effects of high and low exercise intensity. *Arch. Phys. Med. Rehab.* 57:297–300, 1976.
27. Girandola, R. N., and V. Katch. Effects of nine weeks of physical training on aerobic capacity and body composition in college men. *Arch. Phys. Med. Rehab.* 54:521–524, 1973.
28. Hagerman, F. C., G. R. Hagerman, and T. C. Mickelson. Physiological profiles of elite rowers. *Physician Sportsmed.* No. 7:74–83, 1979.
29. Haymes, E. M., and A. L. Dickinson. Characteristics of elite male and female ski racers. *Med. Sci. Sports Exercise* 12:153–158, 1980.
30. Holloszy, J. O., J. S. Skinner, G. Toro, and T. K. Cureton. Effects of a six month program of endurance exercise on the serum lipids of middle-aged men. *Am. J. Cardiol.* 14:753–760, 1964.
31. Jackson, A. S., and M. L. Pollock. Generalized equations for predicting body density of men. *Br. J. Nutr.* 40:497–504, 1978.
32. Jackson, A. S., M. L. Pollock, and A. Ward. Generalized equations for predicting body density of women. *Med. Sci. Sports Exercise* 12:175–182, 1980.
33. Jankowski, L. W., and M. L. Foss. The energy intake of sedentary men after moderate exercise. *Med. Sci. Sports* 4:11–13, 1972.
34. Johnson, R. E., J. A. Mastropaolo, and M. A. Wharton. Exercise, dietary intake, and body composition. *J. Am. Diet Assoc.* 61:399–403, 1972.
35. Katch, F. I., and E. D. Michael. Body composition of high school wrestlers according to age and wrestling weight category. *Med. Sci. Sports* 3:190–194, 1971.
36. Katch, F. I., E. D. Michael, and E. M. Jones. Effects of physical training on the body composition and diet of females. *Res. Quart.* 40:99–104, 1969.
37. Katch, V. L., R. Martin, and J. Martin. Effects of exercise intensity on food consumption in the male rat. *Am. J. Clin. Nutr.* 32:1401–1407, 1979.
38. Keys, A., and J. Brozek. Body fat in adult man. *Physiol. Rev.* 33:245–325, 1953.
39. Kilbom, A., L. H. Hartley, B. Saltin, J. Bjure, G. Grimby, and I. Astrand. Physical training in sedentary middle-aged and older men. *Scand. J. Clin. Lab. Invest.* 24:315–322, 1969.
40. Knowlton, R. G., K. J. Ackerman, P. I. Fitzgerald, S. W. Wilde, and M. V. Tahamont. Physiological and performance characteristics of United States championship class orienteers. *Med. Sci. Sports Exercise* 12:164–169, 1980.
41. Kollias, J., R. A. Boileau, H. L. Barlett, and E. R. Buskirk. Pulmonary function and physical conditioning in lean and obese subjects. *Arch. Environ. Health* 25:146–150, 1972.
42. Kollias, J., J. S. Skinner, H. L. Barlett, B. S. Bergsteinova, and E. R. Buskirk. Cardiorespiratory responses of young overweight women to ergometry following modest weight reduction. *Arch. Environ. Health* 27:61–64, 1973.
43. Kovaleski, J. E., R. B. Parr, J. E. Hornak, and J. L. Roitman. Athletic profile of women college volleyball players. *Physician Sportsmed.* No. 8:112–118, 1980.
44. Krahenbuhl, G. S., C. L. Wells, C. H. Brown, and P. E. Ward. Characteristics of national and world class female pentathletes. *Med. Sci. Sports* 11:20–23, 1979.
45. Lewis, S., W. L. Haskell, H. Klein, J. Halpern, and P. D. Wood. Prediction of body composition in habitually active middle-aged men. *J. Appl. Physiol.* 39:221–225, 1975.
46. Lewis, S., W. L. Haskell, P. D. Wood, N. Manoogian, J. E. Bailey, and M. Pereira. Effects of physical activity on weight reduction in obese middle-aged women. *Am. J. Clin. Nutr.* 29:151–156, 1976.
47. Malina, R. M., A. B. Harper, H. H. Avent, and D. E. Campbell. Physique of female track and field athletes. *Med. Sci. Sports* 3:32–38, 1971.
48. Martin, A. D., D. T. Drinkwater, J. P. Clarys, and W. D. Ross. Estimation of body fat: A new look at some old assumptions. *Physician Sportsmed.* 9, No. 7:21–22, 1981.
49. Mayer, J. Inactivity, and etiological factors in obesity and heart disease. *Nutrition and Physical Activity.* Symposium of the Swedish Nutrition Foundation V. Edited by Gunnar Blix. Uppsala: Almqvist and Wiksells, 1967.
50. Mayer, J., and B. A. Bullen. Nutrition, weight control, and exercise. In, *Science and Medicine of Exercise and Sport.* 2nd edition. Edited by W. R. Johnson and E. R. Buskirk. New York: Harper and Row, 1974.
51. Mayer, J., N. B. Marshall, J. J. Vitale, J. H. Christensen, M. B. Mashayekhi, and F. J. Stare. Exercise, food intake and body weight in normal rats and genetically obese adult mice. *Am. J. Physiol.* 177:544–548, 1954.
52. Mayer, J., P. Roy, and K. P. Mitra. Relation between caloric intake, body weight, and physical work: Studies in an industrial male population in West Bengal. *Am. J. Clin. Nutr.* 4:169–175, 1956.
53. Mayhew, J. L., and P. M. Gross. Body composition changes in young women with high resistance weight training. *Res. Quart.* 45:433–440, 1974.
54. Misner, J. S., R. A. Boileau, B. H. Massey, and J. L. Mayhew. Alterations in the body composition of adult men during selected physical training programs. *J. Am. Geriat. Soc.* 22:33–38, 1974.
55. Moody, D. L., J. Kollias, and E. R. Buskirk. The effect of a moderate exercise program on body weight and skinfold thickness in overweight college women. *Med. Sci. Sports* 1:75–80, 1969.
56. Moody, D. L., J. H. Wilmore, R. N. Girandola, and J. P. Royce. The effects of a jogging program on the body composition of normal and obese high school girls. *Med. Sci. Sports* 4:210–213, 1972.
57. Niinimaa, V., M. Dyon, and R. J. Shephard. Performance and efficiency of intercollegiate cross-country skiers. *Med. Sci. Sports* 10:91–93, 1978.
58. Novak, L. P., R. E. Hyatt, and J. F. Alexander. Body composition and physiologic function of athletes. *J. Amer. Med. Assoc.* 205:764–770, 1968.
59. Oscai, L. B. The role of exercise in weight control. In, *Exercise and Sport Sciences Reviews,* Vol I. Edited by J. H. Wilmore. New York: Academic Press, 1973.

60. Oscai, L. B., and J. O. Holloszy. Effects of weight changes produced by exercise, food restriction, or overeating on body composition. *J. Clin. Invest.* 48:2124–2128, 1969.

61. Oscai, L. B., and B. T. Williams. Effect of exercise on overweight middle-aged males. *J. Am. Geriat. Soc.* 16:794–797, 1968.

62. Oscai, L. B., S. P. Babirak, F. B. Dubach, J. A. McGarr, and C. N. Spirakis. Exercise or food restriction: Effect on adipose tissue cellularity. *Am. J. Physiol.* 227:901–904, 1974.

63. Oscai, L. B., P. A. Mole, and J. O. Holloszy. Effects of exercise on cardiac weight and mitochondria in male and female rats. *Am. J. Physiol.* 220:1944–1948, 1971.

64. Oscai, L. B., C. N. Spirakis, C. A. Wolff, and R. J. Beck. Effects of exercise and of food restriction on adipose tissue cellularity. *J. Lipid Res.* 13:588–592, 1972.

65. Oscai, L. B., B. T. Williams, and B. Hertig. Effect of exercise on blood volume. *J. Appl. Physiol.* 24:622–624, 1968.

66. Parizkova, J. Body composition and exercise during growth and development. In, *Physical Activity, Human Growth and Development,* edited by G. L. Rarick. New York: Academic Press, 1972, pp. 97–124.

67. Parizkova, J., and O. Poupa. Some metabolic consequences of adaptation to muscular work. *Br. J. Nutr.* 17:341–345, 1963.

68. Parr, R. B., J. H. Wilmore, R. Hoover, D. Bachman, and R. Kerlan. Professional basketball players: Athletic profiles. *Physician Sportsmed.* 6:77–84, 1978.

69. Pipes, T. V. The racquetball pro: A physiological profile. *Physician Sportsmed.* 7, No. 10:91–94, 1979.

70. Pipes, T. V. Physiological characteristics of elite body builders. *Physician Sportsmed.* 7, No. 3:116–122, 1979.

71. Pollock, M. L., J. Broida, Z. Kendrick, H. S. Miller, R. Janeway, and A. C. Linnerud. Effects of training two days per week at different intensities on middle-aged men. *Med. Sci. Sports* 4:192–197, 1972.

72. Pollock, M. L., T. K. Cureton, and L. Greninger. Effects of frequency of training on working capacity, cardiovascular function, and body composition of adult men. *Med. Sci. Sports* 1:70–74, 1969.

73. Pollock, M. L., G. A. Dawson, H. S. Miller, Jr., A. Ward, D. Cooper, W. Headley, A. C. Linnerud, and M. M. Nomeir. Physiologic responses of men 49 to 65 years of age to endurance training. *J. Am. Geriatr. Soc.* 24:97–104, 1976.

74. Pollock, M. L., H. S. Miller, and J. Wilmore. Physiological characteristics of champion American track athletes 40 to 75 years of age. *J. Gerontology* 29:645–649, 1974.

75. Pollock, M. L., H. S. Miller, R. Janeway, A. C. Linnerud, B. Robertson, and R. Valentino. Effects of walking on body composition and cardiovascular function of middle-aged men. *J. Appl. Physiol.* 30:126–130, 1971.

76. Pollock, M. L., H. S. Miller, Jr., Z. Kendrick, and A. C. Linnerud. Effects of mode of training on cardiovascular functions and body composition of adult men. *Med. Sci. Sports* 7:139–145, 1975.

77. Pollock, M. L., J. Tiffany, L. Gettman, R. Janeway, and H. Lofland. Effects of frequency of training on serum lipids, cardiovascular function, and body composition. In, *Exercise and Fitness.* Edited by B. D. Franks. Chicago: Athletic Institute, 1969, pp. 161–178.

78. Raven, P. B., L. R. Gettman, M. L. Pollock, and K. H. Cooper. A physiological evaluation of professional soccer players. *British J. Sports Med.* 10:209–216, 1976.

79. Ribisl, P. M. Effects of training upon the maximal oxygen uptake of middle-aged men. *Int. Z. Angew. Physiol.* 27:154–160, 1969.

80. Rusko, H., M. Hara, and E. Karvinen. Aerobic performance capacity in athletes. *Europ. J. Appl. Physiol.* 38:151–159, 1978.

81. Sinning, W. E. Body composition assessment of college wrestlers. *Med. Sci. Sports* 6:139–145, 1974.

82. Sinning, W. E. Body composition, cardiovascular function, and rule changes in women's basketball. *Res. Quart.* 44:313–321, 1973.

83. Sinning, W. E. Use and misuse of anthropometric estimates of body composition. *J. Phys. Ed. Rec.* 51:43–45, 1980.

84. Sinning, W. E., and G. D. Lindberg. Physical characteristics of college age women gymnasts. *Res. Quart.* 43:226–234, 1972.

85. Siri, W. E. Gross composition of the body. In, *Advances in Biological and Medical Physics, IV.* J. H. Lawrence and C. A. Tobias, eds. New York: Academic Press, 1956.

86. Skinner, J. S., J. O. Holloszy, and T. K. Cureton. Effects of a program of endurance exercises on physical work capacity and anthropometric measurements of 15 middle-aged men. *Am. J. Cardiol.* 14:747–752, 1964.

87. Smith, D. P., and F. W. Stransky. The effect of training and detraining on the body composition and cardiovascular response of young women to exercise. *J. Sports Med.* 16:112–120, 1976.

88. Sprynarova, S., and J. Parizkova. Functional capacity and body composition in top weight-lifters, swimmers, runners and skiers. *Int. Z. Angew. Physiol.* 29:184–194, 1971.

89. Stevenson, J. A. F., B. M. Fox, and A. J. Szlavko. *Proc. Soc. Exp. Biol. Med.* 115:424, 1964.

90. Stine, G., R. Ratliff, G. Shierman, and W. A. Grana. Physical profile of the wrestlers at the 1977 NCAA championships. *Physician Sportsmed.* 7, No. 11:98–105, 1979.

91. Taylor, A. W. The effects of different feeding regimens and endurance exercise programs on carbohydrate and lipid metabolism. *Can. J. Appl. Spt. Sci.* 4:126–130, 1979.

92. Taylor, A. W., L. Brassard, L. Proteu, and D. Robin. A physiological profile of Canadian Greco-Roman wrestlers. *Can. J. Appl. Sport Sci.* 4:131–134, 1979.

93. Taylor, A. W., J. Garrod, M. E. McNulty, and D. C. Secord. Regenerating epididymal fat pad cell size and number after exercise training and three different feeding patterns. *Growth* 37:345–354, 1973.

94. Terjung, R. L. Endocrine response to exercise. *Exercise and Sport Sciences Reviews,* Vol. 7. Edited by R. S. Hutton and D. I. Miller. Philadelphia: Franklin Institute Press, 1980.

95. Terjung, R. L., K. M. Baldwin, J. Cooksey, B. Samson, and R. A. Sutter. Cardiovascular adaptations to twelve minutes of mild daily exercise in middle-aged sedentary men. *J. Am. Geriatr. Soc.* 21:164–168, 1973.

96. Thompson, C. W. Changes in body fat, estimated from skinfold measurements of varsity college football players during a season. *Res. Quart.* 30:87–93, 1959.

97. Thompson, C. W., E. R. Buskirk, and R. F. Goldman. Changes in body fat, estimated from skinfold measurements of college basketball and hockey players during a season. *Res. Quart.* 27:418–430, 1956.

98. Vodak, P. A., W. M. Savin, W. L. Haskell, and P. D. Wood. Physiological profile of middle-aged male and female tennis players. *Med. Sci. Sports Exercise* 12:159–163, 1980.

99. Wedgwood, R. J. Inconstancy of the lean body mass. *Ann. N.Y. Acad. Sci.* 110:141–152, 1963.

100. Wickkiser, J. D., and J. M. Kelly. The body composition of a college football team. *Med. Sci. Sports* 7:199–202, 1975.

101. Wilmore, J. H. Alterations in strength, body composition and anthropometric measurements consequent to a 10-week weight training program. *Med. Sci. Sports* 6:133–138, 1974.

102. Wilmore, J. H. Exercise-induced alterations in weight of underweight women. *Arch. Phys. Med. Rehab.* 54:115–119, 1973.

103. Wilmore, J. H. *Training for Sport and Activity: The Physiological Basis of the Conditioning Process.* 2nd ed. Boston: Allyn and Bacon, 1982.

104. Wilmore, J. H., and C. H. Brown. Physiological profiles of women distance runners. *Med. Sci. Sports* 6:178–181, 1974.

105. Wilmore, J. H., and J. J. McNamara. Prevalence of coronary disease risk factors in boys, 8 to 12 years of age. *J. Pediat.* 84:527–533, 1974.

106. Wilmore, J. H., C. H. Brown, and J. A. Davis. Body physique and composition of the female distance runner. *Ann. New York Acad. Sci.* 301:764–776, 1977.

107. Wilmore, J. H., J. A. Davis, R. S. O'Brien, P. A. Vodak, G. R. Walder, and E. A. Amsterdam. Physiological alterations conse-

quent to 20-week conditioning programs of bicycling, tennis, and jogging. *Med. Sci. Sports Exercise* 12:1–8, 1980.

108. Wilmore, J. H., H. L. Miller, and M. L. Pollock. Body composition and physiological characteristics of active endurance athletes in their eighth decade of life. *Med. Sci. Sports* 6:44–48, 1974.
109. Wilmore, J. H., R. B. Parr, R. N. Girandola, P. Ward, P. A. Vodak, T. J. Barstow, T. V. Pipes, G. T. Romero, and P. Leslie. Physiological alterations consequent to circuit weight training. *Med. Sci. Sports* 10:79–84, 1978.
110. Wilmore, J. H., R. B. Parr, W. L. Haskell, D. L. Costill, L. J. Milburn, and R. K. Kerlan. Athletic profile of professional football players. *Physician Sportsmed* 4, No. 10:45–54, 1976.
111. Wilmore, J. H., J. Royce, R. N. Girandola, F. I. Katch, and V. L. Katch. Body composition changes with a 10-week program of jogging. *Med. Sci. Sports* 2:113–117, 1970.
112. Zuti, W. B., and L. A. Golding. Comparing diet and exercise as weight reduction tools. *Physician Sportsmed* 4:49–55, 1976.

Chapter 5

An Integrative Approach to the Noninvasive Assessment of Cardiovascular and Respiratory Function During Exercise

Karlman Wasserman, M.D., Ph.D.
Harbor-UCLA Medical Center
Torrance, California

Exercise testing affords the investigator the opportunity to examine both the cardiovascular and respiratory systems simultaneously under conditions that simulate the metabolic cost of work. To couple external respiration to the respiration of the muscle mitochondria where the chemical energy is produced for muscular contraction, an interaction of physiological mechanisms involving the muscles, peripheral circulation, heart, pulmonary circulation, lungs and chest bellows, and blood is required. Clearly, any disease state that causes one of these interlinked components to malfunction can limit exercise performance. Therefore, in assessing a population such as that being studied in the National Health and Nutrition Examination Survey (NHANES) project, making a single measurement such as heart rate has very limited value. The measurements to be made and the form of ergometry utilized must be thoughtfully considered. The following recommendations are designed to maximize the use of current technology so as to precisely define the pathophysiologic factors limiting exercise in members of the NHANES population while minimizing subject, physician, technician, and analysis time.

The major disorders limiting exercise capacity are ordinarily those of the cardiovascular system, those of the respiratory system, or obesity. In the population being evaluated in the NHANES project, anemia might also be a major contributor to exercise limitation. The aim of this paper is to describe the physiological measurements that might be made in a field laboratory such as NHANES, take advantage of new insights into pathophysiology, and utilize minicomputers for data processing. Protocols that can be used to efficiently test study subjects with minimal testing time are described. Finally, the causes of exercise limitation that can be defined by exercise testing are discussed. First, however, the physiological requirements to perform work are reviewed.

Physiological requirements to perform exercise

Overview

The physiological requirements to perform work involve a functional coupling of accelerated cardiovascular and respiratory activity to achieve higher levels of gas exchange (CO_2 and O_2) between the muscle cells and the atmosphere to accommodate the increased metabolic stress (figure 1). In fact, the primary role of the

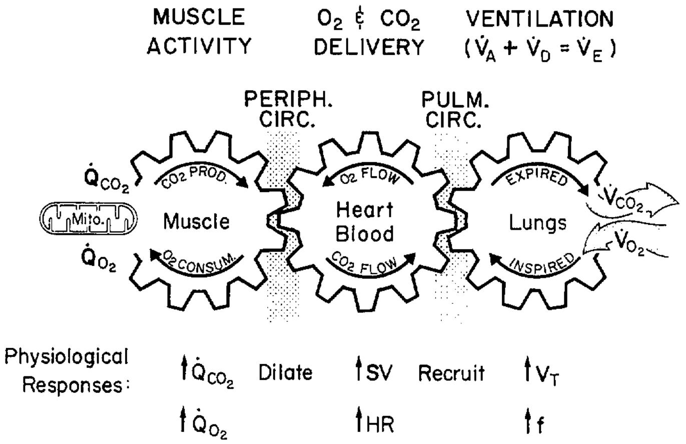


Figure 1. Physiological responses to exercise. Notes: Physiological mechanisms interact to enable oxygen (O_2) uptake ($\dot{V}O_2$) and carbon dioxide (CO_2) output ($\dot{V}CO_2$) to equal muscle O_2 consumption ($\dot{Q}O_2$) and CO_2 production ($\dot{Q}CO_2$). SV = stroke volume; HR = heart rate; V_T = tidal volume; f = breathing rate; $\dot{V}_A$ = alveolar ventilation; V_D = physiological dead-space ventilation; $\dot{V}_E$ = expired minute ventilation. From Wasserman et al. (24).

cardiovascular and ventilatory systems is to support cell respiration through adequate gas exchange.

Normal gas exchange between the cells and the environment requires (a) efficiently operating lungs and chest bellows; (b) effective pulmonary circulation, in which the regional blood flow is matched by an appropriate distribution of ventilation; (c) a heart capable of pumping the quantity of oxygenated blood necessary to sustain tissue energetic processes; (d) an effective system of blood vessels that can selectively distribute blood flow to match tissue gas exchange requirements; and (e) a respiratory control mechanism capable of regulating arterial blood gas tensions and pH. The response of each of these interlinked mechanisms in the healthy individual is quite predictable, depending only on the rate of work being performed and the relative fitness of the subject (1).

Cell respiration

Energy is obtained predominantly from the oxidation of fuel in the mitochondria, with a small additional amount from biochemical mechanisms in the cell cytoplasm (figure 2). This energy is used to form high-energy compounds, predominately adenosine triphosphate (ATP), from which energy stored in the terminal phosphate bond can be made available for cellular reactions involved in synthesis, active transport, and muscle contraction. Exercise entails an acceleration of the energy-yielding reactions in the muscles in order to produce increased amounts of ATP for muscle contraction. This requires an increased utilization of oxygen, which must be met by increased delivery of O_2 to the mitochondria and the simultaneous removal of CO_2, the major catabolic end product of exercise.

Esterification of acetate from the catabolism of carbohydrate, fatty acids, and amino acids (minor) with coenzyme-A (acetyl-CoA) allows acetate to react with oxaloacetate in the mitochondrion to form citrate,

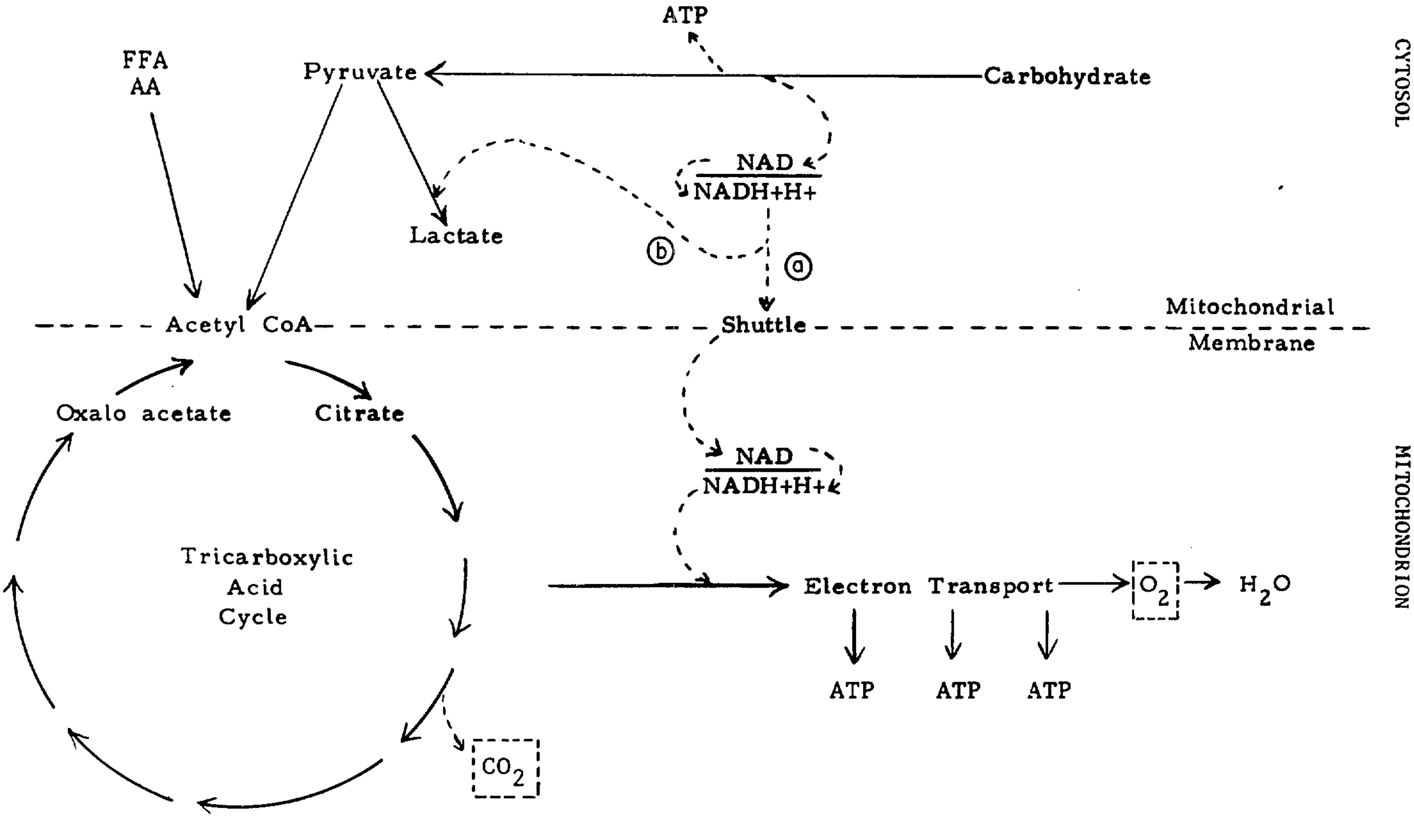


Figure 2. A skeletal version of the biochemical pathways for ATP production. The ability of the "shuttle" mechanisms of the mitochondrial membrane to reoxidize cytoplasmic NAD is a key step in determining lactate accumulation. See text for further details. From Wasserman (25).

thereby entering the tricarboxylic acid (TCA or Krebs) cycle. The subsequent catabolic reactions result in the release of CO_2 and the transfer of a proton pair and its associated electrons down the electron transport (cytochrome) chain to react with cytochrome oxidase and oxygen to form water. For each pair of electrons transferred, energy is released to form approximately 3 ATP molecules. There is a gain of six ATP molecules during the catabolism of glucose to pyruvate if nicotinamide adenine dinucleotide (NAD) in the cytosol, which becomes reduced as a result of glycolysis, is reoxidized by the mitochondrial membrane shuttle (figure 2). This method of regenerating oxidized NAD in the cytosol maintains the cytosol redox state and enables glycolysis to continue (2). The total gain in ATP from the complete oxidation of glucose is 36. Six molecules of O_2 are used for the oxidation steps, and thus three high-energy phosphate bonds are formed for each atom of oxygen consumed ($P{:}O = 3$); there is a concomitant yield of six molecules of CO_2 and H_2O as catabolic end products.

Under conditions in which the mitochondrial FAD fails to reoxidize the shuttle system at a rate sufficient to keep cytosol NADH/NAD normal, the cytosol redox state becomes reduced (figure 2). The mechanism available to buffer this reduction is for pyruvate to react with the reduced cytosol NAD to form lactic acid and reoxidized NAD, an anaerobic oxidation. This is an expensive substrate price to pay compared to the complete oxidation of glucose to CO_2 and H_2O, because the net gain in ATP is only 2 instead of 36, and this causes glucose and glycogen to be utilized at a considerably faster rate. Also, the two lactic acid molecules formed from each glucose molecule cause a disturbance in acid-base balance and, as a consequence, a chain of events is initiated that stimulates breathing and leads to muscle fatigue.

It should be emphasized that the sequence of the turn-on of anaerobic ATP production does not signal the turnoff of aerobic ATP production. O_2 utilization by the mitochondria continues, although there is evidence that the rate of utilization is constrained by inadequate O_2 availability. Providing more O_2 to the metabolizing tissues reduces arterial blood lactate (1, 3).

The anaerobic threshold (AT) is defined as the level of oxygen uptake ($\dot{V}O_2$) above which aerobic energy production is supplemented by anaerobic mechanisms during exercise. Important corollaries of the anaerobic threshold concept are that (a) the O_2 required by the metabolically active muscles can exceed the O_2 supply to the mitochondria when the work rate is sufficiently high; (b) the imbalance between the O_2 supply and O_2 requirement brings about a lowering of the cytosol redox state (increased lactate/pyruvate ratio) associated with increased glucose conversion to lactate; (c) lactate is buffered in the cell primarily by HCO_3^- (figure 3); (d) the CO_2 generated from buffering increases CO_2 output while HCO_3^- exchanges for lactate across the muscle cell membrane, causing blood HCO_3^- to decrease as lactate increases; and (e) the buffering and acid-base disturbances produce predictable changes in gas exchange. The characteristic alterations in gas exchange that occur when the level of work exceeds the anaerobic threshold will be summarized later.

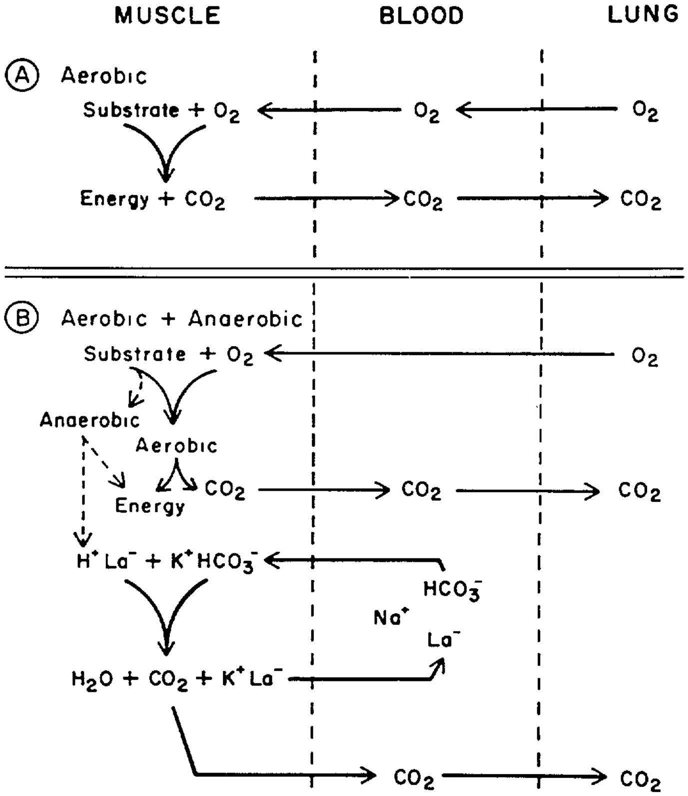


Figure 3. Gas exchange during aerobic exercise (panel A) and aerobic plus anaerobic exercise (panel B). Notes: The acid-base consequence of aerobic plus anaerobic exercise is an increase in cell lactic acid production. The buffering of the newly formed lactic acid takes place in the cell at the site of formation, and predominantly by bicarbonate. The latter mechanisms will increase CO_2 production of the cell by approximately 22 ml per mEq of lactate formed. The increase in cell lactate and decrease in cell bicarbonate will result in chemical concentration gradients, causing lactate to diffuse out of and bicarbonate to diffuse into the cell. From Wasserman et al. (24).

The metabolic-cardiovascular-ventilatory coupling

The large increase in muscle O_2 requirement during exercise demands that O_2 flow to the muscle be increased if muscular contractions are to be sustained. Simultaneously, the large quantity of CO_2 produced in the cell must be removed in order to avoid severe tissue acidosis, with its adverse effects on cell function. Thus, exercise poses a gas transport problem. In order to accomplish the increased gas exchange needed for exercise, a close coupling of physiological mechanisms involved in gas transport involving the lungs, pulmonary circulation, heart, and peripheral circulation is required (figure 1). These subsystems must efficiently accommodate the increased metabolic rate in order to maintain tissue O_2 supply and CO_2 elimination as well as to achieve arterial blood gas homeostasis.

When exercise is initiated, the high-energy stores in the form of creatine phosphate enable muscle contraction to begin immediately. But within seconds, mitochondrial PO_2 must decrease in the metabolically active muscles as O_2 is utilized. To allow this increased O_2 utilization by the muscles, the peripheral vasculature dilates. The initial dilation is neurogenic; subsequent dilation is under local humoral control, enabling blood flow to increase to the muscle units with the highest metabolic activity.

Cardiac output is increased at the start of exercise by an increase in stroke volume and heart rate. Stroke volume appears to increase approximately to maximum as soon as exercise begins. This is accomplished by increased cardiac inotropy and increased venous return secondary to external compression of veins as muscles contract. As exercise progresses, further increases in cardiac output are accomplished almost exclusively by increasing heart rate.

In response to the increase in right ventricular output, the pulmonary vascular bed dilates. This is primarily achieved by perfusing lung units that were unperfused or underperfused at rest. This dilation is essential for the normal exercise response of the left ventricle; without it, the weakly muscled right ventricle could not readily pump the increased venous return through the lungs to the left atrium to effect a normal increase in cardiac output.

Minute ventilation ($\dot{V}_E$) increases at the rate required to remove the CO_2 added to the blood by the tissues. The increase in $\dot{V}_E$ is so precise that arterial isocapnia and isohydria are generally maintained through moderate levels of exercise despite a manifold increase in metabolic rate. Simultaneously, the pulmonary capillary blood is reoxygenated. At work rates above the anaerobic threshold, a metabolic acidosis occurs, which further increases ventilation. The increase in ventilation is accomplished at low and moderate work rates primarily by increasing tidal volume and, to a lesser degree, breathing frequency; the latter increases more significantly at work rates above the anaerobic threshold.

Oxygen cost of work

The oxygen cost of doing work depends on the work rate. Figure 4 shows the $\dot{V}O_2$ for varying levels of cycle ergometer work as related to time for a normal individual. Note that a steady state is reached by 3 minutes at work rates that are not very intense. At high work rates, $\dot{V}O_2$ continues to increase beyond the initial 3 minutes; this upward drift in $\dot{V}O_2$ is seen only for work rates above the anaerobic threshold (4, 5). The higher the work rate above the anaerobic threshold, the greater is the rate of upward drift.

If the steady-state $\dot{V}O_2$ is measured for a range of moderate work rates, a linear relationship between $\dot{V}O_2$ and work rate is obtained, as shown in figure 5. The slope of this relationship is approximately the same for all normal people (10 ml/min/watt).

Work efficiency

Steady-state oxygen consumption and work rate are commonly interchanged when describing the level of

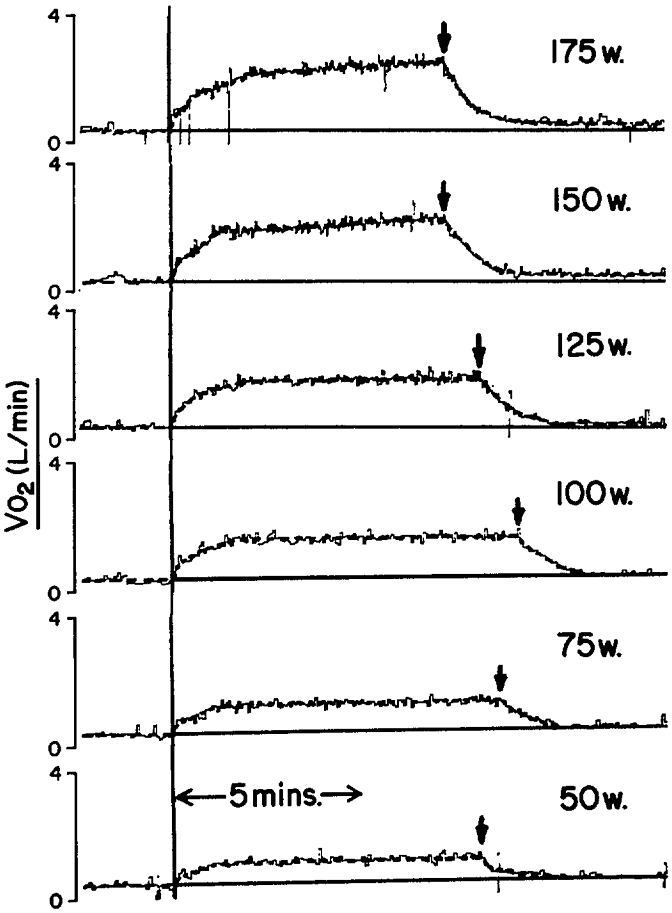


Figure 4. Oxygen uptake kinetics as related to work rate. From Whipp and Wasserman (5).

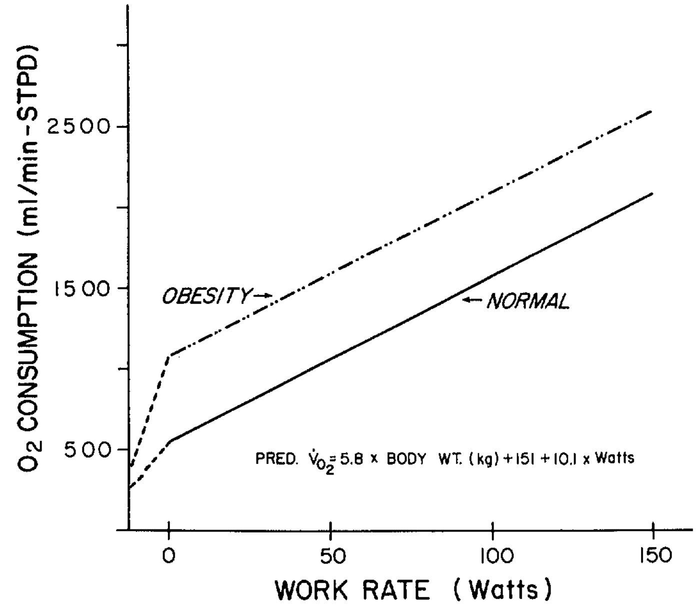


Figure 5. Oxygen (O_2) cost ($\dot{V}O_2$) of performing cycle ergometer work. Slope of work rate – $\dot{V}O_2$ relationship is same for all people and unaffected by training, age, or sex. However, relation is displaced upward by increasing weight. Predicted $\dot{V}O_2$ = 5.8 × body weight (kg) + 151 + 10.1 × watts. STPD indicates standard temperature and pressure, dry. Modified from Wasserman and Whipp (12).

work performed because work efficiency (caloric equivalent of external work divided by caloric equivalent of the oxygen consumption used for work) varies little from individual to individual (6). Thus trained and untrained individuals, old and young, male and female all have similar work efficiencies, suggesting that muscle contraction involves the transduction of common biochemical energy-yielding reactions to achieve physical movement.

To calculate work efficiency, the caloric equivalent of the steady-state oxygen consumption (5 Cal/liter $\dot{V}\hat{O}_2$ at R.Q. = 0.85) and the caloric equivalent of the external work (.014 Cal/min/watt) for at least two work rates must be known (7). When these values are plotted on rectilinear coordinates, the slope is the work efficiency. This is best measured during cycle exercise because the amount of external work can be most accurately estimated with this form of ergometry. For cycle work, normal subjects have an efficiency of approximately 30% (7).

Oxygen supply

The oxygen supply to the metabolically active cells is dependent on five factors: (1) the partial pressure of O_2 in the arterial blood, (2) the hemoglobin concentration, (3) the cardiac output, (4) the distribution of perfusion to the tissues in need of O_2, and (5) the affinity of O_2 for hemoglobin. Each must act to keep capillary $P\hat{O}_2$ at an adequate level to allow diffusion of O_2 from the blood into the muscle mitochondria at a rate sufficient to satisfy the O_2 need. Each of these factors will be discussed separately.

1. *Arterial* PO_2—Mean arterial PO_2 is a function of mean alveolar PO_2. For an idealized lung (all lung units having the same ventilation/perfusion ratio) in which the respiratory quotient is 0.8 and $PaCO_2 = 40$, the P_AO_2 is 100, and PaO_2 is about 90 mmHg. The alveolar-arterial $P\hat{O}_2$ difference of approximately 10 mmHg is caused by a small physiological right-to-left shunt—primarily the Thebesian blood vessels in the heart and the bronchial circulation. A further small increase in the alveolar-arterial $P\hat{O}_2$ difference might result because of a small amount of ventilation/perfusion mismatching present in the normal individual.

   Reductions in arterial $P\hat{O}_2$ relative to alveolar PO_2 are caused by one or more of the following mechanisms: (1) increased right-to-left shunt, (2) diffusion barriers for O_2 at the alveolar-capillary interface, and (3) maldistribution of alveolar ventilation with respect to lung perfusion. The last is clearly the most common cause of hypoxemia.
2. *Arterial* O_2 *content*—The arterial O_2 content depends on the arterial $P\hat{O}_2$ and the hemoglobin content of the arterial blood. Thus, anemias may cause a reduction in the O_2 supply to the tissues as a result of a decreased blood O_2 content.
3. *Cardiac output*—The cardiac output obviously must play a key role in the O_2 supply of the cells. At the start of exercise, stroke volume increases to approximately its maximal value. The size of the stroke volume increase at the start of exercise is dependent on the relative degree of training, age, and build of the individual. (Stroke volume increases by as much as 100% in the exceptionally fit young person but by only 20% in the less fit elderly person.) Further increases in cardiac output are brought about by increasing heart rate. Normally, heart rate increases linearly with $\dot{V}\hat{O}_2$.
4. *Distribution of peripheral blood flow*—During exercise the fraction of the cardiac output that is diverted to the skeletal muscles increases while the fraction of the cardiac output perfusing organs such as the kidney, liver, and gastrointestinal tract decreases. As work rate increases, the vasculature of the exercising muscles dilates further, allowing an even larger fraction of the cardiac output to go through the exercising muscles. The redistribution of perfusion to tissues with high O_2 utilization allows extraction of O_2 from the blood to increase. A useful noninvasive measurement for monitoring the redistribution is the $\dot{V}\hat{O}_2$/HR ratio (O_2-pulse) because it equals the stroke volume × (arterial-mixed venous O_2 content difference). Thus, the O_2-pulse is the amount of O_2 removed from each stroke volume. When stroke volume and the arterial-venous O_2 difference reach their maximums, the O_2-pulse will fail to increase despite increasing work rate.
5. *Altered* O_2*-hemoglobin affinity*—For the normal lung, O_2-hemoglobin affinity differences seen with certain abnormal hemoglobins or altered acid-base balance have little effect on the O_2 content of the blood leaving the lung because of the almost complete saturation of hemoglobin with O_2 at normal alveolar PO_2 values (at sea level). However, the $P\hat{O}_2$ in the tissue capillary is markedly affected by the position of the steep part of the oxyhemoglobin dissociation curve, i.e., the P_{50} of oxyhemoglobin. A shift to the left can markedly impair tissue O_2 availability; a shift to the right allows O_2 to unload from hemoglobin more readily.

CO_2 clearance

Like O_2 consumption, CO_2 production increases during exercise because of the increased metabolic activity in the exercising muscles. The amount of CO_2 generated from this process is related to O_2 consumption by the metabolic respiratory quotient. The ratio of CO_2 produced to O_2 consumed ranges from 0.7 to 1.0, depending on the relative metabolic consumption of fatty acids, amino acids, and carbohydrates. Ordinarily, this ratio ranges from 0.8 to 0.9 during exercise.

At work rates high enough to cause an increase in blood lactate, additional CO_2 is derived from the buffering of this acid by bicarbonate. The only way CO_2 can be eliminated from the body in significant quantity is by its delivery to the lungs in the venous return. In contrast to tissue oxygen supply, the level of cardiac output is not a critical determinant of CO_2 elimination. It is the quantity of ventilation that determines arterial PCO_2 ($PaCO_2$). The tissue PCO_2, in turn, is defined by the $PaCO_2$, blood flow, and metabolic activity.

Determinants of the ventilatory requirement

The quantity of ventilation required to clear a given amount of CO_2 from the blood ($\dot{V}CO_2$) is dependent on the concentration of CO_2 in the alveolar gas (P_ACO_2/P_B). Mass balance considerations dictate that, in an idealized lung (in which gas concentrations are the same in all alveolar spaces), $\dot{V}CO_2 = \dot{V}_A \times P_ACO_2/P_B$, where $\dot{V}_A$ is the alveolar ventilation of the idealized lung. $\dot{V}_A$ is a theoretical ventilation. It is the ventilation that would be required to yield a given alveolar PCO_2 under conditions of a given CO_2 production if the entire ventilation uniformly participated in gas exchange. This important relationship is plotted in figure 6.

However, not all respired air effectively ventilates the lung because some must go to the conducting airways not involved in gas exchange and some to nonperfused alveoli. The fraction of the tidal volume that is wasted from the point of view of gas exchange (V_D/V_T) determines the difference between the actual volume of air respired during breathing ($\dot{V}_E$) and the theoretical alveolar ventilation. The following equation determines the $\dot{V}_E$ needed to eliminate a given quantity of CO_2:

$$\dot{V}_E\ (\mathrm{BTPS}) = \dot{V}CO_2\ (\mathrm{STPD})/PaCO_2 \times k \times 1/(1 - V_D/V_T),$$

where k is the product of the barometric pressure, temperature, and water vapor correction factors needed to express $\dot{V}_E$ in BTPS. From this equation it is evident that the quantity of breathing required for exercise is defined by three factors: (1) the $\dot{V}CO_2$, (2) the level at which $PaCO_2$ is regulated by the respiratory control mechanisms, and (3) the physiological dead space/tidal volume ratio. The effect of these three determinants is illustrated in figure 7. At work rates above the AT, $\dot{V}_E$ and $\dot{V}_A$ increase more steeply as $\dot{V}CO_2$ increases because of the increased ventilatory drive caused by the metabolic acidosis.

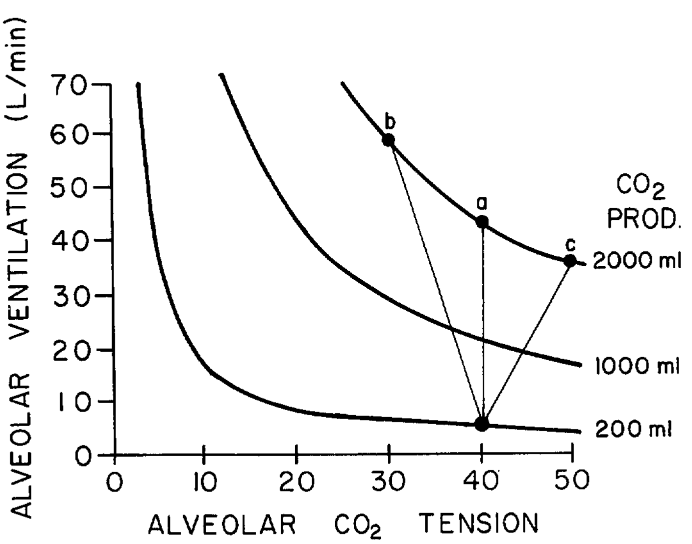


Figure 6. Effect of changing arterial PCO_2 during exercise on alveolar ventilation. Notes: The point on the CO_2 output isopleth of 200 ml/min represents the normal resting value. Points a, b, and c illustrate the alveolar ventilation for isocapnia and hypocapnia (−10 mmHg), and hypercapnia (+10 mmHg) for an exercise CO_2 output of 2,000 ml/min. From Wasserman (14).

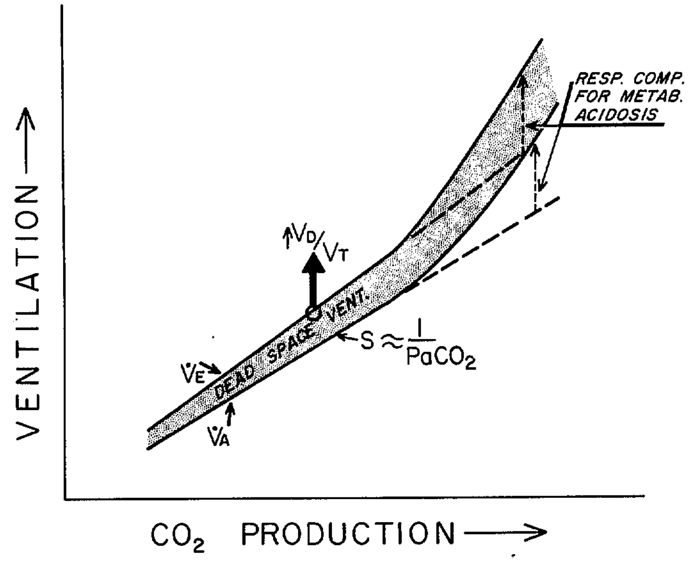


Figure 7. Factors that determine alveolar $\dot{V}_A$ and expired ventilation per minute during exercise ($\dot{V}_A$ and $\dot{V}_E$, respectively). V_D/V_T is the physiological dead space/tidal volume ratio. S is a slope of the curve and is equal to the reciprocal of the arterial PCO_2 (P_aCO_2). Modified from Whipp and Mahler (15).

Control of breathing

Despite a manifold increase in CO_2 production and O_2 utilization during exercise, the ventilatory control mechanisms normally keep arterial PCO_2, H^+, and PO_2 remarkably constant.

Ventilation is manifestly geared to the metabolic requirements of exercise, as previously discussed (figure 1). If $\dot{V}_E$ did not increase in proportion to the rate of CO_2 production during exercise, a respiratory acidosis with associated disturbances in transmembrane potentials and cellular function would result. On the other hand, if ventilation increased out of proportion to metabolism, respiratory alkalosis would result, and this would also impair cell function. However, arterial pH is relatively well controlled during moderate work intensities, and metabolic acidosis occurs only for heavy work intensities (figure 8) because of increased lactic acid production. Respiratory acidosis is present only in patients with abnormal respiratory mechanics or impaired chemoreceptor function. In contrast, respiratory alkalosis does not develop during exercise in normal subjects and is only rarely seen in pathophysiologic states.

Despite intensive research, there is no general agreement on the mechanisms of respiratory control during exercise. However, it appears that the carotid bodies play a role in the tight control of pH, $PaCO_2$, and PaO_2 during

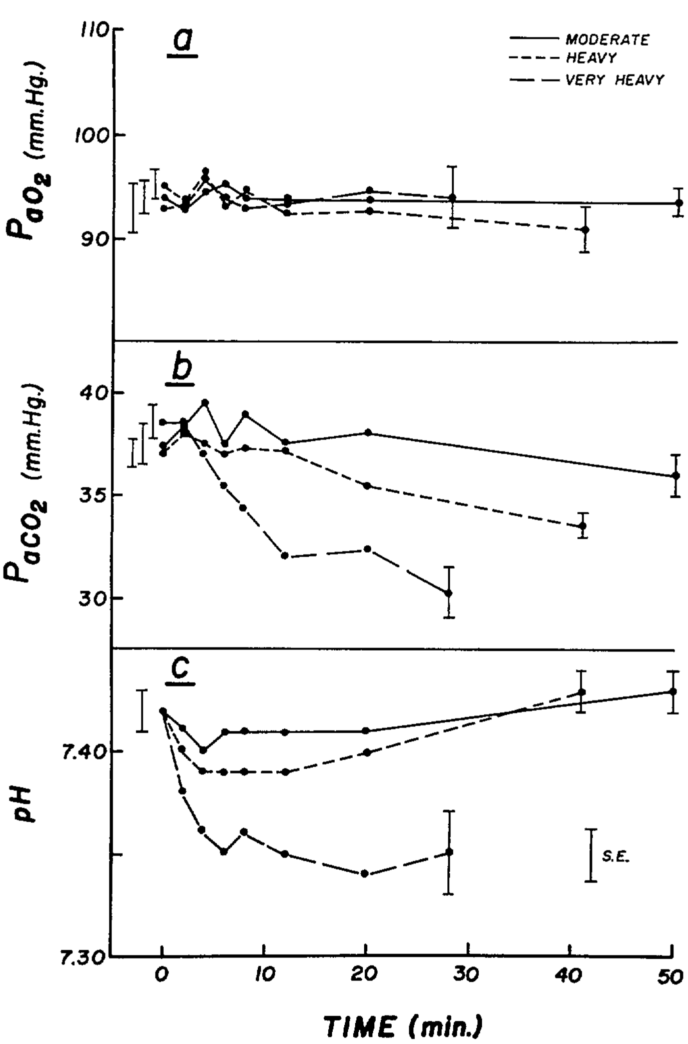


Figure 8. Effect of prolonged constant work rate exercise on arterial blood gases and pH for moderate, heavy, and very heavy work intensity. Each point is the average of 10 subjects, and the bars show plus or minus 1 standard error of the mean for the resting values and the last sampling point during exercise. From Wasserman, Van Kessel, and Burton (26).

exercise. For a more indepth discussion of respiratory control during exercise, see the review by Whipp (8).

Measures of exercise capacity

Exercise testing enables the simultaneous evaluation of the ability of the cardiovascular and respiratory systems to perform their major function, i.e., gas exchange. Noninvasive measurements that have been found to discriminate the adequacy of performance of the components of the metabolic-cardiovascular-ventilatory coupling during exercise are described in this section. The measurements and the functions that they describe are listed in table 1. Accurate measurements can be readily obtained by proper calibration of volume and gas exchange analyzers and with the help of minicomputers programmed to store data and to subsequently provide well-formatted summary reports.

Measurements during incremental exercise testing are very useful because they enable the examiner to (1) titrate the level of the subject's exercise limitation, (2) determine the adequacy of the performance of various components in the external-internal gas exchange coupling, and (3) pinpoint the organ system limiting exercise performance. All of these are best achieved during short non-steady-state rather than prolonged steady-state exercise tests.

The electrocardiogram

The electrocardiogram (ECG) is a valuable measure of the balance between myocardial O_2 availability and O_2 requirement for cardiac work. When the heart muscle contracts without adequate oxygen, it becomes ischemic, causing the muscle cells to alter their transmembrane ion balance (9). Consequently, the electrical charge is altered and abnormal repolarization occurs in the ischemic areas. This causes the T wave and ST segments of the electrocardiogram to change acutely. In addition to T-wave changes and ST-segment shifts, ectopic beats increasing in frequency as the work rate increases should be considered pathologic and are usually indicative of coronary artery disease. Some subjects, however, manifest premature ventricular or atrial contractions at rest that disappear or decrease in frequency during exercise. These ectopic beats are probably benign and unrelated to a disturbance in the myocardial O_2 availability-requirement balance. This benign ectopy can generally be overridden by the sinus tachycardia of exercise.

Maximal $\dot{V}O_2$ ($\dot{V}O_2$max) and maximum $\dot{V}O_2$

The maximal O_2 uptake, or the maximal aerobic power ($\dot{V}O_2$max), is a measure of the greatest capacity of the circulation to deliver O_2 to the tissues under conditions of supramaximal work. $\dot{V}O_2$max was originally defined by the failure of $\dot{V}O_2$ to increase further despite the performance of increasing levels of supramaximal work (10). Often, during an incremental exercise test, a normal subject may put forth his maximal effort, yet the slope of the $\dot{V}O_2$-work rate relationship may not approach a plateau. We use the term *maximum* $\dot{V}O_2$ to describe the subject's highest $\dot{V}O_2$ achieved during an incremental exercise test, whether or not a plateau in $\dot{V}O_2$ was reached, and reserve the term *maximal* $\dot{V}O_2$ ($\dot{V}O_2$max) for the condition in which $\dot{V}O_2$ fails to increase despite increasing work rate. All subjects will have a maximum $\dot{V}O_2$, but not all subjects will be able to stress themselves sufficiently to elicit a $\dot{V}O_2$max.

The $\dot{V}O_2$max and maximum $\dot{V}O_2$ become reduced in any condition where O_2 flow to or O_2 utilization by the tissues is impaired, e.g., cardiovascular diseases, lung diseases, or anemias. Note that the $\dot{V}O_2$max does not describe all of the energy expended by the subject because it does not account for the anaerobic (lactic-acid-generating) contribution to energy production, an important energy source at high work rates (11).

Table 1. Assessing function with physiological measurements

Measurement	Function
Electrocardiogram	Myocardial O_2 availability–O_2 requirement balance
$\dot{V}\hat{O}_2$	Cardiac output × (arterial-mixed venous O_2 content)
Maximum $\dot{V}\hat{O}_2$	Highest $\dot{V}\hat{O}_2$ achieved during presumed maximal effort for an incremental exercise test (specific for type of work)
$\dot{V}\hat{O}_2$ max	Maximal aerobic power (specific for type of work)
$\dot{V}\hat{O}_2$/work rate relationship ($\triangle\dot{V}\hat{O}_2/\triangle WR$) during incremental exercises	Aerobic contribution to exercise (low value suggests high anaerobic contribution)
Anaerobic threshold ($\underline{AT}$)	Maximum $\dot{V}\hat{O}_2$ that can be utilized without developing a metabolic acidosis; important determinant of potential for endurance work (specific for form of work)
O_2-pulse ($\dot{V}\hat{O}_2/HR$) = $SV \times (a - \bar{v})\,O_2$	SV and arterial-mixed venous O_2 difference
$\frac{\text{HR pred max} - \text{HR ex max}}{\text{HR pred max}}$	Heart rate reserve at maximum exercise
$\text{MVV} - \dot{V}_E \text{ max}$ and $\frac{\text{MVV} - \dot{V}_E \text{ max}}{\text{MVV}}$	Breathing reserve
Ventilatory equivalent for O_2 and CO_2	Ventilation/perfusion mismatching and/or hypoventilation
V_T/IC	Restricted lung expansion
$\triangle\dot{V}\hat{O}_2(6-3)$ – constant WR	Relationship of steady-state WR to $\underline{AT}$

A single incremental exercise test to determine the subject's maximum $\dot{V}\hat{O}_2$ has a number of advantages: (1) It starts out at relatively low work rates, so the test does not require the application of great muscle force or sudden, large cardiorespiratory stress; (2) the maximum $\dot{V}\hat{O}_2$ can be determined from a single exercise test, lasting approximately 10 minutes, in which the subject is heavily stressed for only a few minutes; (3) the $\dot{V}\hat{O}_2$-work rate relationship can be determined in which the work rate is accurately known.

The maximum $\dot{V}\hat{O}_2$ establishes whether the subject's physiological responses allow normal maximal aerobic function. Other measurements are needed to differentiate the cause of a reduced maximum $\dot{V}\hat{O}_2$.

$\dot{V}\hat{O}_2$-work rate relationship

Although the O_2 uptake measurements are made at the mouth, they reflect the O_2 utilization by the cells. Because $\dot{V}\hat{O}_2$ equals cardiac output × the arterial-mixed venous oxygen difference (Fick equation), this measurement depends on cardiovascular performance. Factors affecting the position, slope, and linearity of the $\dot{V}\hat{O}_2$-work rate relationship are discussed next.

Obesity

The position of the $\dot{V}\hat{O}_2$-work rate relationship is dependent on body weight, as shown in figure 5 (12). Obese subjects require increased $\dot{V}\hat{O}_2$ to do a given amount of external work. The increase in $\dot{V}\hat{O}_2$ during exercise caused by obesity is considerably greater than the increase that would be predicted from the increase in resting $\dot{V}\hat{O}_2$ because the latter does not include the additional cost of moving the limbs with cycle ergometry and the entire body mass with treadmill exercise. For cycle ergometry, $\dot{V}\hat{O}_2$ has been found to be increased at all work levels by an average of 5.8 ml/min of oxygen for each kg increase over ideal body weight. The effect of obesity on $\dot{V}\hat{O}_2$ during treadmill exercise cannot be described with any precision because of the difficulty in accurately measuring the subject's actual power output for this form of ergometry. It is not possible to get a fully accurate correction for treadmill exercise because of complex mechanical factors, such as the subject's varying center of gravity as the angle of the treadmill is changed and the variable length of the stride as the speed and/or grade are altered.

Cardiovascular diseases

The slope of the $\dot{V}\hat{O}_2$-work rate relationship is also important because it is a measure of the aerobic work efficiency. $\dot{V}\hat{O}_2$ has been found to increase by about 10.1 ml/min/watt (12, 13). However, if the muscle fails to utilize oxygen because of inadequate oxygen delivery caused by stenosis of conducting vessels, heart disease, or other factors that might impair O_2 utilization, then the slope will be more shallow than normal as the maximum $\dot{V}\hat{O}_2$ is approached, and the predicted $\dot{V}\hat{O}_2$max will not be reached. Although, theoretically, a number of reasons might cause this slope to be reduced, a reduction is most likely to be evident in conditions of impaired O_2 flow to the exercising extremities.

In disorders of the cardiovascular system, the $\dot{V}O_2$ may increase normally as work rate is incremented over low levels but slow its rate of increase as the maximum $\dot{V}O_2$ is approached. This decline in slope for the $\dot{V}O_2$-work rate relationship is accompanied by a relatively steep $\dot{V}CO_2$-work rate slope, reflecting the simultaneous buffering of lactic acid generated by anaerobiosis. In this situation, the subject's $\dot{V}O_2max$ is clearly reduced.

Anaerobic threshold

The anaerobic threshold is defined as the level of exercise $\dot{V}O_2$ above which aerobic energy production is supplemented by anaerobic mechanisms (figure 3). Like the $\dot{V}O_2max$, the AT is influenced by the size of the muscle groups involved in the activity.

Information derived from the anaerobic threshold measurement

Blood lactate—The AT is the $\dot{V}O_2$ above which there is a sustained increase in blood lactate and lactate/pyruvate ratio.

$\dot{V}O_2$ above which metabolic acidosis occurs—There is a close reciprocal change in lactate and bicarbonate (figure 9) and decrease in pH during the performance of work above the AT.

$\dot{V}O_2$ above which there is a delay in steady state—The AT demarcates the work rate above which $\dot{V}O_2$ kinetics are changed. The steady state in $\dot{V}O_2$, found to occur by 2–3 minutes below the AT, is delayed above the AT. The increase in $\dot{V}O_2$ between the third and sixth minute of constant work rate exercise was found to significantly correlate with the lactate increase (4).

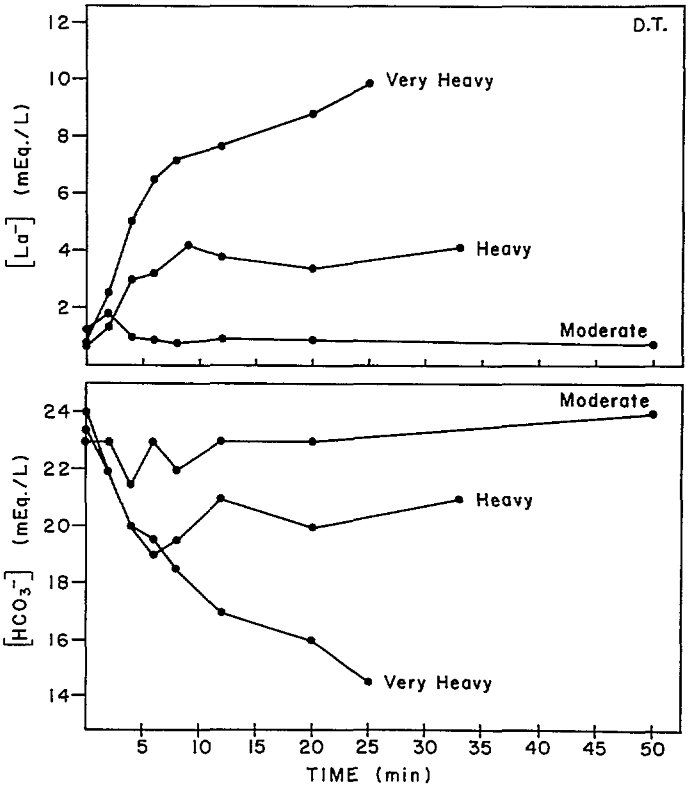


Figure 9. Lactate increase and bicarbonate decrease with time for moderate, heavy, and very heavy work intensities for a normal subject. Bicarbonate changes in opposite direction to lactate and in a quantitatively similar manner. Although the target work rate was 50 minutes for each work rate, the endurance time was reduced for the heavy and very heavy work rates. From Wasserman et al. (24).

Work rate above which $\dot{V}_E$ increases with time—Ventilatory drive is stimulated by the metabolic acidosis resulting from increased lactate production. Thus, $\dot{V}_E$ increases progressively as work rates above the AT are sustained (figure 10). This progressive increase in $\dot{V}_E$ is usually effected by an increase in breathing frequency with little further increase in tidal volume.

Sustainable work rate—Endurance depends on work intensity. Work rates below the anaerobic threshold can be sustained for prolonged (marathon) periods. The greater the increase in arterial lactate, the less the endurance. In endurance cycling studies on motivated young men, it was found that none could sustain pedaling for 50 minutes with an increase in lactate above 2.5 mM/L (figure 11).

Methods of measurement

Because the increase in muscle lactate and H+ is associated with an obligatory increase in CO_2 production, it is possible to detect the cellular acidosis by measuring gas exchange at the mouth. A relatively short, progressive work rate test can rapidly determine the $\dot{V}O_2$ at the anaerobic threshold when gas exchange is measured breath by breath or as the average of several breaths. The appropriate measurements in such a test allow detection of the characteristic gas exchange phenomena associated with developing metabolic acidosis.

A flow diagram describing the gas exchange sequence and ventilation changes for a 1-minute incremental exercise test is illustrated in figure 12. The record of an actual study is shown in figure 13. The linear pattern of increase in $\dot{V}O_2$ and $\dot{V}CO_2$ seen initially, as work rate is incremented, changes above the anaerobic threshold. Lactic acid increase results in an acceleration of the increase in $\dot{V}CO_2$ relative to the $\dot{V}O_2$. Because $\dot{V}_E$ and $\dot{V}CO_2$ initially accelerate in a parallel manner above the AT, there is a short period in which $\dot{V}_E/\dot{V}CO_2$ and $P_{ET}CO_2$ do not change while $\dot{V}_E/\dot{V}O_2$ and $P_{ET}O_2$ increase. Thus, there is hyperventilation with respect to O_2 but not CO_2, initially, as the AT is exceeded. This normally lasts about 2 minutes and is referred to as the period of isocapnic buffering because of the lack of hyperventilation with respect to CO_2 despite the development of metabolic acidosis (14). The abrupt increase in $\dot{V}_E/\dot{V}O_2$ without an increase in $\dot{V}_E/\dot{V}CO_2$ indicates buffering of acid and is not consistent with other causes of increased ventilation (i.e., hypoxemia, pain, psychogenic), which will cause $\dot{V}_E/\dot{V}CO_2$ to increase as well. Thus, the isocapnic buffering period is a sensitive gas-exchange demonstration that the AT has been exceeded. As the work rate is increased further, the carotid bodies respond to the decreasing pH and cause ventilation to increase faster than CO_2 production. This causes $PaCO_2$ to decrease and the reduction in pH to be constrained.

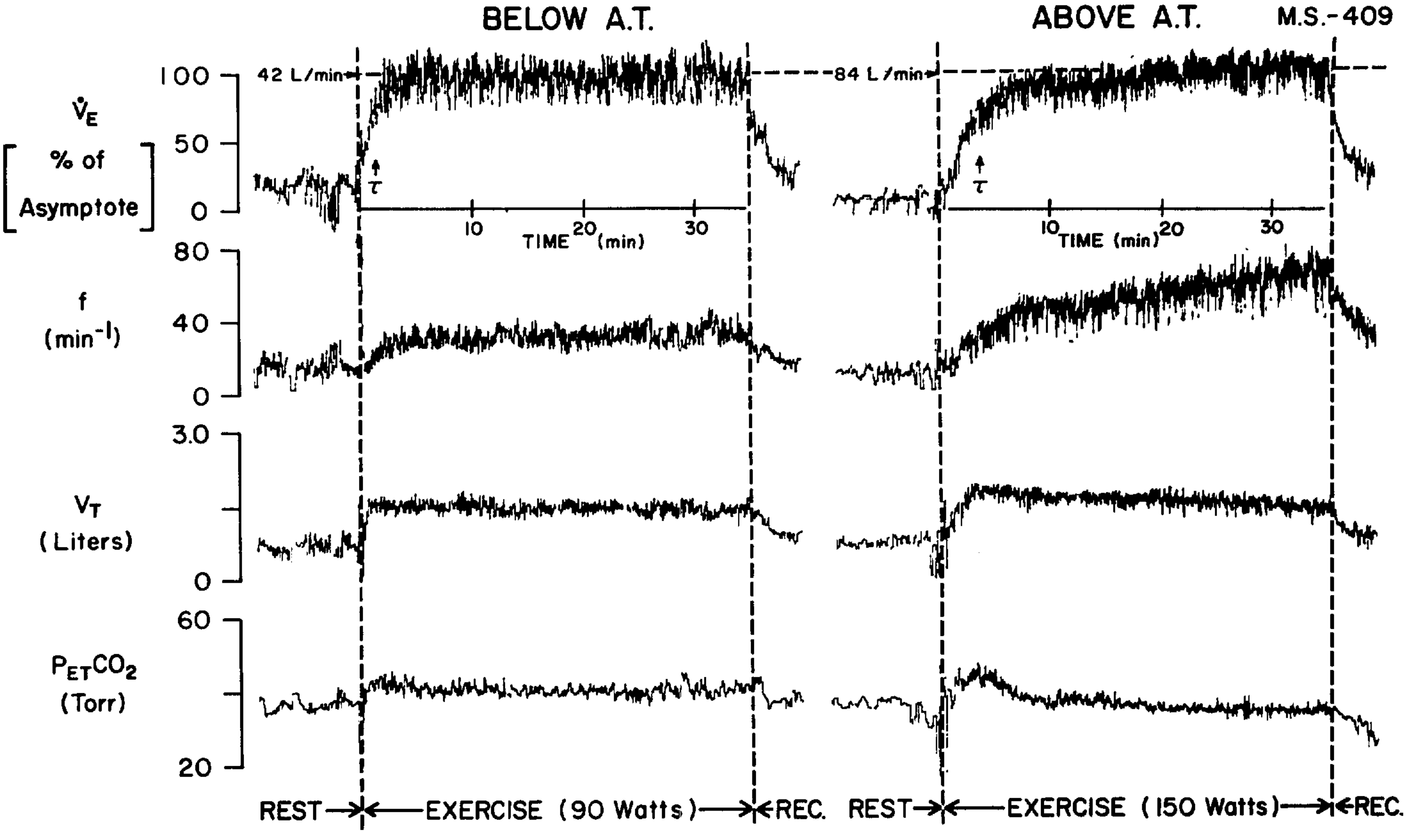


Figure 10. Effect of prolonged exercise below and above the anaerobic threshold (AT) on $\dot{V}_E$, V_T, f, and $P_{ET}CO_2$. The pattern of breathing remains unaltered during prolonged exercise below the AT. In contrast, for work above the AT, the $\dot{V}_E$ and *f* increase with time and $P_{ET}CO_2$ decreases, presumably in response to the accompanying metabolic acidosis. From Wasserman (14).

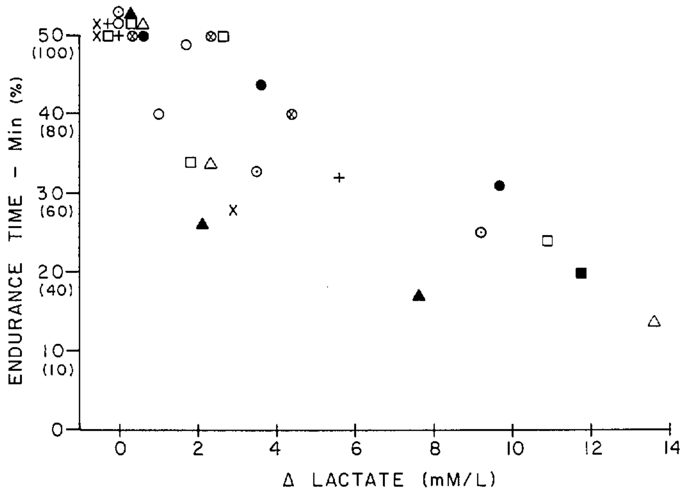


Figure 11. The endurance time as related to the increase lactate above the control value at the end of constant work rate cycle ergometer exercise. Data are from 30 experiments on 10 male subjects studied at 3 work rates, each for a target time of 50 minutes. Endurance is reduced when lactate is increased. From Wasserman (25).

This respiratory compensation for the lactic acidosis is reflected by an increase in $\dot{V}_E/\dot{V}CO_2$ as well as by a further increase in $\dot{V}_E/\dot{V}O_2$. (See figure 13.) R, which ordinarily increases as work rate is incremented, increases more rapidly above the AT. When the AT is measured as a metabolic stress (i.e., in units of O_2 consumption), it is unaffected by the rate with which the work rate is incremented (15, 16).

The reason for preferring a test in which the work rate is rapidly incremented is to take advantage of the fact that the CO_2 contribution from buffering is observed only *during* the buffering process and not after the lactate has been buffered, i.e., during the period of *decreasing* bicarbonate. (An unchanging, although elevated, lactate does not generate additional CO_2 because it is already buffered.)

Application of the anaerobic threshold measurement

The AT aids in the differential diagnosis of disorders of cardiorespiratory coupling to cellular respiration. Conditions that limit O_2 flow to the exercising muscles during exercise are likely to cause the AT to be low. In most subjects, it can be measured noninvasively using the gas exchange measurements described earlier.

Heart rate-$\dot{V}O_2$ relationship and heart rate reserve

Heart rate normally increases linearly with $\dot{V}O_2$ during exercise. In the cardiovascular disorders, the heart rate increase is relatively steep for the increase in $\dot{V}O_2$ because the stroke volume is low (12). As subjects with heart disease approach their maximum work rate, $\dot{V}O_2$ commonly slows its rate of increase while the heart rate continues to increase. Therefore, the rate of heart rate increase relative to $\dot{V}O_2$ becomes even more steep and may deviate from linearity. Although this phenom-

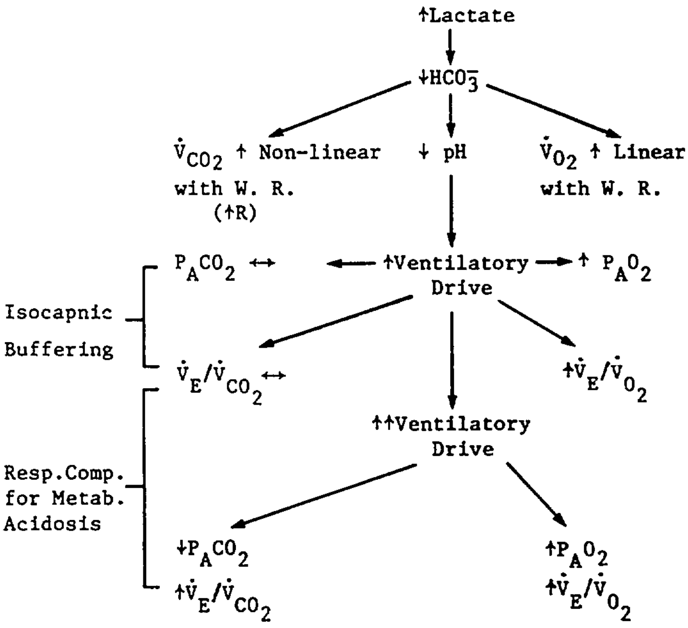


Figure 12. Effect of lactate increase on gas exchange during a 1-minute incremental test. Lactate increase causes an approximately equimolar decrease in biocarbonate. This causes $\dot{V}CO_2$ and R to increase more steeply as work rate in incremented while $\dot{V}O_2$ continues to increase in an essentially linear pattern. At the start of metabolic acidosis, ventilation ($\dot{V}_E$) increases are approximately parallel to the increase in $\dot{V}CO_2$, resulting in no change in end tidal or arterial PCO_2 (isocapnic buffering); alveolar oxygen tension (P_AO_2) is increased secondary to the disproportionate increase in ventilation relative to the increase in $\dot{V}O_2$. The increase in ventilation is reflected in an increase in the ventilatory equivalent for oxygen ($\dot{V}_E/\dot{V}O_2$); the ventilatory equivalent for CO_2 ($\dot{V}_E/\dot{V}CO_2$) does not change. After several minutes, ventilatory drive increases further and $\dot{V}_E$ increases disproportionately to the increase in $\dot{V}CO_2$, resulting in hyperventilation with respect to both O_2 and CO_2. The latter achieves some degree of respiratory compensation for the metabolic acidosis. From Wasserman, Whipp, and Davis (1).

enon is not always seen in subjects with a cardiovascular disease, it is a diagnostically useful observation.

The heart rate reserve expresses the fraction of the predicted heart rate increase that remains at the end of maximum exercise:

$$\frac{\text{HR predicted max} - \text{max exercise HR}}{\text{HR predicted max}}$$

We find that the formula 220 − age (years) predicts normal maximum heart rate fairly well. In normal subjects, the heart rate reserve is relatively small (less than 0.1). It is usually also small in subjects with myocardial dysfunction and in subjects with disorders of the pulmonary circulation. However, subjects with peripheral vascular disease or coronary artery disease may discontinue exercise because of pain before the maximal heart rate is reached. Persons who take beta-adrenergic blocking drugs will have a low maximum heart rate and therefore a high heart rate reserve. Also, subjects limited in exercise because of primary lung disease usually have a large heart rate reserve. Of course, subjects who make poor effort also will manifest an increased heart rate reserve.

O_2-pulse ($\dot{V}O_2$/HR)

The oxygen-pulse is calculated by dividing the oxygen consumption by heart rate. It is the volume of O_2 extracted by the tissues per heart beat and can be shown to be equal to the product of stroke volume and the arterial-mixed venous O_2 difference. (See the earlier section on physiological requirements to perform exercise.) As work rate is increased, the O_2-pulse rises curvilinearly (figure 14), primarily because the arterial-mixed venous O_2 difference increases. When the stroke volume is reduced, the arterial-mixed venous oxygen difference and therefore the O_2-pulse reach maximal values at a relatively low work rate, and the O_2-pulse asymptotes at a low value (figure 15).

Breathing reserve

The breathing reserve is measured either as the difference between the maximal voluntary ventilation and the maximum exercise ventilation in absolute terms or as a fraction of the maximal voluntary ventilation (i.e., analogous to heart rate reserve). Except in extremely fit individuals who can attain high levels of $\dot{V}_E$, normal males have a breathing reserve of at least 15 liters per minute or 20–40% of the maximal voluntary venti-

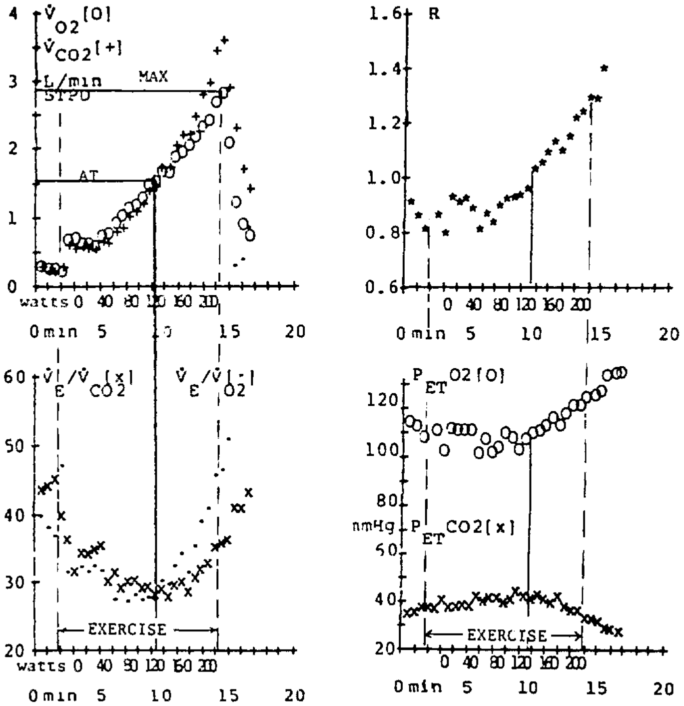


Figure 13. Gas exchange for a normal subject during a 1-minute incremental exercise test. The left vertical dashed line indicates the start of the incremental period of exercise, and the right vertical dashed line indicates the end of exercise. The maximum $\dot{V}O_2$ is indicated in the left upper quadrant. The anaerobic threshold (AT) is the point of the nadir of the $\dot{V}_E/\dot{V}O_2$ curve (vertical solid line). The nadir of the $\dot{V}_E/\dot{V}CO_2$ curve does not occur until a higher work rate is reached and reflects the start of respiratory compensation for the metabolic acidosis. At the AT, P_{ET} O_2 increases, reflecting the hyperventilation with respect to $\dot{V}O_2$, but there is no change in $P_{ET}CO_2$ until approximately 2 minutes later. At the AT, R increases more rapidly, reflecting the increase in $\dot{V}CO_2$ relative to $\dot{V}O_2$. From Wasserman (25).

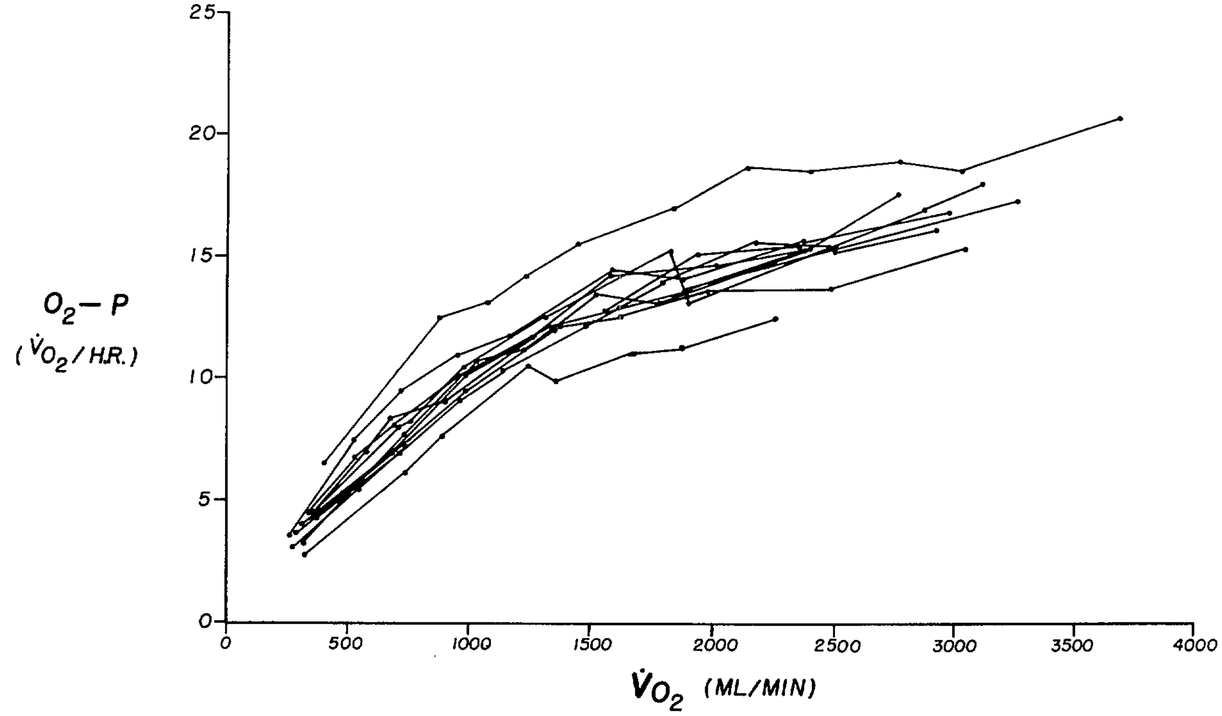


Figure 14. Oxygen pulse for 10 subjects during graded exercise (4 minutes at each load). Each point is the value between the 3rd and 4th minute at the indicated work rate. The highest $\dot{V}O_2$ achieved by a given subject at maximal work capacity. From Wasserman, Van Kessel, and Burton (26).

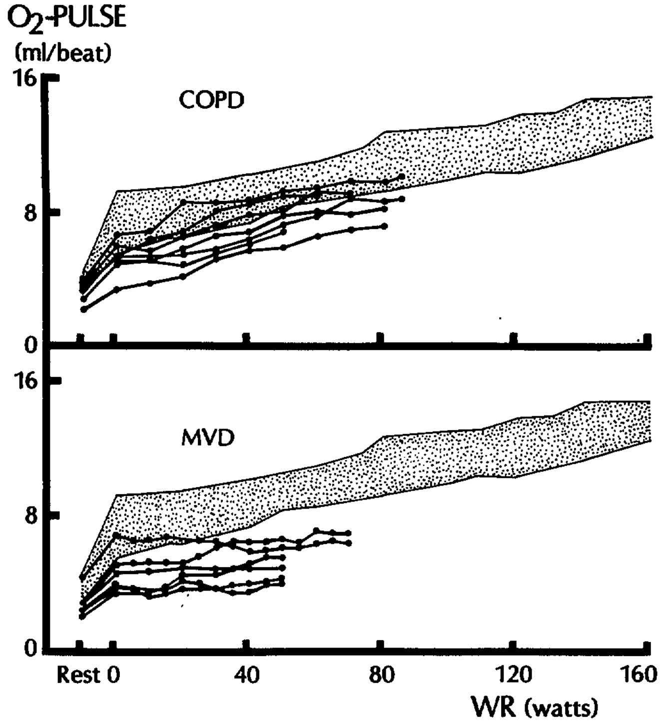


Figure 15. O_2-pulse response to incremental exercise in chronic obstructive pulmonary disease (COPD) (upper panel) and MVD patients (lower panel), compared with the range of values of a control group (stippled area). From Nery et al. (27).

lation (17). A low breathing reserve is characteristic of persons with primary lung disease, as their ventilation is limited. A low breathing reserve is not characteristic of persons with cardiovascular diseases or other disorders that limit exercise performance.

Ventilatory equivalents as indexes of increased V_D/V_T

It is helpful to know if V_D/V_T is abnormal in subjects because abnormality suggests lung or pulmonary vascular disease. The nadir of the ventilatory equivalent for CO_2 or O_2 ($\dot{V}_E/\dot{V}CO_2$ and $\dot{V}_E/\dot{V}O_2$, respectively) can be used as a noninvasive guide to ventilation/perfusion unevenness (increased V_D/V_T) in population screening. Ordinarily, $\dot{V}_E/\dot{V}CO_2$ changes as illustrated in figure 13. The $\dot{V}_E/\dot{V}CO_2$ normally decreases to a nadir of approximately 28, with a range of 26–30. Normal values for ventilatory equivalent with a $P_{ET}CO_2$ of approximately 40 suggest normal V_D/V_T.

Tidal volume/inspiratory capacity ratio ($V\hat{T}$/IC)

This ratio is usually abnormal in subjects with pulmonary fibrosis or chest wall disease. Normally, $V\hat{T}$ increases during exercise, but it rarely exceeds 70% of the IC (determined from resting pulmonary tests) at the subject's maximal work rate. However, subjects with restrictive lung disease have a reduced IC and a limited ability to increase their $V\hat{T}$ in response to exercise. As work rate is incremented, the $V\hat{T}$/IC characteristically exceeds 70% in the restricted subject and commonly asymptotes at a relatively low work rate at the IC volume. Because the increase in $V\hat{T}$ is limited, breathing rate is disproportionately high. Thus a plot of $V\hat{T}$ against $\dot{V}_E$ provides diagnostically useful information, particularly when the IC and MVV limits are plotted on the $V\hat{T}$ and $\dot{V}_E$ axes, respectively.

Anaerobic threshold from constant work rate tests

Measurement of gas exchange, particularly $\dot{V}\hat{O}_2$, is valuable in determining if a subject can sustain a work task. When subjects exercise below their anaerobic threshold, $\dot{V}\hat{O}_2$ reaches a steady state within 3 minutes (figure 4). This means that cellular oxidative mechanisms are being satisfied by atmospheric O_2. In a steady state, there are no anaerobic mechanisms supporting the energetics and the O_2 debt has reached a maximum (18). However, exercise performed above the anaerobic threshold results in a delay in $\dot{V}\hat{O}_2$ steady state. $\dot{V}\hat{O}_2$ continues to increase, the rate of rise depending on the level above the AT. Metabolic acidosis develops under these conditions and breathing rate continues to rise. To determine if the work rate is above the AT, measurement of $\dot{V}\hat{O}_2$ at 3 and 6 minutes during a 6-minute constant work rate test is very helpful. If the 6-minute $\dot{V}\hat{O}_2$ is greater than the 3-minute $\dot{V}\hat{O}_2$ ($\Delta\dot{V}\hat{O}_2$ (6 − 3)), then the work is being performed above the AT. A high correlation between the $\Delta\dot{V}\hat{O}_2$ (6 − 3) and the lactate concentration has been demonstrated (4).

Arterial blood pressure

This may be quite difficult to record during exercise because of the noise of the ergometer (particularly with treadmill exercise) and arm movement (treadmill exercise). However, documenting unusual elevations or decreases in blood pressure when work rate is increased can aid in the diagnosis of cardiovascular dysfunction. Abnormally large fluctuations in blood pressure with respiration are characteristic of pericardial disease and obstructive and restrictive lung disease.

Testing protocols

Form of ergometry

To properly stress the cardiovascular and respiratory systems and to determine limiting factors, exercise testing must use large muscle groups. The most practical forms of ergometry in a laboratory situation are the cycle ergometer and treadmill. The cycle ergometer has advantages over the treadmill in that the work rate performed can be precisely defined. Therefore, the oxygen uptake-work rate relationship, an essential measurement in assessing exercise limitation, can be determined.

Practical considerations also exist that make the cycle somewhat more desirable for the NHANES project. The treadmill is heavier and more expensive than the cycle. Also, it is unlikely that the ceiling clearance of the NHANES project trailer would allow a treadmill to be used for an exercise test of a relatively tall person. (The height of the trailer room would have to exceed 8 feet.) Although it has been argued that the treadmill allows a more normal form of exercise, walking at a pace set by a running belt can be difficult for some subjects unless they hold on to a rail for balance. If this is done, the actual work performed for a given treadmill height or speed is reduced. On the other hand, if elderly subjects or subjects with poor balance do not hold on to the rail, they may fall. Some investigators use a harness suspended from the ceiling to catch the patient in case they lose balance. This is probably impractical in the NHANES trailer. Thus, with respect to subject safety, the cycle ergometer has significant advantages over the treadmill.

Exercise protocol

The most sensitive way to determine abnormalities in the cardiovascular and respiratory systems during exercise is to make measurements breath by breath during an incremental exercise test. This allows the examiner to determine if there is an appropriate cardiovascular and respiratory response to the changing metabolic demand. Mixing-chamber measurements are less reliable than measuring gas exchange at the mouth because the concentrations in the mixing chamber are dissociated in time from the expired air flow and because the concentration measured is derived from variable numbers of breaths and fractions of breaths.

It is strongly recommended that short incremental exercise tests be used with equal work rate increments increased at 1-minute intervals to the subject's maximum during a period of 6–12 minutes. To achieve maximal work rates in this period of time when starting from unloaded cycling, a reasonable practice is to increment the work rate at 10 watts per minute for women, 15 watts per minute for men above age 40, and 25 watts per minute for men less than age 40. This is equivalent to increasing the $\dot{V}\hat{O}_2$ by 10 times the watts each minute. This should be preceded by a short (3-minute) warmup at unloaded cycling or very low level treadmill exercise, as described by Hansen (19). This short protocol allows the complete assessment of each component of the metabolic-cardiovascular-ventilatory coupling. Maximum $\dot{V}\hat{O}_2$, anaerobic threshold, work efficiency, O_2-pulse, heart rate reserve, breathing reserve, and the ventilatory equivalents can be determined from the data. Yet the test is sufficiently short that the subject recovers quickly after the test is completed.

An incremental exercise protocol allows the examiner to simultaneously titrate the performance of the heart, circulation, and respiratory system to the subject's maximal effort and to determine if there is a problem in oxygen flow to the tissues or a breathing limitation. This test is sufficient to alert the examiner to abnormalities. When appropriate measurements are made (such as those described in the sections on measures of exercise capacity and causes of exercise limitations to be detected during testing), the examiner can often distinguish among the many possible causes of exercise limitation.

If the examiner is mainly concerned with the ability of a subject to perform at a given work level requiring a certain metabolic cost, this can be assessed by making measurements at the work rate requiring the metabolic cost of interest. The key measurement in determining if the subject can sustain a given work rate is the ability to maintain constant $\dot{V}O_2$. The $\dot{V}O_2$ reaches a constant level by 3 minutes of exercise if the work rate is performed at or below the anaerobic threshold. Any drift upward in $\dot{V}O_2$ after 3 minutes of sustained work indicates that the work rate is above the subject's anaerobic threshold and is accompanied by metabolic acidosis. The subject should be employed at a job requiring a metabolic rate that can be performed without engendering a metabolic acidosis, as reflected by no change in $\dot{V}O_2$ after 3 minutes of continuous work.

Measurement system

The devices used for making measurements during exercise testing should be determined by the measurements to be made. Therefore, the recommendations for measurements in this paper are not predicated on the availability of a given commercial system. Rather, they are predicated on the idea that equipment manufacturers wanting to sell equipment to NHANES will provide a product that can meet the NHANES standards and requirements. Vendors selling a product to the NHANES project should be obligated to prove that their system can make the recommended measurements. Several manufacturers now market systems that purport to make all the measurements outlined previously.

It should be noted that only bloodless measurements have been described because it is assumed that sampling blood during exercise will considerably complicate the testing for NHANES. To improve accuracy and speed of reporting, the system chosen should use a computer to access and process the physiological signals. Many minicomputers are quite adequate for this purpose. Some manufacturers report that their device can measure physiological dead space and/or cardiac output without arterial blood sampling. I feel that these are exaggerated claims, and selection of a measurement device should not be predicated on these purported abilities. The breathing valve should be kept as small as possible (100 ml) to keep from increasing the anatomical dead space ventilation unreasonably. However, exceptionally high exercise performances may require larger dead space valves to accommodate high airflow without impeding ventilation.

The system used should measure O_2 and CO_2 in the breath with a gas analyzer with a 95% response time of less than 0.1 second and a sampling delay time of less than about 0.3 second. The sampling rate from the mouthpiece should be as small as possible and should not exceed 250 ml/min. The system should be sufficiently flexible to determine delay and response time as part of the daily calibration procedures because this measurement is needed for breath-by-breath gas exchange calculations. Expired and/or inspired ventilation should be determined with a device that measures instantaneous volume or flow, e.g., a pneumotachograph or turbine. The computer should time-align the gas concentrations with exhaled volume so that the sum of the instantaneous products of the derivatives of each can be obtained. Thus, the $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}_E$, V_T, f, HR, and $\dot{V}O_2$/HR can be obtained for each breath. The graphical output can be breath by breath, averages of several whole breaths, or a moving average to reduce noise.

The following computer-generated graphs provide valuable insight into the coupling of external to internal regulation:

- $\dot{V}O_2$ and $\dot{V}CO_2$ versus work rate.
- HR and O_2 pulse versus work rate.
- HR versus $\dot{V}O_2$.
- $\dot{V}_E/\dot{V}O_2$ and $\dot{V}_E/\dot{V}CO_2$ versus work rate.
- $P_{ET}O_2$ and $P_{ET}CO_2$ versus work rate.
- R versus work rate.
- $\dot{V}_E$ versus work rate.
- $\dot{V}_E$ versus $\dot{V}CO_2$.
- V_T versus $\dot{V}_E$.

The ergometer must be accurately and linearly calibrated and should have an unloaded cycling work-rate equivalent of less than 15 watts.

Causes of exercise limitation to be detected during exercise testing

Many disorders interfere with the normal metabolic-cardiovascular-ventilatory coupling needed to perform exercise, including primary disorders of red blood cell production, the peripheral circulation, the heart, the pulmonary circulation, the lungs, the chest wall, respiratory control, and metabolism. Limiting symptoms are dyspnea, fatigue, and/or pain. The pathophysiological mechanisms contributing to the exercise intolerance of each of these disorders are briefly described here and summarized in table 2.

Heart disease

The major and most immediate role of the heart is to pump blood in order to provide oxygen to the tissues to support their metabolic requirements. During exercise, the heart is under particular stress. All primary disorders of the heart cause limitation in the amount of blood that can be pumped to the tissues. Therefore, exercise testing provides a specific approach to detecting diseases of the heart.

Coronary artery disease

Coronary artery disease can best be detected under conditions that cause an increase in cardiac work. When the blood flow perfusing the myocardium is inadequate to provide the oxygen required to support the bioenergetics needed for generation of ATP for cardiac contraction, repolarization of the myocardium is slow. This usually results in abnormal electrocardiographic complexes during exercise performance, such as ST-segment elevations and T-wave changes. Also, ventricular arrhythmias developing during exercise should be regarded as evidence of myocardial ischemia. Myocardial

Table 2. Disorders and mechanisms impairing work tolerance

Disorder	Mechanism
Obesity	Increased metabolic requirement; cardiorespiratory restriction
Peripheral vascular disease	Prevents normal vasodilation, thereby limiting muscle O_2 supply
Heart diseases	Limits cardiac output (stroke volume) increase and thereby O_2 delivery to tissues
Anemia (or carboxyhemoglobinemia)	Limits muscle oxygenation; stimulates breathing at low WR because metabolic acidosis develops, leading to dyspnea
Pulmonary vascular occlusion	Prevents normal increase in $\dot{Q}$; ventilation of nonperfused lung (increased V_D/V_T); causes hypoxemia with exercise
Lung diseases	Reduced maximum ability to breathe; decreased efficiency of gas exchange (increased V_D/V_T); exercise-induced hypoxemia
Chest wall disease	Reduced maximum ability to breathe

ischemia will result in a reduced stroke volume and limited increase in cardiac output. Thus, maximum aerobic capacity (maximum $\dot{V}O_2$ relative to the predicted maximum $\dot{V}O_2$) and the anaerobic threshold will be reduced. Also, $\dot{V}O_2$ may fail to increase normally as work rate is incremented.

Valvular heart disease

The heart may not be able to eject a normal stroke volume because of dysfunction of one of the heart valves, either stenosis or incompetence. This limits the increase in cardiac output that can be generated during exercise, and O_2 transport to the tissues becomes inadequate. Thus the maximum $\dot{V}O_2$ and $\underline{\text{AT}}$ will be reduced, the heart rate response with respect to $\dot{V}O_2$ will be high and the O_2 pulse will be low at all work rates, and a flat O_2-pulse response will commonly be elicited.

Cardiomyopathy

A disorder of the myocardium caused by metabolic dysfunction, hypertrophy, dilatation, or fibrosis replacing muscle will cause stroke volume to be reduced and limit cardiac output increase. The physiological measurements during exercise testing that will be abnormal are similar to those described for valvular heart disease; however, the heart rate reserve may be abnormally high in subjects in whom abnormalities in the conduction system prevent the normal increase in heart rate.

Congenital heart disease

Most varieties of congenital heart disease will result in limited exercise tolerance because of failure to develop a normal cardiac output. In the case of subjects with cyanotic congenital heart disease, the problem is compounded because of the increased O_2 desaturation that occurs during exercise, as a larger fraction of the cardiac output bypasses the lungs. The physiological measurements during exercise testing will be similar to those for valvular heart disease. In addition, strikingly increased ventilatory requirements will result when the pulmonary circulation is involved and when a right-to-left shunt is present or develops during exercise.

Arrhythmias

Heart block, failure to increase heart rate appropriately consequent to sinoauricular node dysfunction, or the development of inappropriate tachycardia in which cardiac ejection is inefficient can limit exercise tolerance. These become evident when monitoring the heart rate-$\dot{V}O_2$ relationship.

Lung diseases

Most persons with lung disease are limited by shortness of breath. The persons with obstructive lung diseases (bronchial asthma, bronchitis, and emphysema) have the predominant problem of being unable to ventilate enough to eliminate CO_2 produced from metabolism. The pathophysiology of exertional dyspnea in subjects with chronic obstructive airway disease is shown in figure 16. Two basic problems exist in these subjects, both limiting the maximum metabolic rate that

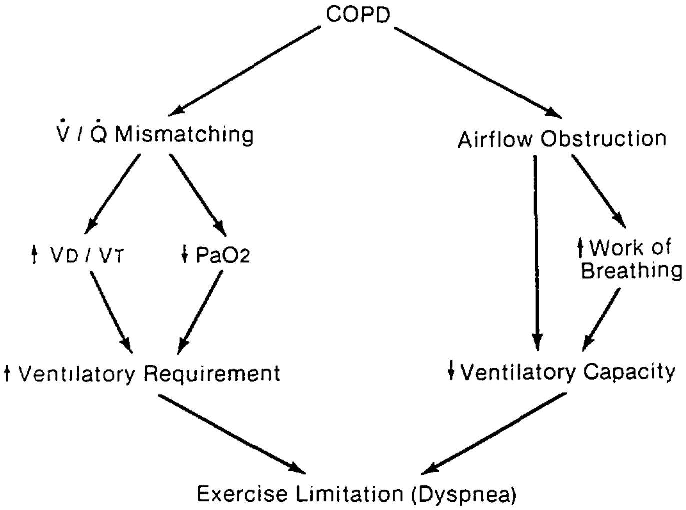


Figure 16. Pathophysiology of exercise limitation in patients with chronic obstructive pulmonary disease (COPD). V_D/V_T indicates physiological dead space/tidal volume ratio; $\dot{V}/\dot{Q}$, ratio of ventilation to perfusion in lung gas exchange units. From Brown and Wasserman (28).

they can endure: (a) Because of limited ability to breathe, the maximum ventilation that the subject can develop is reduced; (b) the ventilation required by the subject to perform exercise is increased because ventilation becomes inefficient as a result of mismatching of ventilation relative to perfusion in persons with obstructive lung disease. The maximum $\dot{V}O_2$ is reduced in persons with obstructive lung diseases, but the $\dot{V}O_2$-work rate relationship does not approach a plateau, as it commonly does in patients with circulatory limitation. Generally, the AT, if it can be determined, is on the low side of normal, consistent with a typically sedentary subject. The heart rate at maximum work rate is generally low (high heart rate reserve) but can be increased if the maximum work rate can be improved by appropriate therapy. Because of the mismatching of ventilation relative to perfusion, persons with primary lung disease generally have high ventilatory equivalents for $\dot{V}O_2$ or $\dot{V}CO_2$ at all work rate levels. Also, subjects with obstructive lung disease often do not develop an appreciable respiratory compensation for metabolic acidosis. Consequently, the ventilatory equivalents do not increase, even at the subject's maximum work rate.

The ventilatory response of subjects with pulmonary fibrosis is steep because of hyperventilation (reduced $PaCO_2$) and high V_D/V_T. Ventilation at their maximum work rate commonly approaches the MVV.

Obesity

Despite some increase in the resting metabolic rate ($\dot{V}O_2$) relative to the lean body mass in obese subjects, the increase in $\dot{V}O_2$ is even more marked during dynamic exercise because of the additional energy needed to move their large body segments. As adipose tissue is added to the body, there is no commensurate growth of the heart or blood vessels. Consequently, when the obese individual does any form of physical work, there is a greater than normal cardiovascular and respiratory stress.

Because of fat deposition in the chest wall and abdomen, a constraint is also imposed on the maximal performance of the cardiovascular and respiratory systems by obesity. Because of the large body mass, resting cardiac output per kg lean body weight is already high. During exercise, the further increase in cardiac output is limited (20).

Hypoxemia is commonly present at rest, consequent to atelectasis of peripheral lung units resulting from the pressure imposed on the chest by the obesity. During exercise, presumably because of the deep breathing, the hypoxemia disappears due to re-aeration of atelectatic alveoli.

The increased O_2 cost of performing mechanical work is predictable and well worked out for cycle ergometer work (12, 13). The $\dot{V}O_2$-work rate relationship is displaced upward in obese subjects without a discernible change in slope, the extent depending on the degree of obesity (figure 5).

The $\dot{V}O_2$max and AT are low when related to actual body weight, but are normal when related to height (13) and lean body mass (21). Ventilation-perfusion relationships are usually normal during exercise, resulting in normal ventilatory equivalents for O_2 and CO_2. The O_2-pulse is normal or high when related to predicted values determined from normal body weight. The heart rate reserve and breathing reserve are generally normal.

Peripheral vascular disease

The normal increase in peripheral blood flow essential for increasing O_2 flow to the working muscles is not possible in subjects with peripheral vascular diseases. Thus, the O_2 supply to the exercising muscles may not adequately meet the high O_2 requirement. Consequently, the ischemic muscles produce lactic acid. The failure of the circulation to efficiently remove the acid may induce pain. When the lactic acid reaches the central circulation, breathing is further stimulated. Although there is a compensatory increase in mitochondrial number in the ischemic muscle, it is inadequate to make up for the deficiency in O_2 flow (22).

The $\dot{V}O_2$max and the anaerobic threshold are reduced, although the latter can be so low that it often cannot be detected. Because lactate may not enter the central circulation in detectable quantity because of reduced muscle perfusion, evidence of a systemic lactic acidosis may be minimal. The slope of the $\dot{V}O_2$/work rate relationship may be more shallow than normal, reflecting diminished O_2 utilization and increased anaerobic energy production.

It is important to make blood pressure measurements. Arterial hypertension commonly accompanies peripheral vascular occlusive disorders.

Although resting and exercise blood pressures are elevated in subjects with essential hypertension, the maximum $\dot{V}O_2$, AT, and work efficiency are normal. The heart rate reserve and the O_2-pulse will be high in those subjects receiving beta-adrenergic blocking drug therapy.

Pulmonary vascular disease

Diseases of the pulmonary circulation characteristically result in reduced perfusion of ventilated alveoli. Because these alveoli are so poorly perfused, a large ventilatory component is wasted (alveolar dead space). Likewise, alveoli that are perfused must be ventilated to a greater degree than normal to remove the metabolic CO_2 and maintain $PaCO_2$ at a normal level. Because minute ventilation is the sum of the alveolar ventilation and the physiological dead-space ventilation, it is increased in subjects with pulmonary vascular diseases.- Thus the ventilatory equivalents of O_2 and CO_2 will be abnormally high.

Hemodynamically, pulmonary vascular occlusive diseases represent a stenosis in the central circulation, making it difficult for the right ventricle to deliver blood to the left atrium at a rate sufficient to meet the increased cardiac output needed for exercise. The cardiac output increase in subjects with pulmonary vascular disease can be quite limited, and reductions in $\dot{V}\hat{O}_2$max, $\underline{AT}$, and O_2-pulse similar to those found in persons with primary heart disease are usually observed.

Anemia and abnormal hemoglobins

Anemic states (blood with reduced O_2 content) compromise delivery of O_2 to the mitochondria because the blood $P\hat{O}_2$ falls more rapidly than normal as the red cells travel through the capillaries. Because of the reduced capillary blood $P\hat{O}_2$, the diffusion gradient for O_2 from the blood to the mitochondria might reach critically low levels, causing anaerobic energy generation at relatively low work rates.

Similarly, the presence of abnormal hemoglobins that have a leftward shift in the O_2 dissociation curve (a reduced P_{50}) result in a decrease in capillary blood $P\hat{O}_2$ at a given O_2 consumption. Because, at high rates of O_2 consumption, the partial pressure difference for oxygen between the capillary and mitochondrion may not provide adequate O_2 flow, the mitochondrial membrane shuttle would not get reoxidized rapidly enough to maintain the cytosol redox state. Thus lactic acidosis occurs at reduced work rates (23).

Breathlessness during exercise in subjects with anemia is well recognized. The increased ventilatory drive (mediated by the carotid bodies) is presumably caused by the metabolic acidosis that accompanies the low anaerobic threshold.

Subjects with reduced O_2 carrying capacity have a relatively high cardiac output, with a higher than expected heart rate for a given work rate, i.e., a relative tachycardia. In contrast to subjects with cardiac diseases and disorders of the pulmonary circulation, in which the stroke volume is reduced, the stroke volume is normal or even increased in this disorder. However, as both the arterial and venous O_2 contents are low to begin with, the potential increase in the arterial-venous O_2 difference is limited. Thus, the O_2-pulse increases to a maximum at a relatively low work rate in the anemic subject and plateaus at a low value.

Summary

Assessment of the cardiorespiratory health and work potential of a population requires the ability to determine the work capacity of the subject in a short period of testing without discomfort. The predicted maximum $\dot{V}\hat{O}_2$ should be reached with appropriate cardiovascular and ventilatory responses. It is not adequate to simply determine that the predicted $\dot{V}\hat{O}_2$max was reached, because subjects with considerable disease can reach the normal value. To assess the health of the organ systems involved in exercise performance, it is essential to determine whether the predicted maximum $\dot{V}\hat{O}_2$ was reached with *normal* cardiovascular and ventilatory responses.

References

1. Wasserman, K., B.J. Whipp, and J.A. Davis. Respiration in exercise. In: *Internat. Rev. Physiol., Respiration Physiology*, Vol. 3. Edited by J.G. Widdicombe. Baltimore: Univ. Park Press, pp. 149–211, 1981.
2. McGilvery, R.W. *Biochemistry: A Functional Approach.* 3rd Ed., 1983; p. 469–470. W.B. Saunders Co., Philadelphia.
3. Perret, C. Hyperoxie et regulation de la ventilation durant l'exercise musculaire. *Helvetica Physiologica Pharmacologica Acta.*, 18:72–97, 1960.
4. Roston, W.L., B.J. Whipp, J.A. Davis, D.A. Cunningham, R.M. Effros, and K. Wasserman. Oxygen uptake kinetics and lactate concentration during exercise in humans. *Am. Rev. Resp. Dis.* 135:1080–1084, 1987.
5. Whipp, B.J., and K. Wasserman. Oxygen uptake kinetics for various intensities of constant-load work. *J. Appl. Physiol.* 33:351–356, 1972.
6. Astrand, P.O., and K. Rodahl. *Textbook of Work Physiology.* 2nd Edition, pp. 393–411, McGraw-Hill Book Co., New York, 1977.
7. Whipp, B.J., and K. Wasserman. Efficiency of muscular work. *J. Appl. Physiol.* 26:644–648, 1969.
8. Whipp, B.J. The control of the exercise hyperpnea. In: *The Regulation of Breathing.* Edited by T. Hornbein. New York: Dekker, pp. 1069–1139, 1981.
9. Wallace, A.G. Electrical activity of the heart. In: *The Heart.* Edited by J.W. Hurst. New York, McGraw-Hill, 1982.
10. Taylor, H.L., E. Buskirk, and A. Henschel. Maximal oxygen intake as an objective measure of cardiorespiratory performance. *J. Appl. Physiol.*, 8:73–80, 1955.
11. DiPrampero, P.E. Energetics of muscular exercise. *Rev. Physiol. Biochem. Pharamacol.*, 88:143–222, 1981.
12. Wasserman, K., and B.J. Whipp. Exercise Physiology in Health and Disease. *Am. Rev. Resp. Dis.*, 112:219–249, 1975.
13. Hansen, J.E., D.Y. Sue, and K. Wasserman. Predicted values for clinical exercise testing. *Am. Rev. Resp. Dis.*, 129:S–49–S55, 1984.
14. Wasserman, K. Breathing during exercise. *N. Engl. J. Med.*, 298–780–785, 1978.
15. Whipp, B.J., and M. Mahler. Dynamics of gas exchange during exercise. In: *Pulmonary Gas Exchange,* Vol. II. Edited by J.B. West. New York, Academic Press, 1980.
16. Buchfuhrer, M.J., J.E. Hansen, T.E. Robinson, D.Y. Sue, K. Wasserman, and B.J. Whipp. Optimizing the exercise protocol for cardiopulmonary assessment. *J. Appl. Physiol.: Respi. Environ. Exer. Physiol.*, 55(5):1558–1564, 1983.
17. Sue, D.Y., and J.E. Hansen. Normal values in adults during exercise testing. In: *Clinics in Chest Medicine.* Symposium on Exercise: Physiology and Clinical Applications. Edited by J. Loke. 5:89–97, 1984.
18. Schneider, E.G., S. Robinson, and J.L. Newton. Oxygen debt in aerobic work. *J. Appl. Physiol.*, 25:58–62, 1968.
19. Hansen, J.E. Exercise instruments, schemes, and protocols for evaluating the dyspneic patient. *Am. Rev. Resp. Dis.* (Suppl.), 129:S25–S27, 1984.
20. Alexander, J.K., K.H. Amad, and V.W. Cole. Observations in some clinical features of extreme obesity, with particular reference to circulatory effect. *Am. J. Med.*, 32:512–524, 1962.
21. Buskirk, E., and H.L. Taylor. Maximal oxygen intake and its relation to body composition, with special reference to chronic physical activity and obesity. *J. Appl. Physiol.*, 1172–78, 1957.
22. Bylund-Fellenius, A.C., P.M. Walker, A. Elander, and T. Schersten. Peripheral vascular disease. *Am. Rev. Resp. Dis.* (Suppl), 129:S65–S67, 1984.
23. Butler, W.M., L.A. Spratling, J.A. Kark, and E.B. Schoomaker. Hemoglobin Osler: Report of a new family with exercise studies before and after phlebotomy. *Am. J. Hemat.*, 13:293–301, 1982.
24. Wasserman, K., J.E. Hansen, D.Y. Sue, and B.J. Whipp. *Principles of Exercise Testing and Interpretation.* Lea and Febiger Publishers. Philadelphia, 1987.

25. Wasserman, K. The anaerobic threshold measurement to evaluate exercise performance. *Am. Rev. Resp. Dis.,* 129:Suppl:S35–S40, 1984.
26. Wasserman, K., A.L. Van Kessel, and G.G. Burton. Interaction of physiological mechanisms during exercise. *J. Appl. Physiol.* 22:71–85, 1967.
27. Nery, L.E., K. Wasserman, W. French, A. Oren, and J.A. Davis. Contrasting cardiovascular and respiratory responses to exercise in mitral valve and chronic obstructive pulmonary diseases. *Chest* 83:446–453, 1983.
28. Brown, H.V., and K. Wasserman. Exercise performance in chronic obstructive pulmonary diseases. *Med. Clinics of North America* 65:525–547, 1981.

Part III

Fundamental Perspectives on Energy Balance, Dietary Patterns, and Physical Activity

Chapter **6**

General Considerations Related to Assessing Energy Turnover: Energy Intake or Energy Expenditure

E. R. Buskirk, Ph.D.
The Pennsylvania State University

Introduction

If the assessment of energy turnover is important in a national survey—and it appears that it is for purposes of ascertaining what people are eating in relation to what they do—then it is important to obtain relatively accurate indications of either energy intake or energy expenditure. These variables are related to food intake and physical activity, respectively; the latter energy output is superimposed on the resting metabolic rate. Such data, complemented by data on body composition and nutritional status, provide reasonable estimates of the energetics of large groups and populations. The energetics of different population segments, regional trends, etc. can also be assessed. At the outset, it must be realized that methods applicable to surveys are generally less accurate than those used in the laboratory or the metabolic ward.

The energy required for growth, maturation, daily organ function, and maintenance is ultimately derived from the chemical transformations of the absorbed nutrients—a process that follows the first law of thermodynamics, i.e., the law of conservation of energy, which states that energy can neither be created nor destroyed. The second law of thermodynamics stipulates that all forms of energy are quantitatively convertible to heat. Thus, any input of food energy augments body energy content and that energy is lost from the body in the form of heat, although the performance of work modifies how energy is transferred. It should be remembered that energy represents an abstraction because energy can be measured only when it is converted to heat or the transformation process can be related to some intermediate step such as oxidation. For comparative purposes, the units for energy transfer and heat exchange are the kilocalorie (kcal) and the kilojoule (kJ).

Energy balance

The basic equation for energy balance appears in figure 1. Because energy intake must equal energy output (commonly referred to as expenditure) plus losses and storage, either side of the equation can be evaluated to obtain an estimate of energy turnover. An interesting perspective on use of the equation has been provided by Beaton (1), who questions its interpretation without input from both the biological and psychosocial sciences. General methods for measuring the intake and output (expenditure) sides of the equation are listed in table 1. The listing for general methods of assessment of energy intake is expanded in table 2. Similarly, an expanded listing of methods for measurement of energy output, or expenditure, appears in table 3. These relatively inclusive listings are offered as a survey of the options available for assessment of daily energy turn-

Energy balance

Intake ± E^1 = expenditure ± E

Food intake = work + heat + waste ± tissue storage
Lipid
Carbohydrate
Protein
Other substrates
Heat

[1] Energy can be gained from the environment, e.g., solar radiation.

Figure 1. Basic equation for energy balance.

over. When it comes to choosing possibilities for population surveys, the choices are more limited and the investigator is forced to evaluate the utility of the various options, as has been attempted in table 4.

In regard to energy losses from the body, they are several and in some instances subtle. In conventional studies of energy balance, all losses of energy are rarely measured. It is relatively common to measure losses in urine and feces, but the following losses are usually disregarded: Perspiration, tears, sputum, flatus, epidermal sloughing, nail clippings, male and female discharges, and blood loss. Because these losses are often neglected, daily energy intake usually exceeds energy output or expenditure when both are carefully controlled under laboratory conditions (2, 3). Because of greater measurement error, the calculation of energy balance under field conditions may not reflect such a small but consistent difference.

Table 1. General methods of measurement of energy intake and expenditure

- Energy intake:
  - Dietary survey
  - Survey with weighed portions
  - Chemical analysis of representative food samples
  - Liquid formulas and powdered meals
  - Bomb calorimetry of replicate meals or aliquots
- Energy expenditure:
  - Activity diary
  - Activity survey
  - Indirect heat loss
  - Direct heat loss
  - Indirect heat production
  - Activity-related, i.e., heart rate, pulmonary ventilation, etc.
  - Body-fat and lean-tissue loss

Source: Adapted from E. R. Buskirk and J. Mendez, 1980. Copyright; used by permission.

Table 2. Methods of assessment of energy intake

- Dietary survey:
  - Self-administered recall questionnaires
  - Self-prepared food-intake record
  - Recall questionnaire with interview
  - Observation during meal eating
  - Group logistics and ration analysis
- Survey with weighed portions [1]:
  - Individual data
  - Group data
    - Procurement
    - Distribution
    - Waste
- Chemical analysis of weighed samples:
  - Individual meals
  - Daily composites
  - Multiday composites
  - Analysis of food purchases and food inventory
- Liquid formulas and prepared meals:
  - Diets from recipes
  - Commercial preparations
    - Liquid
    - Powder
    - Meals such as TV dinners
- Bomb calorimetry:
  - Sample foods
  - Composite meals
  - Multiday composites

[1] Recipe method of food inventory: records of food preparation and consumption.

Source: Adapted from E. R. Buskirk and J. Mendez, 1980. Copyright; used by permission.

There are a variety of studies of energy balance, i.e., of energy intake and output, in the literature. Two of the most comprehensive are those conducted by Edholm et al. (4, 5). In general, better results (i.e., closer agreement) were obtained for women than for men. The differences between energy intake and output ranged from plus 735 kcal d^{-1} (energy intake exceeded output) to minus 170 kcal d^{-1} (energy output exceeded intake). Plotting of individual data showed considerable interindividual variability, with a seemingly random distribution. Comparison of daily kcal intake with output in the same individual over several days also revealed relatively large differences. It is of note that energy output tends to decrease with less than maintenance caloric intake (3). Data reviewed by Apfelbaum and Bostsarron (6) clearly indicate such a decrease. Failure to recognize this physiological and behavioral adaptation could lead to misinterpretation of energy balance data, leading to an overestimate of negative balance. Such a consideration is also important because nutritional status may well modify energy output, i.e., induced by another deficit such as iron deficiency.

Several gross factors influence energy turnover, and some of them are listed in table 5. Most of these factors have been taken into consideration when the committee from the Food and Nutrition Board has prepared the Recommended Dietary Allowances (7). Thus, it is important that these factors be taken into account when surveying energy turnover in any population.

Durnin and Ferro-Luzzi (8) have prepared recommendations for conducting and reporting studies on energy intake and output. Many of their recommenda-

Table 3. Methods for measurement of energy expenditure and heat exchange

- Physical activity diary:
  - Recall at intervals
  - Personal time-motion study
  - 15-minute checksheet
- Physical activity survey:
  - Self-administered recall questionnaire
  - Questionnaire with interview
  - Direct-observation time-motion study
  - Photographic time-motion study
- Physical job classification:
  - Pedometers
  - Cumulative heartbeats
  - Cumulative joint rotation
  - Cumulative extremity acceleration
  - Ventilation ($\dot{V}_E$)
- Indirect heat loss:
  - Body temperature changes
    - Radiometers
    - Thermocouples
    - Thermistors
  - Humidity detectors
- Direct heat loss:
  - Calorimeters—partial or total body
    - Radiation
    - Convection and conduction
    - Evaporation
- Indirect heat production:
  - Oxygen consumption ($\dot{V}O_2$) and/or carbon dioxide production ($\dot{V}CO_2$)
- $\dot{V}$ and FO_2:
  - $\dot{V}O_2$, $\dot{V}CO_2$, and nitrogen excretion
  - ^{18}O turnover ($^2H_2{}^{18}O$ or $^3H_2{}^{18}O$)
  - Insensible water loss
- Body-fat loss (change in):
  - Relative body weight
  - Anthropometric measurements
  - Somatotype
  - Subcutaneous fatness
    - Skinfolds
    - Soft-tissue x ray
    - Ultrasonic pattern
    - Tissue impedance
  - Fat-soluble gas dilution
  - Densitometry, air and water
  - Total-body water and other fluids
  - CAT scan
- Lean-tissue loss (change in):
  - Representative biopsies
  - Nitrogen excretion
  - Creatinine excretion
  - ^{40}K and exchangeable ^{42}K
  - Potassium balance
  - 3-Methylhistidine excretion

Source: Adapted from E. R. Buskirk and J. Mendez, 1980. Copyright used by permission.

Table 4. Options for daily measurement of energy expenditure using methods for activity assessment ranking

Activity	I	II	Energy expenditure	I	II
Free diary	5	2	Dietary assessment	5	4
Questionnaires	4	1	Heart-rate accumulation	4	2
Prepared diary	3	3	Joint rotation	3	3
Time-motion, observer	2	5	$\dot{V}$ monitor	2	1
Time-motion, movie	1	4	Continuous $\dot{V}O_2$, $\dot{V}CO_2$, RQ	1	5

Note: I, accuracy of measurement: Superior, 1 to inferior, 5; II, practicality of measurement: relatively easy, 1 to difficult, 5.
Source: Adapted from E. R. Buskirk and J. Mendez, 1980. Copyright; used by permission.

Table 5. Factors modifying energy turnover and energy balance

Physical activity
Body size
Growth and development
Aging
Sex
Climate or environment
Disability or disease
Bed rest
Pregnancy and lactation

Note: Not all factors listed are independent.
Source: Adapted from E. R. Buskirk and J. Mendez, 1980.
Copyright; used by permission.

tions are included in the tables that follow. For the most part, the various considerations need to be addressed during the planning phases of any study or survey. Important information to be considered for inclusion in energy studies appears in table 6. In terms of energy intake appraisal, a number of crucial questions need to be answered during the planning phase of a survey. Some of these questions appear in table 7. On the other side of the energy balance equation, special considerations and crucial questions with respect to energy output or expenditure appear in table 8.

As one looks at the available literature on energy turnover, there appear to be a variety of shortcomings. Some of these shortcomings are

Poor methodology utilized
Methodology inadequately described
Subjects, group, or population inadequately described
Careless pooling of data
No explanation of aberrant findings
Lifestyles inadequately described

Many of the concerns expressed here have been addressed in the other, more specific manuscripts that are part of this publication. Suffice it to say that appropriate planning is all important in designing and programming an effective population survey designed to assess energy turnover by either energy intake or expenditure.

Table 6. Important general information in energy turnover studies

- Purpose of the survey

- Why a particular group is selected
  - Method of subject selection
  - Why number is chosen
  - Whether the sample is biased; if so, in what way

- Why longitudinal or cross-sectional

- Socioeconomic information:
  - Occupation
  - Household grouping
  - Living conditions
  - Food availability and nutrition
  - Physical activity

- Anthropometry and physical characteristics

- Time of year, climate, and environment

- Duration of study

Table 8. Considerations for an energy expenditure appraisal

- Need to know for a diary-type record:
  - Physical activity components
  - Critical periods of time
  - Who completed the diary
  - Whether there were observers and who they were
  - Who checked the diary and how frequently

- Need to know for indirect calorimetry:
  - Equipment used
  - How equipment was calibrated
  - Duration of measurements
  - How activities were selected
  - How activity was standardized
  - How representative the activity was
  - Number of measurements
  - How measurements were made
  - What equations were used

- Need to know for use of energy expenditure tables:
  - Source of energy expenditure tables
  - Whether adjustments were made for age, sex, body size, etc.
  - How representative the tables are regarding subject activities

Summary

A brief review has been presented of a variety of considerations important in planning and conducting studies of energy turnover. Ultimately, their relative worth will be shown in the solution of practical problems related to appropriate nutrition and the prevention and outpatient management of disease. Our understanding of procedure applicability in an effort to improve survey results should ultimately prove of value to the populace at large. At present our methodologies and the rigor with which they can be applied need improvement. In general, coefficients of variation of from 10% to 20% can be expected, but we hope that our sustained efforts will place our future best procedures in the more acceptable range of 5% to 10%. The latter range for coefficients of variation would appear adequate for energy turnover surveys of large groups of people.

Table 7. Considerations for an energy intake appraisal

- Need precise descriptions of:
  - How food intake was measured
  - Whether standardized procedures were used
  - What exactly was measured
  - Who made the measurements
  - Where it was done
  - How long it took
  - Whether there was replication
  - Whether procedures were validated and how validation was conducted

- Need to know:
  - What kind of scales or balances were employed
  - How accurately the balances were read
  - How the balances were calibrated
    - Whether food item mixing was done
    - Whether composites were separated
  - How many and what kind of containers were used
  - How waste was handled
  - What sort of records were kept and whether the subject was involved in recordkeeping

- Instructions given to subjects:
  - How and where instructions were given
  - What supervision was involved and for how long
  - Period of subject involvement
  - Whether the period was representative
  - Which tables or computer programs were used and how
  - Whether allowance was made for digestion, absorption, and excretion
  - Whether they reflect food or available energy

References

1. Beaton, G. H. Energy in human nutrition: perspectives and problems. *Nutr. Rev.* 41:325–340, 1983.
2. Buskirk, E. R., J. L. Hodgson, and D. Blair. Assessment of daily energy balance: Some observations on the methodology for indirect determinations of energy intake and expenditure. Chap. in J. M. Kinney (Ed.). *Assessment of Energy Metabolism in Health and Disease.* Columbus, OH: Ross Laboratories, 1980, pp. 113–117.
3. Buskirk, E. R., R. H. Thompson, R. Moore, and G. D. Whedon. Human energy expenditure studies in the National Institute of Arthritis and Metabolic Diseases metabolic chamber. *Am. J. Clin. Nutr.* 8:602–613, 1960.
4. Edholm, O. G., J. M. Adam, M. J. R. Healy, H. S. Wolff, R. Goldsmith, and T. W. Best. Food intake and energy expenditure of army recruits. *Brit. J. Nutr.* 24:1091–1107, 1970.
5. Edholm, O. G., S. Humphrey, J. A. Loure, B. E. Tredre, and J. Brotherhood. VI. Energy expenditure and climate exposure of Yemenite and Kurdish Jews in Israel. *Philos. Trans. R. Soc. London* (B) 266:127–140, 1973.
6. Apfelbaum, M., and J. Bostsarron. The energy balance of obese patients on a reducing diet. *La Presse Med.* 77:1941, 1969.
7. Food and Nutrition Board. *Recommended Dietary Allowances.* 9th Edition. National Academy of Sciences, National Research Council, Washington, DC, 1980.
8. Durnin, J. V. G. A. and A. Ferro-Luzzi. Conducting and reporting studies on human energy intake and output: Suggested standards. *Am. J. Clin. Nutr.* 35:624–626, 1982.
9. Buskirk, E. R. and J. Mendez. Energy: Caloric requirements. Chap. 2 in R. F. Alfin-Slater and D. Kritchevsky (Eds.). *Human Nutrition: A Comprehensive Treatise, 3A, Nutrition and the Adult Macronutrients.* New York: Plenum Press, 1980, pp. 49–95.

Chapter 7

Issues Related to Measuring Energy Balance for the National Health and Nutrition Examination Survey

Dorothy Blair, Ph.D.
The Pennsylvania State University

Introduction

The question most relevant to a discussion of dietary methodologies for measuring energy balance in the National Health and Nutrition Examination Survey (NHANES) is not which method is most appropriate, but rather, whether an attempt to measure energy balance is an appropriate aim for a large survey such as NHANES. To answer this question, we must first have a basic understanding of what energy balance is, what affects it, and who may be defined as being in energy balance. Then we should ask with which methods and under what conditions energy balance has been successfully measured and whether those methods and conditions can be duplicated or approximated in NHANES.

This paper attempts to answer these questions and ends with a discussion of the dietary methodologies that are most appropriate for the estimation of total energy intake for NHANES.

Definition of energy balance

Energy balance can be defined as a close agreement between caloric intake and caloric expenditure. Garrow (1) defines energy balance as an agreement between energy intake and energy expenditure within ± 50 kcal/day. Although this is somewhat arbitrary, it demarcates the limits of our accuracy of measurement under very controlled circumstances. An imbalance of lesser magnitude would nevertheless have physiological consequences. A positive imbalance of 30 kcal/day could lead to an accumulation of 1 kg of body fat per year.

The test of energy balance in an individual is the maintenance of stable body fat stores. Fluctuations in carbohydrate stores may cause variations in body weight as great as 2 kg (1), with a 2,000 kcal fluctuation in body energy stores. Such changes in carbohydrate stores invalidate body weight as a measure of energy balance (2). The choice of stable body fat as the test for energy balance puts limits on who can be considered to be in energy balance. Growth and reproductive-related changes in body fat stores rule out the possibility of being in energy balance for large segments of the population. The use of percent body fat is less valid in the context of energy balance than is total body fat in kilograms, as the percent body fat may fluctuate without

---

Note: Thanks are due to Dr. Helen S. Wright for our conversations about dietary methodology and to Dr. John Beard for his comments on the manuscript. This paper was originally prepared for the NHANES III Dietary Survey Methodology Workshop held at Airlie House in Warrenton, Virginia, March 16–18, 1986.

Table 1. Components and effectors of energy balance

Intake effectors	Caloric intake components	Body energy stores Body energy losses	Caloric expenditure components	Expenditure effectors
Conscious restraint or augmentation of intake	Fat	Feces Urine	Basal needs	Fat free mass
Palatability	Carbohydrate		Adaptive thermogenesis: Food induced Nonspecific	Long-term kcal
Mood, stress, illness, drugs	Protein		Immune response, fever	Short- and long-term intake; meal composition
Time and financial constraints	Alcohol		Growth, reproductive function; body building and training	Smoking, tension, drugs, thermal regulation
Weekend, weekday, season, holiday			Movement and exercise	Cyclic variations
Somatic controls: Energy expenditure and exercise Physiological feedback Cyclical variations (menstrual cycle)				Fat free mass; weight, day, season
Alcohol				

a necessary change in total fat stores. An example is the aging-related loss of lean tissue, which leads to an increased percent of body tissue as fat without an increase in body energy stores.

Beaton makes a point worth noting: Energy balance is maintained in some cases at the expense of optimal physiological and mental functioning (3). An example is the adaptation of children to severely reduced energy intake through the cessation of growth and reduced activity. More commonly observed in affluent societies is the adaptation to low caloric intakes with severe and long-term dieting. Thus, energy balance is not of itself a desirable state but may be an adaptive response of the body to adverse circumstances.

The components and effectors of energy intake and energy expenditure

Table 1 shows the components of energy intake and energy expenditure and lists those factors that influence either side of the equation. Energy expenditure is under much tighter bounds than energy intake. It is a function of the Fat Free Mass, FFM (4, 5), of short- and long-term caloric intake (6, 7) and food composition, of illness and growth, and of movement (2). Movement would seem to be a discretionary category, but for most of us the caloric demands of the day are somewhat fixed. Leisure-time activity is the one dimension of energy expenditure we are free to define. Energy expenditure may vary from slightly greater than 1 × BMR, basal metabolic rate, (assuming complete bed rest without food intake) to greater than 2 × BMR. Most of us live in the range of 1.3 to 1.5 × BMR (sedentary to light activity). Not even a 6-mile jog would hoist us into the heavy activity category. Mean caloric intake estimates of population groups that are under 1.4 × BMR must be viewed with suspicion.

Effectors of energy intake are much more varied and may be situational as well as under cognitive, emotional, or somatic controls. Physiological feedback mechanisms described by Van Itallie (8, 9) and Booth (10) may be overridden by external stimuli: Palatability, other peoples' behavior, etc. Conscious controls may be overridden by emotionality, alcohol, or drugs. To most, eating is a social, emotional, and sensual activity that is synchronized with physiological needs only by dint of constant learned attention to physiological cues (10). For those who are less finely in tune with their body's needs, the salient cue may be the body composition change that is the negative outcome of ignoring earlier physiological cues.

The propensity toward obesity may be a function of the degree to which cognitive and somatic controls can be overpowered by external cues and emotional needs (11). The obese also appear to differ from normal persons in their exercise-induced eating. Increased exercise does not appear to influence the caloric intake of obese women, but normal persons increase intake in relation to exercise intensity (12). Intake may therefore be less tightly coupled to energy expenditure in obese than in normal persons.

Most single effectors of both intake and energy expenditure have a coupled influence so that the chance for energy balance is heightened. For example, severe decrease in intake decreases both basal metabolism (6, 7) and food-induced thermogenesis, which is approximately 10% of caloric intake (1), depending on meal composition. The energy expended in exercise drops as weight is lost (6). Smoking increases metabolism and increases consumption (13, 14), although the increase in intake may result from a relaxation of necessary restraint. The menstrual cycle causes a coupled variability in energy expenditure and intake (15). Exercise also increases intake over the long term in normal persons (12), although the relationship is not closely coupled within a 2-day period (16).

On the other hand, some effectors uncouple energy intake and expenditure. Illness increases metabolic requirements but reduces caloric intake through reduction in appetite (17). In some, stress has an appetite-depressing effect, but muscular tension increases energy expenditure (18). A yearly cycle of weight loss and gain may occur because of seasonal and holiday overeating when exercise is reserved for the warmer months. Restrained eating and the conscious augmentation of energy expenditure could cause a particularly sharp energy imbalance.

Thus, though there are some physiological and cognitive reasons for a close correlation between energy intake and expenditure, there are other forces at work that disrupt this coupling and make a close match between energy intake and expenditure unlikely. It also seems to be true that the time required for energy balance to occur may differ between obese and lean individuals because of a difference in intake cue responsiveness. In a sedentary society where food is plentiful, varied, and easily attained, energy balance on the short, measurable term may be the exception rather than the rule.

Precise estimations of energy balance and the categorization of individuals: Where do they get us?

The point of measuring energy balance is not only to determine with some acceptable degree of reliability and validity those individuals who are in energy balance and those who are not. The broader goal would seem to be to increase our understanding of the mechanisms of body fat maintenance and the effects of body fat fluctuations on disease processes and outcomes. It must be stressed that energy balance, unless broadly defined and appropriately measured, is not the same as body fat maintenance.

It can be assumed that fat-stable individuals, whether normal or obese, fluctuate with some regularity around a set body composition. We do not know the periodicity of this fluctuation or the height of the arc, i.e., the degree to which body composition change is tolerated by the individual before corrective mechanisms are employed. There are probably three basic types of body composition maintenance cycles, with many variations on these themes: (1) A close coupling of intake and output, resulting in slight weight fluctuations from day to day but no substantive change in body fat stores, i.e., true energy balance; (2) a short-term but measurable fluctuation of body fatness around a long-term stable body composition; and (3) a seasonal cycle of imbalance, with an increase in winter holiday eating and cold-related decrease in exercise, followed by a reverse summer trend. These three styles are shown in figure 1. By definition, those individuals whose body composition fluctuates (types 2 and 3) cannot be considered to be in energy balance, nor will they be measured as such by precise estimates of intake, output, or body composition assessments. Thus, those individuals whose body composition is, by any other criteria, stable will necessarily be considered out of energy balance if their body composition fluctuates regularly. The time period chosen for measurement will also have a profound effect on who is considered to be in energy balance. Those who balance over a longer period and who are caught on the upswing or downswing will be considered out of balance.

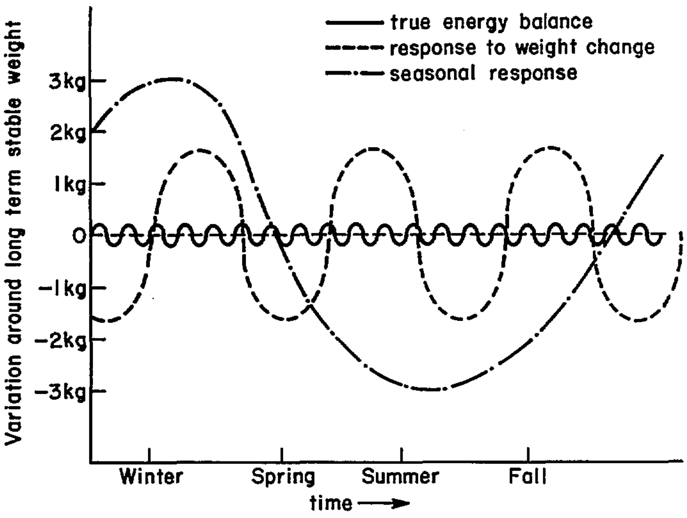


Figure 1. Three styles of long-term body composition maintenance

Thus, many long-term fat-stable individuals will be designated as out of balance by precise, short-term studies with low tolerances for body composition change. The more stringent the criteria for energy balance, the more false negatives will result. Positive and negative energy imbalances will exclude an individual from the maintenance criteria regardless of the outcome of that imbalance. Short-term energy balance studies have neither the sensitivity nor specificity to put individuals in categories that can serve a useful predictive function. The concept of body fat maintenance is more easily operationalized via serial measurements of body composition, which would be a more direct measure of the phenomenon of interest than energy intake and expenditure.

Review of the attempts to measure energy balance: The realities of imprecision

The literature on energy balance can be divided into two categories: Those studies performed on small samples under fairly rigorous field conditions and those studies using methods more applicable to epidemiological studies such as NHANES. Metabolic chamber studies are not included in this review because of their impact on habitual behavior. The more rigorous field studies will be examined first to see if any techniques for measuring energy balance can yield results that are both valid and useful for the purposes of NHANES.

The most exacting field studies have obtained activity records and food records or weighed intakes over 7 days or longer. Energy expenditure has typically been calculated by the factorial method (19). Generally, group means for energy intake have been very close to energy expenditure, but some individual subjects appear to be considerably out of balance over the time frame of a week. Durnin's review of six 7-day studies shows that, of 69 individuals studied, only 6% had daily intakes that correlated positively with daily energy expenditure (17). Twenty-six percent had significant differences in the mean energy intake minus energy expenditure after 7 days. Harries et al. (20), reviewing studies done from 1955 to 1962, felt that 1 week was not a sufficient time to estimate energy balance because of the day-to-day variability in energy intake. The variability of energy expenditure was considerably lower than that of intake.

None of the early estimates of energy balance used a body composition methodology more sophisticated than body weight. In an exceedingly long study of energy balance on a captive group of men in Antarctica, Acheson et al. (21) used a four-site skinfold estimate of body fat, repeated every week over a period of 6–12 months. However, none of their varied techniques for measuring energy balance was sufficiently accurate to predict individual fat gains or losses over the study period.

Both Acheson et al. (21) and Borel et al. (22) have utilized long-term caloric intake data along with long-term data on body composition to estimate the caloric intake required for energy balance. This method, the intake balance technique, corrects caloric intake for the caloric cost of changes in body composition. The technique has the advantage of correcting for body composition changes that may be the result of altered intake caused by the self-monitoring of food consumption. However, the correction factor is only as good as the method for estimating body composition. Certainly, skinfolds are too imprecise to be used in this context, but even densitometry could result in substantial errors in the estimation of the caloric cost of body composition change. The technique does have the advantage of eliminating the matching of two messy methodologies—dietary intake and energy expenditure—as a criterion for success.

Over a week, energy balance may be within ±10–20%. The error of field energy expenditure techniques is at best ±10% (23, 1). Assuming accurate reporting, 3 weeks of dietary data via 24-hour recall or food record are necessary to estimate usual energy intakes with confidence limits of ±10% (24). Poor subject cooperation and recall abilities can further decrease validity, along with errors in food composition estimates and coding. In both cases, recording may change behavior. If randomness of error can be assumed, then use of group data would be possible, though the findings would be weak. If randomness cannot be assumed, then an error of this magnitude could obscure even a group comparison.

Thus, 7 days or more of continuous data collection of a type requiring intense subject cooperation and not a small intrusion on people's daily routine—methods totally unsuitable for NHANES—are not sufficiently valid to classify individuals into energy balance categories that are consistent with changes in their body fat stores.

The second area for examination in this review of the literature on energy balance studies is whether estimations of energy intake and expenditure more applicable to a large-scale epidemiological survey such as NHANES yield both valid and useful results. Because energy balance techniques are tedious for the subject and intrude on normal behavior, many investigators have looked for easier methods to measure one or both sides of the energy balance equation, with a resulting increase in the error of estimation. Usually the energy expenditure side of the equation suffers the most from approximation. The subject may spend tedious hours recording dietary intake over many days, yet energy expenditure is estimated by use of an average daily heart rate and one heart rate-energy expenditure regression equation obtained in a laboratory under controlled conditions (6, 25, 26). More frequently, activity levels have become the surrogate measure for energy expenditure, presumably as the one variable in the energy expenditure equation under individual control. However, no attempt has been made to convert activity into calories expended, which leaves the impression that a given activity has an equal metabolic cost for all persons. The unconverted activity levels are then compared with caloric consumption.

A number of small- and large-scale studies have examined dietary intake and activity compared with body fatness. The dietary methods have ranged from 1-day or repeat 7-day food records and food histories to 24-hour recalls. Activity levels have been estimated through diaries, observation, or a mechanical device such as a pedometer. The results of these studies are as inconsistent as the methods are varied, but results can be grouped into the following three categories.

Body fat is related to decreased caloric intake and decreased activity:

Johnson et al. (1956): Diet histories, teenagers (27).
Rose and Mayer (1968): Diet histories, infants (28).
Hutson et al. (1965): 24-hour recall, adult men (29).
Montoye et al. (1976): 24-hour recall, adult men (30).
Baecke et al. (1983): 2-day food records, adult men (31).

Body fat is related to decreased caloric intake but not to differences in reported activity:

Hutson et al. (1965): 24-hour recall, adult females (29).
Baecke et al. (1983): 2-day food records, adult females (31).
Huenemann (1967): 7-day food records, mixed teenagers (32).

Table 2. Caloric intake and sum of 4 skinfolds by activity group in the Tecumseh, Michigan, males

Activity category	Kcal/kg	*N*	Sum of skinfolds(mm)[1]	*N*
Most active	38.2	192	83.6	275
Intermediate	37.4	516	86.7	813
Least active	33.9	175	92.4	273

[1] Triceps, subscapular, suprailium, and juxtaumbilicus.
Source: Montoye et al. (30).

Stefanik et al. (1959): Diet histories, teenage boys (33).
Bradfield et al. (1971): 3-day records, teenage girls (34).
Body fat is unrelated to caloric intake or activity:
McCarthy (1966): 7-day diet record, adult females (35).
Maxfield and Konishi (1966): 7-day diet record, adult females (36).

The methods used in these studies are so imprecise that biases in measurement between obese and lean subjects could easily be the major source of significant results, other errors being random. There is a tendency, at least in adults, to find no energy intake differences between obese and normal persons using a 7-day diet record, and 24-hour recalls have routinely categorized the obese as eating less. This tendency was not true for obese children and adolescents, who were found to eat less than normal individuals by every dietary method employed.

The temptation is strong to use very short-term dietary data from cross-sectional studies to draw conclusions relevant to energy balance and weight maintenance, or even to categorize obese and lean individuals. We are ever hopeful that some relationship between caloric intake and body fat will be found, even if the meaning of those relationships is equivocal. The problems of drawing conclusions from such data can be illustrated by the analysis of the Tecumseh, Michigan, data by Montoye et al. (30) and the analysis of data from the first National Health and Nutrition Examination Survey (NHANES I) by Braitman et al. (37). Both studies employed a 24-hour recall to approximate energy intake. The Tecumseh study employed a rigorous 30–60 minute recall of the preceding year's activities to classify 1,000 men into active, moderately active, or sedentary categories (table 2). In NHANES I, more modest estimates of perceived recreational and vocational activity on a scale from 0–2 were employed. These two activity scales were summed for an activity index ranging from 0–4.

Montoye et al. found an inverse relationship between a year's average for work and leisure activity, expressed as a work-to-basal ratio and a 1-day estimate of caloric intake. A diet history would have provided a more appropriate comparison to a year's worth of activity data. The authors state that their work-to-basal ratio estimates energy expenditure, but in fact it tends to equalize the relationship between energy expenditure and activity among groups of different body compositions. The ratio penalizes the obese because the greater caloric cost of their exertion is balanced by their higher metabolic rate. It is not surprising that inactivity, so defined, was found to be related to a higher sum of skinfolds. Caloric intake was standardized by kilograms of body weight. A comparison based on caloric intake per FFM would have kept the comparison congruous. An even more congruent comparison would be caloric intake compared with an estimate of average daily energy expenditure based on an estimate of metabolism and the average basal-to-work ratio.

The sample of Montoye et al. ranged from 16–64 years of age. Decreasing activity and decreasing caloric intake are expected trends with aging, as in an increase in adipose tissue, at least until age 40 in men (38). Thus, the strength of these trends may be accounted for by aging itself and not any unique differences between obese and normal individuals. The lack of consistent significance across age groups strengthens this possibility.

The methodological difficulties, the standardization problems, and the comparison of a year's activity to 1 day of caloric intake all serve to make the data of Montoye et al. very difficult to interpret and accept at face value. Even if the comparisons were valid, could we infer from this data that increasing fatness is associated with decreasing caloric needs? No, because activity constitutes only a fraction of energy needs and because we know nothing about the variability in caloric intake in this sample. There is reason to believe that this variability is greater for obese than normal individuals, as was stated in the section on components and effectors of energy intake and expenditures. The study tells us nothing at all about dietary variability and nothing about energy balance.

If more days of dietary intake were assessed or a diet history for the past year were performed, if the activity and metabolic data were converted to kilocalories, would the data then give us information on energy balance? The energy expenditure data might give an estimation of energy needs, which by one estimate is accurate within ±200 kcal (39). This is useful. The estimate of caloric intake would be considerably more valid and useful if variability over time were considered (40). However, the difference between the two estimates would not be of sufficient validity and reliability to determine the direction of imbalance, predict weight maintenance, or predict changes in body weight over time.

The data of Braitman et al. from NHANES I are presented in table 3 (females) and table 4 (males). Reported caloric intake of females by 24-hour recall was inversely related to the percent optimum weight category. This relationship remained significant after adjustment for age and the perceived activity index. The relationship between caloric intake and weight category was not significant for men.

Can it be inferred from these data that obese females require less calories than normal females to maintain

Table 3. Unadjusted and adjusted caloric intake of females from the first National Health and Nutrition Examination Survey at 5 levels of percent of optimum weight: 1971–75

Percent of optimum weight	Number of subjects	Unadjusted caloric intake	Caloric intake adjusted for physical activity and age [1]
<101	1,246	1,743	1,689
101–119	1,321	1,591	1,595
120–134	453	1,480	1,550
135–149	252	1,411	1,488
>149	245	1,478	1,525

[1] For the analysis of covariance, F = 9.269 and $p < 0.001$.
Source: Braitman et al. (37).

their weight? Is caloric intake in males unrelated to weight maintenance? The authors state that "there are two possible interpretations of these data: either obese adults eat no more than nonobese and therefore maintain their greater weight without excessive energy intake, or the estimates of food intake in this study differ in accuracy between obese and nonobese adults, either being underestimated by the former or overestimated by the latter" (37). At least two other hypotheses are possible. One is that the obese are restrained eaters for the most part. Thus, a 24-hour recall may underrepresent the impact of their unrestrained days and underestimate true consumption. A second hypothesis is that the assumption of weight maintenance in each group is false. The tendency in cross-sectional studies is to assume that weight fluctuations in each group are equal, canceling each other, and that the caloric intake reported is for weight or energy store maintenance. There is cross-sectional evidence from the National Health Examination Survey of 1960–62 and NHANES I, 1971–74, that over a 10-year time span, female age cohorts have increased energy stores in the form of body fat in all but the age group 55–74 years. Men increased their subscapular skinfold in each age category, but triceps did not increase after ages 35–44 years (41). Increases in internal body fat stores would be expected with increasing age. Other studies have shown significant changes in body composition over time (2). Obviously, 10-year changes are not in a league with the short-term fluctuations that also could cause erroneous data interpretation, but both types of changes do lend credibility to the fourth hypothesis. We do not know which group, the obese or normal, has a more consistent upward trend in body fat.

The degree of energy imbalance and the distribution of energy imbalance within a range of body fatness would be crucially important for the interpretation of the Braitman data. However, if longer term, valid data were available on energy intake and expenditure for NHANES III, we still could not determine energy balance with the degree of accuracy necessary to estimate the direction of true fat fluctuation. Only serial body composition estimates, the test of energy balance, would provide the information needed to establish both rapidly and unequivocally if a meaningful energy imbalance exists.

Table 4. Unadjusted and adjusted caloric intake of males from the first National Health and Nutrition Examination Survey at 5 levels of percent of optimum weight: 1971–75

Percent of optimum weight	Number of subjects	Unadjusted caloric intake	Caloric intake adjusted for physical activity and age [1]
<101	708	2,423	2,359
101–119	1,241	2,366	2,375
120–134	512	2,264	2,310
135–149	155	2,214	2,278
>149	79	2,406	2,411

[1] For the analysis of covariance, F = 0.795 and $p = 0.53 > 0.05$.
Source: Braitman et al. (37).

Appropriate caloric intake methodology for NHANES III

The precise and valid measurement of caloric consumption is an appropriate goal for NHANES III, but the limitations of available methods must restrain the uses to which the data are put. No method available is ideal, and some are not even possible within the context of NHANES. Short- and long-term diaries and weighed records are not applicable because of their reliance on subject cooperation and recording abilities. They also tend to interfere with normal eating patterns. NHANES III must rely on methods that put the burden of work on the nutritionist rather than the interviewee. Rigorous quality control will have to be maintained in interviewer selection and training and through the collection of replicate data.

There are several criteria which should be considered when choosing the best dietary methodology for estimating energy consumption:

1. The data should reflect the full range of variability in the individual's diet, including weekdays, weekends, and other points of variability.

Dietary histories obtain a subject's estimate of habitual intake (42), although the method has also been used to obtain weekend-weekday estimates, which can then be weighed (43). The 24-hour recall has the advantage of giving data for discrete days so that group variability can be estimated if a suitable number of days are sampled. (24, 40).

NHANES II performed 24-hour recalls on Tuesdays through Saturdays (44), so the weekend variability in intake observed in other studies (45, 46) was not recorded. Information on weekend-weekday patterns must be part of the dietary method. Each subject should have at least one 24-hour recall from a weekend day. If a weekend-weekday eating pattern is shown to exist, the intake estimate should be a weighted average of weekends and weekdays.

2. The method should minimize reporting bias, particularly differences between obese and normal subjects.

Twenty-four-hour recalls may underestimate caloric intake of the obese versus intake of normal persons (47–49), but diet histories are subject to the same dependence on subject recall (50). When compared with a caloric intake that had maintained subject weight within ±2% over at least 90 days, the diet history technique overestimated the normal subjects' needs and underestimated the obese subjects' needs (39). Whatever dietary method is employed, it should use special probing techniques aimed at uncovering the intake of high-calorie food items, such as snack foods, desserts, and alcohol, foods frequently underestimated (49, 50). A food frequency cross-check designed to elucidate the consumption of these particular high-calorie food items may provide a useful probing tool.

3. The situation should maximize the subject's ability to remember both the types of foods consumed and the amounts.

Unless most meals are consumed away from the home, the home environment would provide the best site for the initial dietary interview. Here the setting and the contents of the refrigerator and cupboards provide memory cues. The size of portions can be approximated directly from the actual serving utensils. If interviews were computerized, information on usual container and plate sizes could be retained on the computer for subsequent interviews in the mobile caravans and perhaps by telephone (51, 52).

Because intakes on discrete days can help provide the estimates of day-to-day variability necessary to generate a statistical approximation of the distribution of usual intakes (24), repeat 24-hour recalls should be the method of choice for a cross-sectional comparison of group data. However, 2–3 days of dietary intake data cannot give a precise estimate of individual intake. As a result of the large intraindividual variation in caloric consumption, correlations and regressions of caloric intake with relevant physiological parameters are unlikely to yield more than marginally significant results (53). The probability of type 2 error is high.

In the right hands and with cooperative subjects, the diet history technique would seem to be a better choice than the 24-hour recall for prospective studies. Several authors have shown its repeatability over time, especially for energy intake (54–56). Changes in energy intake as estimated by diet history have been reasonably congruent with changes in body weight (57, 58). A similar correlation using a 24-hour recall technique could not be expected. Diet histories give a higher estimate of caloric intake than 24-hour recalls (59, 60), especially in the obese (48). One negative aspect of diet histories is the possible difficulty of obtaining information from school-age children (61). Diet histories are more time consuming to code and to convert to nutrient intake; therefore, they are more expensive to perform than 24-hour recalls. However, if an important part of an already expensive survey is to have comparable energy consumption data over time, the extra expense of thorough baseline dietary data would be compensated for by usable data.

Conclusion

NHANES can never hope to achieve the level of precision required for the measurement of energy intake and expenditure to estimate energy balance; nor is it clear that energy balance data would provide more useful information than body composition measured over time. The attempt to measure energy balance is an academic exercise more suited to testing the validity of methodologies than meaningfully assessing individuals.

The accurate measurement of total energy intake is an appropriate and useful goal for NHANES. Although no existing method is perfectly suited to the measurement of energy intake, the precision with which this variable is presently measured in NHANES could be considerably improved. Repeat 24-hour recalls with improved cross-checks are recommended for cross-sectional design. A diet history technique is recommended if a longitudinal design is adopted.

References

1. Garrow JS. *Energy balance and obesity in man.* 2nd edition. Amsterdam, The Netherlands. Elsevier/North-Holland Biomedical Press. 1978.
2. Acheson KJ, Campbell IT, Edholm OG, Miller DS, Stock MJ. A longitudinal study of body weight and body fat changes in Antarctica. *Am J Clin Nutr* 1980;33:972–7.
3. Beaton G. Energy in human nutrition: perspectives and problems. *Nutr Rev* 1983;41:325–40.
4. Cunningham JJ. A reanalysis of the factors influencing basal metabolic rate in normal adults. *Am J Clin Nutr* 1980;33:2372–74.
5. Ravussin E, Burnand B, Schutz Y, Jequier E. Twenty-four-hour energy expenditure and resting metabolic rate in obese, moderately obese, and control subjects. *Am J Clin Nutr* 1982;35:566–73.
6. Warnold I, Carlgren G, Krotkiewski M. Energy expenditure and body composition during weight reduction in hyperplastic obese women. *Am J Clin Nutr* 1978;31:750–63.
7. Liebel RL, Hirsch J. Diminished energy requirements in reduced-obese patients. *Metabolism* 1984;33:164–70.
8. Van Itallie TB, Kissileff HR. The physiological control of energy intake: an econometric perspective. *Am J Clin Nutr* 1983;38:978–88.
9. Van Itallie TB, Kissileff HR. Physiology of energy intake: an inventory control model. *Am J Clin Nutr* 1985;42:914–23.
10. Booth DA. Acquired behavior controlling energy intake and output. In: Stunkard AJ, ed. *Obesity.* Philadelphia, PA: WB Saunders, 1980:101–43.
11. Rodin J. The externality theory today. In: Stunkard AJ, ed. *Obesity.* Philadelphia, PA: WB Saunders, 1980:226–39.
12. Pi-Sunyer TX, Woo R. Effect of exercise on food intake in human subjects. *Am J Clin Nutr* 1985;42:983–90.
13. Hofstetter A, Schutz Y, Jequier E, Wahren J. Increased 24-hour energy expenditure in cigarette smokers. *N Engl J Med* 1986;314:79–82.
14. Jacobs DR Jr, Gottenborg S. Smoking and weight: the Minnesota Lipid Research Clinic. *Am J Public Health* 1981;71:391–6.

15. Anonymous. Changes in nitrogen and energy metabolism during the menstrual cycle. *Nutr Rev* 1983;41:116–8.
16. Durnin JVGA. "Appetite" and the relationships between expenditure and intake of calories in man. *J Physiol* 1961;156:294–306.
17. Chandra RK. Nutrition, immunity and infection: present knowledge and future directions. *Lancet* 1986;1(March 26):688–91.
18. Miller DS, Mumford P. Obesity: physical activity and nutrition. *Proc Nutr Soc* 1966;25:100–7.
19. Durnin JVGA, Brockway JM. Determination of the total daily energy expenditure in man by indirect calorimetry: assessment of the accuracy of a modern technique. *Brit J Nutr* 1959;13:41–53.
20. Harries JM, Hobson EA, Hollingsworth DF. Individual variations in energy expenditure and intake. *Proc Nutr Soc* 1962;21:157–68.
21. Acheson KJ, Campbell IT, Edholm OG, Miller DS, Stock MJ. The measurement of daily energy expenditure—an evaluation of some techniques. *Am J Clin Nutr* 1980;33:1155–64.
22. Borel MJ, Riley RE, Snook JT. Estimation of energy expenditure and maintenance energy requirements of college-age men and women. *Am J Clin Nutr* 1984;40:1264–72.
23. Passmore R. Energy balances in man. *Proc Nutr Soc* 1967;26:97–101.
24. National Research Council. Comm. on Life Sciences. Food and Nutrition Board. Coord. Comm. on Eval. of Food Cons. Surveys. Subcomm. on Criteria for Diet Eval. Nutrient adequacy. Assessment using food consumption surveys. Washington DC. National Academy Press. 1986.
25. Bradfield RB, Jorden M. Energy expenditure of obese women during weight loss. *Am J Clin Nutr* 1972;25:971–5.
26. Griffiths M, Payne PR. Energy expenditure in small children of obese and non-obese parents. *Nature* 1976;260:698–700.
27. Johnson ML, Burke S, Mayer J. Relative importance of inactivity and overeating in the energy balance of obese high school girls. *Am J Clin Nutr* 1956;4:37–44.
28. Rose HE, Mayer J. Activity, caloric intake, fat storage, and energy balance of infants. *Pediatrics* 1968;41:18–28.
29. Hutson EM, Cohen NL, Kunkel ND, Steinkamp RS, Rourke MH, Walsh HE. Measures of body fat and related factors in normal adults. *J Am Diet Assoc* 1965;47:179–86.
30. Montoye HJ, Block WD, Metzner HL, Keller JB. Habitual physical activity and serum lipids: males, age 16–64 in a total community. *J Chron Dis* 1976;29:697–709.
31. Baecke JAH, van Staveren WA, Burema J. Food consumption, habitual physical activity, and body fatness in young Dutch adults. *Am J Clin Nutr* 1983;37:278–86.
32. Heunemann RL. Teenagers' activities and attitudes toward activity. *J Am Diet Assoc* 1967;51:433–40.
33. Stefanik PA, Heald FP, Mayer J. Caloric intake in relation to energy output and obese and nonobese adolescent boys. *Am J Clin Nutr* 1959;7:55–61.
34. Bradfield RB, Paulos J, Grossman L. Energy expenditure and heart rate of obese high school girls. *Am J Clin Nutr* 1971;24:1482–8.
35. McCarthy MC. Dietary and activity patterns of obese women in Trinidad. *J Am Diet Assoc* 1966;48:33–7.
36. Maxfield E, Konishi F. Patterns of food intake and physical activity in obesity. *J Am Diet Assoc* 1966;49:406–8.
37. Braitman LE, Adlin EV, Stanton JL Jr. Obesity and caloric intake: the National Health and Nutrition Examination Survey of 1971–1975 (HANES I). *J Chron Dis* 1985;9:727–32.
38. Blair D, Habicht J-P, Sims EAH, Sylwester D, Abraham S. Evidence for an increased risk for hypertension with centrally located body fat and the effect of race and sex on this risk. *Am J Epidemiol* 1984;119:526–40.
39. Mahalko JR, Johnson LK. Accuracy of predictions of long-term energy needs. *J Amer Diet Assoc* 1980;77:557–61.
40. Beaton GH, Milner BA, Corey P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am J Clin Nutr* 1979;32:2546–59.
41. Bowen PE, Custer PB. Reference values and age related trends for arm muscle area, arm fat area and sum of skinfolds for United States adults. *J Amer Col Nutr* 1984;3:357–76.
42. Beal VA. The nutritional history in longitudinal research. *J Amer Diet Assoc* 1967;51:426–32.
43. van Staveren WA, deBoer JO, Burema J. Validity and reproducibility of a dietary history method estimating the usual food intake during one month. *Am J Clin Nutr* 1985;42:554–9.
44. National Center for Health Statistics, A McDowell, A Engle, JT Massey, K. Maurer. Plan and operation of the Second National Health and Nutrition Examination Survey, 1976–80. (*Vital and Health Statistics,* Series 1–15). [DHEW publication no. (PHS)81-1317].
45. Beaton GH, Milner J, McGuire V, Feather TE, Little JA. Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins and minerals. *Am J Clin Nutr* 1983;37:986–95.
46. Richard L, Roberge AG. Comparison of caloric and nutrient intake of adults during week and weekend days. *Nutr Res* 1982;2:661–8.
47. Beaudoin R, Mayer J. Food intakes of obese and non-obese women. *J Amer Diet Assoc* 1953;29:29–33.
48. van den Berg AS, Mayer J. Comparison of a one-day food record and research dietary history on a group of obese pregnant women. *J Amer Diet Assoc* 1954;30:1239–44.
49. Linusson EEI, Sanjur D, Erickson EC. Validating the 24-hour recall method as a dietary survey tool. *Arch Latinoam Nutr* 1975;24:277–94.
50. Bray GA, Zachary B, Dahms WT, Atkinson RI, Oddie TH. Eating patterns of massively obese individuals. Direct vs. indirect measurements. *J Amer Diet Assoc* 1978;72:24–7.
51. Krantzler NJ, Mullen BJ, Schutz HG, Grivetti LE, Holden MS, Meiselman HL. Validity of telephoned diet recalls and records for assessment of individual food intake. *Am J Clin Nutr* 1982;36:1234–42.
52. Posner BM, Borman CL, Morgan JL, Borden WS, Ohls JC. The validity of a telephone-administered 24-hour dietary recall methodology. *Am J Clin Nutr* 1982;36:546–53.
53. Sempos CT, Johnson NE, Smith EL, Gilligan C. Effects of intraindividual and interindividual variation in repeated dietary records. *Am J Epidemiol* 1985;121:120–30.
54. Dawber, TR, Pearson G, Anderson P, et al. Dietary assessment in the epidemiologic study of coronary heart disease: the Framingham study. II Reliability of measurement. *Am J Clin Nutr* 1962;11:226–34.
55. Reshef A, Epstein LM. Reliability of a dietary questionnaire. *Am J Clin Nutr* 1972;25:91–5.
56. Jain M, Howe GR, Johnson KC, Miller AB. Evaluation of a diet history questionnaire for epidemiologic studies. *Am J Epidemiol* 1980;111:212–19.
57. Trulson MF, McCann MB. Comparison of dietry survey methods. *J Amer Diet Assoc* 1959;35:672–6.
58. Block G. A review of validations of dietary assessment methods. *Am J Epidemiol* 1982;115:492–505.
59. Young CM, Hagan GC, Tucker RE, Foster WD. A comparison of dietary study methods. II Dietary history vs. seven-day record vs. 24-hour recall. *J Amer Diet Assoc* 1952;28:218–21.
60. Morgan RW, Jain M, Miller AB, Choi NW, Matthews V, Munan L, Burch JD, Feather J, Howe GR, Kelly A. A comparison of dietary methods in epidemiological studies. *Am J Epidemiol* 1978;107:488–98.
61. Rasanen L. Nutrition survey of Finnish rural children. VI Methodological study comparing the 24-hour recall and the dietary history interview. *Am J Clin Nutr* 1979;32:2560–7.

Chapter 8

Measuring Dietary Patterns in Surveys of Physical Fitness and Activity

Catherine E. Woteki, Ph.D.
National Center for Health Statistics

Introduction

In nutrition surveys, five types of measurements are needed to completely characterize the nutritional status of a person or a population: food and nutrient intakes, body measurements, hematological and biochemical tests, physical examination for the presence of clinical signs of deficiency or toxicity, and medical history. No one method alone is sufficient for assessing the nutritional status of an individual or a group of individuals.

In approaching the design of a survey in which nutrition, physical fitness, and activity are to be assessed, the selection of which nutritional assessment methods are to be used depends on the answers to two questions: What aspects of nutrition are of interest? How suitable are the available measures? A clear statement of the goals of the survey and of hypotheses to be tested is helpful in deciding which aspects of nutrition are of interest and which broad categories of measurements should be included. Consideration as to the practicality, reliability, and validity of available measurements will determine their suitability for the survey under design.

In a recent article, Ware et. al. (1) discuss issues of practicality, reliability, and validity when selecting measures of health status. Their discussions are also pertinent to the selection of nutrition status indicators. The authors caution that health status, like nutritional status, cannot be observed directly and that investigators can only make inferences about health or nutritional status from fallible indicators. Whether a given measurement is practical depends upon the total measurement resources available, respondent burden, and analytical resources. Total measurement resources and the priority assigned to nutritional assessment will to a large degree determine the amount of staff time and resources to be dedicated to nutritional assessment. The burden on the respondent in terms of time, inconvenience, and physical discomfort will affect refusal rates, rates of missing responses, and administration time in followup of nonrespondents. The degree of accuracy needed depends on the purpose of the study, with more sensitive and specific measures required for clinical decisionmaking than for comparing groups of people. Validity of nutrition status indicators should be evaluated within the context of the particular health or nutritional components of interest to the study. The literature is limited with respect to reliability and validity of nutritional status measurements and offers opportunities for future research.

Data available from national nutrition surveys, with particular emphasis on dietary surveys and their strengths, weaknesses, and limits to interpretation are described here. The food frequency questionnaires and the 24-hour recall of food consumption are discussed in detail. The dietary methodology used in previous Health and Nutrition Examination Surveys is described. Other aspects of nutritional assessment—hematological and biochemical tests, body measurements, physical examination, and medical history—are reviewed elsewhere in this chapter.

Data available from national surveys

The Federal Government collects a broad range of nutrition-related information under the general heading of the National Nutrition Monitoring System (NNMS). Five major categories of information are collected: Health status measurements, food consumption measurements, food composition, dietary knowledge and attitudes, and food supply determinations. The activities in each of these categories and the agencies responsible for them are described below.

Health status measurements

The Department of Health and Human Services (DHHS) through the National Center for Health Statistics (NCHS), the Centers for Disease Control (CDC), and other agencies collects a broad array of data on the health status of the American population. Two programs are specifically oriented to nutrition. These are the National Health and Nutrition Examination Surveys (NHANES) conducted by NCHS and the Coordinated State Surveillance System (CSSS) conducted by CDC. Specific nutrition-related indicators for which data are available from NHANES and CSSS are summarized in table 1.

Table 1. Sources of data on indicators of nutritional status

	National Health and Nutrition Examination Survey (NHANES)		Coordinated State Surveillance System (CSSS)	
Indicator	I	II	Pediatric	Pregnant
Height	*	*	*	—
Weight	*	*	*	—
Skinfold thickness				
Triceps	*	*	—	—
Subscapular	*	*	—	—
Head circumference	—	*	*	—
Hemoglobin	*	*	*	*
Hematocrit	*	*	*	*
RBC count	—	*	—	—
WBC count	*	*	—	—
MCV	—	*	—	—
MCHC	—	*	—	—
RBC protoporphyrin	—	*	—	—
Serum iron	*	*	—	—
Serum TIBC	*	*	—	—
Transferrin saturation	*	3–74 yr	—	—
Ferritin	—	[a]	—	—
Serum zinc	—	3–74 yr	—	—
Serum copper	—	3–74 yr	—	—
Serum vitamin C	—	3–74 yr	—	—
Serum albumin	—	3–74 yr	—	—
Serum vitamin A	—	3–11 yr	—	—
RBC folate	—	[a]	—	—
Serum folate	—	[a]	—	—
Serum cholesterol	*	20–74 yr	—	—
Serum triglycerides	—	20–74 yr	—	—
High-density lipoprotein	—	20–74 yr	—	—
Serum vitamin B_{12}	—	[a]	—	—
Breast feeding	—	[a]	—	—
Low birth weight	—	—	—	—

* Data available, for NHANES I, 1–74 years; for NHANES II, 6 months–74 years.

— Data not available

[a] Performed on a subsample of persons 3–74 years.

NHANES is conducted periodically by NCHS and is the cornerstone of Federal efforts to monitor the overall nutritional status of the American people. NHANES consists of a series of surveys carried out on a representative sample of the civilian noninstitutionalized U.S. population, comprising over 20,000 persons each. The surveys, which include health histories, dietary interviews, physical examinations, and laboratory measurements, provide information on national health and nutritional status. Two national surveys have been completed—NHANES I (1971–1974) and NHANES II (1976–1980)—and a survey of the health and nutritional status of Hispanic Americans was conducted in 1982–84. Descriptions of the plan and operation of these surveys are published by NCHS (2–4).

Through NHANES, physical and biochemical measurements are made that provide information about a number of nutrition-related conditions, including growth retardation, anemia, obesity, heart disease, hypertension, diabetes mellitus, vitamin and mineral deficiency or toxicity, and heavy metal and pesticide exposures.

The types of measurements include physical examinations, anthropometry, hematological assessments, biochemical analyses of blood and urine, x rays, and functional assessments, as well as a health history.

The Nutrition Division, Center for Health Promotion and Education, CDC contributes to nutrition monitoring through a nutrition surveillance program. The CSSS monitors the nutritional status of high risk pediatric and pregnant populations through the collection of measurements readily available, such as height, weight, hemoglobin and/or hematocrit. The system uses information from service delivery programs operated by selected State and metropolitan health jurisdictions to provide data on the prevalence of major nutritional problems in the targeted groups. The composition of the groups under surveillance is determined by their socioeconomic status, their proximity to a nonprivate outpatient clinic, and by the fact that selected State health departments are cooperating with the CDC program in obtaining and utilizing nutrition-related data. The indicators of health status measured by the CSSS are limited to a relatively few indexes related to nutrition problems identified in the NHANES activities. The CSSS provides information about the prevalence of overweight, underweight, retarded linear growth, and anemia among high risk children. In addition, pregnant women are kept under surveillance with attention to indicators such as anemia, abnormal weight changes, fetal survival, birth weight of the child, and whether breast or bottle feeding is used and for how long. CDC publishes annual reports on its findings (5).

Food consumption measurements

Four surveys provide information on the food people consume. They are the Nationwide Food Consumption Survey (NFCS) and the Continuing Survey of Food Intakes by Individuals (CSFII) conducted by the Human Nutrition Information Service (HNIS) of the U.S. Department of Agriculture (USDA), NHANES (already described), and the Food Usage Survey conducted by the Food and Drug Administration (FDA). Nutrients for which estimates are made in NFCS, NHANES, and the National Food Supply (discussed in "Food Supply determination"), with an indication of the completeness of the food composition data underlying the estimates, are shown in table 2.

The NFCS has been conducted at roughly 10-year intervals since the mid-1930's. It is actually two surveys: a survey of household food use that has been conducted

Table 2. Sources of data on nutrients and other food constituents[a]

Nutrient or food constituent	Food supply	Nationwide Food Consumption Survey (NFCS)		NHANES		Total diet study	Food composition
		Household	Individual	I	II		
Water	—	—	—	—	—	—	*
Energy (kcal)	*	*	*	*	*	—	*
Protein, total	*	*	*	*	*	—	*
Amino acids	—	—	—	—	—	—	b,c
Carbohydrate, total	*	*	*	—	*	—	*
Sugars	*	—	c	—	—	—	—
Lipids:							
Total fat	*	*	*	—	*	—	*
Saturated fat	*	—	—	—	*	—	b
Oleic acid	*	—	—	—	*	—	b
Total monounsaturated	—	—	—	—	—	—	b
Linoleic acid	*	—	—	—	*	—	b
Total polyunsaturated	—	—	—	—	—	—	b
Cholesterol	*	—	c	—	*	—	b
Vitamins:							
A value, IU	*	*	*	*	*	—	*
RE	—	—	—	—	—	—	b
Carotene	—	—	—	—	—	—	—
E	—	—	—	—	—	—	b, c
Thiamin (B_1)	*	*	*	*	*	—	*
Riboflavin (B_2)	*	*	*	*	*	—	*
Niacin (preformed)	*	*	*	*	*	—	*
Pantothenic acid	c	—	—	—	—	—	b,c
B_6	*	*	*	—	—	—	*
Folic acid	c	—	—	—	—	—	b,c
B_{12}	*	*	*	—	—	—	*
C	*	*	*	*	*	—	*
Minerals:							
Calcium	*	*	*	*	*	*	*
Phosphorus	*	*	*	—	*	*	*
Magnesium	*	*	*	—	—	*	*
Iron	*	*	*	*	*	*	*
Iodine	—	—	—	—	—	—	—
Sodium	*	—	c	—	*	*	*
Potassium	*	—	—	—	*	*	*
Copper	—	—	—	—	—	*	b,c
Zinc	*	—	—	—	—	*	b
Manganese	—	—	—	—	—	*	b,c
Selenium	—	—	—	—	—	*	—
Chromium	—	—	—	—	—	*	—
Fiber, crude	*	—	—	—	—	—	*
Dietary	—	—	—	—	—	—	c
Alcoholic beverage	—	*	*	—	*	*	*

* Data available

— Data unavailable

[a] Alcoholic beverages are included.

[b] Nutrient data available at the completion of U.S. Department of Agriculture Handbook No. 8 revision.

[c] Incomplete or questionable.

six times and a survey of individuals' food intakes that has been conducted twice.

The most recent NFCS was conducted in 1977–78 and consisted of a basic survey of 15,000 household members and five additional surveys in Alaska, Hawaii, Puerto Rico, and in the continental U.S. of low-income and elderly households. In the NFCS, the method of collecting household data is a 7-day recall of food used. The individual household members recall 1 day's food intake and keep a diary for 2 additional days. Descriptions of the survey design and results are published (6, 7) and are also available on computer tapes for researchers.

Recognizing the need for more frequent food consumption information, in 1985 the Department of Agriculture initiated an annual survey called the Continuing Survey of Food Intakes by Individuals (CSFII). The survey selects three panels: women 19–50 years of age and their children 1–5 years, all incomes; men 19–50 years, all incomes; and a low-income sample of women 19–50 years and their children 1–5 years. They are interviewed six times over the course of a year and are asked to report on the food eaten on the previous day. Descriptions of the 1985 and 1986 surveys have been published (8, 9).

Using commercial market research data bases (A.C. Nielsen Co.), the FDA conducts, on a biennial basis, a survey of a statistically representative sample of products representing major food classes from the total packaged food supply. The Food Usage Survey involves approximately 1,700 individual food brands representing about 44 percent of the packaged food supply in retail dollar terms, which in turn is generalizable to the total packaged food supply. The ingredient data are the basis for the FDA food ingredient data bank, and they are used for multiple special studies. Significant changes in aggregated public-purchasing patterns and the food industry's reactions can be followed. For example, changes in public purchasing practices associated with avoiding specific components of foods (e.g., sugars, food additives) may be quickly identified. The same data base is used to measure changes in aggregated public purchasing practices of nutritionally modified foods, such as fortified foods, low sodium and reduced sodium foods, low cholesterol and reduced calorie foods. Through this means, public responses to nutrition information and education programs, new labeling approaches, media coverage, and other societal events can be measured. This type of information permits estimates of the impact of programs initiated by Government and the private sector in the interest of improving nutritional health.

Food composition

Assessing the nutritional adequacy of diets reported in NHANES and NFCS would be impossible without information on the nutrient content of foods. Four different activities contribute to our knowledge of food composition: the Nutrient Data Bank, the Total Diet Study, Labeled Food Surveillance, and research.

The USDA Nutrient Data Bank (NDB) is the major mechanism for collecting, evaluating, storing, and collating nutrient composition data for individual foods (10). The task is substantial, given the fact that there are some 10,000–15,000 food items in the U.S. food supply, and data are being acquired for 60–100 nutrients or other food components. Data are being collected and entered into the NDB on a continual basis, but the availability of data on some nutrients is limited by the lack of suitable methods. Sources of data include a number of Federal Government laboratories, including USDA's Nutrient Composition Laboratory; university research under Government sponsorship; and food nutrient analyses conducted by industry in support of the nutrition labeling program. The development of the NDB is keyed to the process of revising Agriculture Handbook No. 8, "Composition of Foods . . . Raw, Processed, Prepared" (11), which is the standard reference table on food composition.

FDA conducts annually its Total Diet Study to estimate average consumption of important components of the diet. The study provides a tracking system for specific indicators of significant changes in the nutritional quality of the national food supply. For example, through this study, FDA has documented the existence of a higher quantity than desirable of iodine in the food supply (12). The quantity of iodine is now being monitored, particularly for some of the food classes noted previously to have high levels, such as milk and cereal grain products (13).

The scientific base for the Total Diet Study has been updated from food consumption data obtained in 1965 to the recently available NHANES II and NFCS data (14). Approximately 200 individual foods are involved that represent about 90 percent of the total foods consumed in the United States. Sampling is done in 30 urban areas in the United States, and analyses for dietary content of seven minerals (iodine, iron, sodium, potassium, copper, magnesium, and zinc) are performed. This is the only extant system for annual chemical analytical measurement of average intakes of pesticides, heavy metals, and environmental contaminants. The study has recently been converted from the measurement of nutrients, pesticides, and contaminants in food composites to estimation in individuals foods.

The FDA has maintained since 1977 both a surveillance and a compliance program for nutrition labeling. A statistical sample of the 40 percent of processed foods that bear nutrition labels is analyzed for many nutrients on a continuous basis. Annually, approximately 300 foods are analyzed for eight nutrients, involving in excess of 2,000 individual analyses. This surveillance program permits FDA to track the evolution of nutrition labeling in the food supply, assure necessary levels of accuracy of label values, and identify segments of the industry that require encouragement. This activity also permits early identification by FDA of new fortification practices by industry. When combined with consumer studies, reasonable assessments of the value of nutrition labeling are possible.

USDA's Nutrient Composition Laboratory, located in Beltsville, Md., provides essential data on the nutrient content of foods as consumed in the United States by analyzing the nutrient content of foods with tested, dependable assay techniques; developing either new or improved methods for the analysis of nutrients in foods; developing sound sampling techniques to ensure that representative samples are analyzed; and conducting research on the effect of food processing procedures and transportation and marketing methods, as well as home, institution, and restaurant food preparation procedures on the nutrient composition of foods. Data from these studies are used to update the food composition values in the Nutrient Data Bank.

Dietary knowledge and attitudes

Annually since 1978, FDA has conducted the Public Attitudes Survey based on a national probability sample of food purchasers to measure public attitudes, knowledge, and practices about food and nutrition. About one–half of the survey content is concerned with such matters as opinions about nutrition, food quality, and food regulation; and it is repeated every year for the purpose of tracking changes over time. The other half involves new areas of interest or concern to FDA.

This survey involves detailed interviews in the homes of approximately 1,500 individuals primarily responsible for household food purchases, about 85 percent of whom are women and 15 percent are men. Studies of this type permit monitoring of public attitudes and practices about foods and nutrition, as well as identification of concerns and elements of confusion. It is through this mechanism that the predominance of avoidance practices in food purchases through use of food labels was identified (e.g., avoidance of fats and sugars). In addition, assessments are made of the influence of nutrition misinformation and the public's ability to comprehend food label information.

Food supply determination

Each year since 1909, USDA has calculated the nutrients available for daily per capita consumption from estimates of per capita food availability (retail weight). No deductions are made for waste of food in the distribution system or in the home, for use as pet food, or for loss of nutrients during the preparation of food. Adjustments are made for nutrients added to the food supply through enrichment and fortification, e.g., iron, thiamin, riboflavin, and niacin added to flour and cereal products (15).

Estimates of nutrient availability are based on the quantities of 350 foods that "disappear" into the U.S. food-marketing system (15). Hence, the name *disappearance data* is frequently used in referring to these estimates. Levels of food energy (kilocalories) and 15 nutrients (table 2) are calculated using food composition data collected and published by USDA. (See discussion above under "Food composition".)

Dietary questionnaires used in previous National Health and Nutrition Examination Surveys

The five major components of previous National Health and Nutrition Examination Surveys were a household questionnaire, a medical history questionnaire, a dietary questionnaire, examination by a physician, and special procedures and tests. The household questionnaire consisted of questions about family relationship; age, sex, and race of family members; housing information; occupation, income, and education level of each family member; and participation in the food stamp program and school breakfast and lunch programs. Separate medical history questionnaires were used depending on the age of the sample person, one questionnaire for children 6 months through 11 years and another for persons 12–74 years. Both the household questionnaire and the medical history were administered in the respondent's home.

When the sample persons arrived at the mobile examination center, they were scheduled through the dietary interview, physician's examination, and special procedures and tests. Procedures and tests for the nutritional assessment included body measurements and urine and blood tests. From blood samples taken in the center, a number of nutrition-related assays were done. These included serum albumin, serum vitamins A and C, serum lipids (cholesterol, triglycerides, and high-density lipoproteins), protoporphyrin, serum iron, total iron binding capacity, serum zinc, and serum copper. Red cell folates, serum folates, serum ferritin, and serum vitamin B_{12} were determined on blood samples with abnormal complete blood count, hemoglobin, hematocrit, or MCV and on a subsample of all other blood samples.

The dietary questionnaires consisted of a 24-hour recall, a food frequency, a supplemental dietary questionnaire, and specific questions on medication and vitamin and mineral supplement usage. All interviews were conducted by trained interviewers who held at least a bachelor's degree in home economics. Copies of the questionnaires used in NHANES II and HHANES can be found in the program description reports (3).

In the 24-hour recall, respondents were asked to report all foods and beverages consumed on the previous day. Respondents estimated the size of the portions consumed by referring to food models. In addition to foods and portion sizes, interviewers asked about what time of day the food was eaten and its source. The time of day was coded as one of five ingestion periods: morning, noon, between meals, evening, or total day. The source of the food was coded as home, school, restaurant, or other.

Each food item was coded by the interviewer within 72 hours of the interview. The food code book developed for the NHANES II contained 5-digit food codes for approximately 2,500 food items. Each food item was identified by name, including brand names if appropriate; by whether it was raw, dry, canned or frozen; by how it was prepared; and for mixed dishes without food codes, by the major ingredients. A food composition data base updated from NHANES I was used to calculate the energy, vitamin, and mineral content of the reported foods. Modifications to the NHANES I data base included new data from USDA's revised Handbook No. 8, and food composition data from food companies on new products and brand name products of unique formulation. The food composition data base for HHANES and future surveys will be the data base in current use by the U.S. Department of Agriculture.

The questionnaire used to determine the frequency of food consumption has changed with each survey. In NHANES II and HHANES, the food frequency elicited information about the consumption of 18 food groups over the previous 3 months. These groups can be related to the 13 major groups in NHANES I, but they are progressively more detailed. The frequency was coded as a whole number, never, or unknown. The interval at which the food was usually eaten was coded as never, daily, weekly, or less than weekly. One question was asked about how often the salt shaker was used at the table. Responses to this last question could be assigned to one of three codes: rarely or never, occasionally or seldom, frequently or always.

The supplemental dietary questionnaire contained questions about whether the respondent was on a special diet, what type, and for how long. One question asked about use of nine medications in the previous week. These were commonly prescribed medications that might interfere with test results or affect interpretation of results. Another question related to problems preventing the respondent from obtaining needed groceries. The final question asked about trouble swallowing, pain, nausea and vomiting following eating, and loss of appetite.

The medication, vitamin, and mineral usage questionnaire has evolved over time. NHANES II requested specific information about the brand name, manufacturer's name, and reason for using vitamin or mineral supplements and medications. Sample persons were left the form and asked to fill it out and bring it with them to the examination centers. Because of low response, the procedures for collecting medication and supplement usage were changed in HHANES. During the household interview, sample persons were asked to bring their medicines and supplements to the interviewer who recorded the label information.

The quality of the dietary component was controlled at several levels. Before the survey began, the dietary interviewers were trained in interview techniques and in how to code the 24-hour recall. A manual describing the procedures to be followed was issued to each interviewer (16). Periodically, the forms were reviewed and evaluated, and instructions were issued to the interviewers to promote consistency. Interviewers exchanged coded 24-hour recall forms to check each other's work, and forms were also reviewed by the field staff before forwarding to headquarters. At every location, each interviewer tape recorded two interviews with randomly selected subjects. The recordings were evaluated at headquarters for adherence to procedures. Comparisons were made at headquarters of the mean values and frequency distributions by stand location and by interviewer to detect unusual results by location and systematic errors by interviewers. Foods for which no appropriate food codes existed were forwarded to headquarters for assignment of new code numbers.

Uses and limitations of survey methods

Uses of dietary data

NHANES dietary data have been put to four types of uses: relating diet and demographic characteristics, relating diet and health characteristics, determining interactions of diet and nutritional status indicators, and tracking trends in diet and nutrient intakes over time.

In relating diet to demographic characteristics of the population, the major question to be asked is: What are the food consumption patterns and nutrient intakes of subpopulations of the United States categorized by such characteristics as age, race, sex, income, occupation, and education? The NHANES dietary data can answer questions such as: How do nutrient intakes and food consumption patterns of persons differ by level of education? What are the regional differences in consumption of certain food groups?

NHANES data have been used to relate food consumption patterns and nutrient intakes of subpopulations of the United States to indicators of health status. With the addition of an assessment of physical fitness and activity, another dimension will be added to the possible statistical analyses. Specific questions that could be addressed include: How do nutrient intakes compare with the recommended dietary allowances and other dietary guidelines? What dietary patterns and nutrient intakes are associated with differing levels of fitness or activity?

Examining interactions among nutrition-related variables, NHANES data could compare among dietary intake, biochemical status, anthropometry, and presence or absence of health conditions. Questions that could be addressed by the data include: What are the relationships among dietary intake, biochemical indicators, and fitness status for persons who smoke, use vitamin or mineral supplements, or use oral contraceptives? Are those who take vitamins and other dietary supplements the ones who need them? Do the diets, fitness levels, and activity of subpopulations with high serum cholesterol levels differ from those with lower levels?

Changes over time in food and nutrient intakes could be tracked and correlations made with health variables. Examples of questions that could be posed to the data include: What changes in obesity, diet, and activity patterns will take place in the next 10 years? Will activity levels and diet help explain the continuing decline in serum cholesterol values among men and women?

Estimates of food and nutrient intakes by individuals

Of the methods used by Federal surveys, only one would be appropriate for studies of the association between diet, physical fitness, and activity—estimates of food and nutrient intakes by individuals.

National surveys collect information about food and nutrient intakes by individuals for two purposes—to estimate intakes by groups of individuals who share some characteristic (e.g., age, sex, race, income, education) and to relate estimated intakes by individuals to some condition (e.g., cholesterol intake and serum cholesterol level).

The national surveys use three methods to collect information from individuals about their food intakes: 24-hour recalls (NHANES, NFCS, and CSFII), food diaries (NFCS), and food frequency questionnaires (NHANES). The 24-hour recall and food diaries attempt to capture actual dietary intakes during specified periods of time; and frequency questionnaires attempt to capture usual intakes over longer time spans, ranging from a few months to a year. The 24-hour recall and diaries yield quantitative estimates of foods and nutrients ingested. Several articles (17–20) have appeared that review the reliability and validity of these methods. The literature is more extensive for the 24-hour recall method than for diaries or frequencies.

The 24-hour recall and diaries can be used in cross-sectional surveys to determine the foods ingested and to calculate nutrient intakes. Estimates can be made of the mean, median, and distribution of intakes by specific groups described by age, sex, race, income, or other characteristics. The average of repeated recalls and diaries can be used to determine usual food intake by individuals. The minimum number of days of observation required to characterize usual intake varies with the nutrient of interest and by the amount of intraindividual and interindividual variations.

Both the 24-hour recall and diaries are limited in that they rely on respondents' ability to remember and describe to an interviewer what they ate and drank, or to remember to record what was eaten. Probably more frequently than one would like to admit, the diary method becomes a recall when the respondent fails to record what was eaten and relies on memory to complete the diary before it is returned to the interviewer.

The precision of estimates of individuals' usual intakes based on single, 1-day observation is low, largely because of high intraindividual variance (21,22). The major components of variance in the 24-hour recall are sex, day of the week, interindividual and intraindividual variations. Men tend to eat more than women, so most studies of variability in intakes treat the sexes separately. The day of the week for which the recall was obtained can be a significant source of variance for women with weekend intakes being higher than weekdays (21). However, the two largest sources of variance are interindividual and intraindividual. Interindividual variations of energy and nutrient intakes are relatively constant, but intraindividual variations differ with the nutrient under consideration. Intraindividual variation is usually larger than interindividual variation, and it includes methodological errors as well as the true day-to-day variation in intake within respondents. The high intraindividual variability noted in many studies suggests a wide daily variation in the composition of self-selected diets.

Table 3. Food composition knowledge as percent of applicable food categories containing substantial data

Nutrient	Percent of food categories with substantial data
Total protein	82
Total fat	54
Niacin	53
Riboflavin (vitamin B_2)	53
Thiamin (vitamin B_1)	53
Calcium	46
Iron	44
Phosphorus	44
Sodium	44
Potassium	44
Cholesterol	42
Magnesium	40
Zinc	40
Copper	37
Fatty acids	34
Vitamin A	27
Vitamin B_6	20
Simple sugars	10

Note: U.S. Department of Agriculture has divided food supply into 42 broad categories. Because some nutrients are known to be absent from some foods and food categories, not all nutrient analyses are desired or needed for all 42 categories. Percent of food categories with substantial data is calculated only for categories in which nutrient is suspected to be present in important amount. "Substantial data" means sufficient information to establish normal amount of nutrient in most foods in category.

Source: National Research Council, 1984 (25).

The food composition data base used to calculate nutrient levels from reported foods is a contributor to error associated with 24-hour recalls and diaries. As indicated in table 3, there are substantial gaps in the tables of food composition. For some nutrients of interest only scanty data exist. Taking simple sugars as an example, only 10 percent of food categories suspected to contain simple sugars in important amounts have sufficient data upon which to establish an estimate.

One of the major dietary research questions has been how many days of observation (obtained by recall or by record) are needed to characterize an individual's "usual" intake of nutrients. In studies of diet and disease, the large intraindividual variation in daily food intake may mask any significant associations with a disease, biochemical marker, or other physiological outcomes of interest. This occurs because as the ratio of intraindividual variation to interindividual variation increases, the power to distinguish between individuals decreases (23). If one knows the magnitude and potential sources of variation in the diet, it is theoretically possible to minimize the effects of intraindividual variation by repeating the measurement a sufficient number of times to estimate the usual intake of the individual. Using published variance component ratios (23), one can estimate the number of days of observation needed in studies of the relationship between diet and disease.

Food frequency questionnaires are useful to classify individuals and groups by food intake characteristics (18). Frequencies can be designed to cover the entire diet (as has been done by NHANES), or they can be targeted to one or more foods or groups of foods of research interest. Information on the frequency of consumption of foods can be combined with estimates of usual portion sizes to yield estimated quantities of foods consumed, and with special food composition tables to estimate usual intake of nutrients.

Although food frequency questionnaires are thought to provide data more representative of respondents' usual diets than recalls or records, frequencies suffer from a number of limitations (20). Because frequencies cover longer periods of time than recalls or diaries, they are more difficult to validate. Usually, validation studies have attempted to compare frequencies to another questionnaire method. Possibly the greatest problem for frequency questionnaires is the inaccuracy of estimates of intake. The relative accuracy of nutrient intake data for specific days is less than for recalls or records, and the estimated nutrient values are less absolute. Frequency questionnaires may preserve individuals' rank ordering (24), but not to the extent that repeated recalls or records do (20). And, like the recall, frequencies rely on respondents' ability to accurately remember and report their usual food intake.

Frequency questionnaires have several merits when compared with recalls and records (18,20). The "true mean intake" of an individual is estimated directly. Frequencies are easier to administer and code. And, most important of all, problems caused by intraindividual variation are less serious. They are also more likely to include foods that are infrequently consumed or eaten only when in season.

Summary

Reviewed here are the strengths, weaknesses, and limits to interpretation of dietary intake methods used by national nutrition surveys, emphasizing the 24-hour recall and food frequency methods. These dietary assessment methods are but one of five categories of nutritional assessment techniques. The other four include body measurements, hematological and biochemical tests, medical examination for the presence of clinical signs of deficiency or toxicity, and medical history. No one method alone is sufficient for assessing the nutritional status of individuals or groups.

In the hypothesis-generating phases of research on diet and cardiovascular diseases and diet and cancer, correlations of per capita availability of food and nutrients with mortality from these diseases in several countries proved useful. However, the approach is limited to the extent that similar data are available from several other countries and that the association observed is true and not spurious. The 24-hour recall method has proved to provide reproducible estimates of the mean intakes of population groups, but multiple days' information are necessary for characterizing an individual's usual nutrient intake. Food frequency questionnaires are useful when one is interested in food rather than nutrient consumption of individuals or groups. The 24-hour recall, food diary, and food frequency methods can be used to develop a descriptive epidemiology of food intake within U.S. population groups with specified fitness and activity characteristics or to monitor the prevalence of dietary risk factors within the population. They are limited to the extent that food composition data are available. When selecting a dietary assessment method to be used in a survey of diet and physical fitness, three points must be kept in mind: practicality in terms of respondent burden and analysis resources, reliability, and validity.

References

1. Ware, J.E., Brook, R.H., Davis, A.R., and Lohr, K.N.: Choosing measures of health status for individuals in general populations. *Am. J. Pub. Health* 71:620–625, 1981.
2. National Center for Health Statistics: Plan and operation of the Health and Nutrition Examination Survey, 1971–73. *Vital and Health Statistics.* Series 1, No. 10a. DHHS Pub. No. (PHS) 79–1310. Public Health Service. Washington. U.S. Government Printing Office, 1973.
3. National Center for Health Statistics: Plan and operation of the second National Health and Nutrition Examination Survey, 1976–80. *Vital and Health Statistics.* Series 1, No. 15. DHHS, Pub. No. (PHS) 81–1317. Public Health Service. Washington. U.S. Government Printing Office, 1981.
4. National Center for Health Statistics: Plan and Operation of the Hispanic Health and Nutrition Examination Survey, 1982–84. *Vital and Health Statistics.* Series 1, No. 19. DHHS Pub. No. (PHS) 85–1321. Public Health Service. Washington. U.S. Government Printing Office, 1976.
5. Centers for Disease Control: Nutrition surveillance 1980. DHHS Pub. No. (CDC) 83–8295. Public Health Service. Washington. U.S. Government Printing Office, 1982.
6. Human Nutrition Information Service: Food consumption: Households in the United States, Spring 1977. Pub. No. H-1. USDA. Washington. U.S. Government Printing Office, 1982.
7. Human Nutrition Information Service: Food intakes: Individuals in 48 States, Year 1977–78. Pub. No. I-1. USDA. Washington. U.S. Government Printing Office, 1983.

8. U.S. Department of Agriculture, Human Nutrition Information Service: Nationwide Food Consumption Survey, Continuing Survey of Food Intakes by Individuals: Women 19–50 Years and Their Children 1–5 Years, 1 Day, 1985. U.S. Department of Agriculture, CSFII Rpt. No. 85–1, 1985.
9. U.S. Department of Agriculture, Human Nutrition Information Service: Nationwide Food Consumption Survey, Continuing Survey of Food Intakes by Individuals: Women 19–50 Years and Their Children 1–5 Years, 1 Day, 1986. U.S. Department of Agriculture, CSFII Rpt. No. 86–1, 1987.
10. Hepburn, F. N.: The USDA National Nutrient Data Bank. *Am. J. Clin. Nutr.* 35:1297–1301, 1982.
11. U. S. Department of Agriculture: Composition of foods: Raw, processed, prepared. Agriculture Handbook 8–1 through 8–12. Washington. U.S. Government Printing Office, 1976–1984.
12. Park, Y.K., Harland, B.F., Vanderveen, J.E., Shank, F.R., and Prosky, L.: Estimation of dietary iodine intake of Americans in recent years. *J. Am. Dietetic Assoc.* 79:17–24, 1981.
13. Allegrini, M., Pennington, J.A.T., and Tanner, J.T.: Total Diet Study: Determination of iodine intake by neutron activation analysis. *J. Am. Dietetic Assoc.* 83:18–24, 1983.
14. Pennington, J.A.T.: Revision of the Total Diet Study food lists and diets. *J. Am. Dietetic Assoc.* 82:166–173, 1983.
15. Welsh, S.O., and Marston, R.M.: Review of trends in food use in the United States, 1909 to 1980. *J. Am. Dietetic Assoc.* 81:120–125, 1982.
16. National Center for Health Statistics: *Examination Staff Procedures Manual for the Health and Nutrition Examination Survey,* 1976–1979. Instruction Manual, Part 15a. Public Health Service. Washington. U.S. Government Printing Office, 1976.
17. Burk, M. C., and Pao, E. M.: Methodology for large-scale surveys of household and individual diets. Home Economics Research Report No. 40. Washington. U.S. Government Printing Office, 1976.
18. Block, G.: A review of validations of dietary assessment methods. *Am. J. Epidemiol.* 115:492–505, 1982.
19. Bazzarre, T.L., and Myers, M.P.: The collection of food intake data in cancer epidemiology studies. *Nutrition and Cancer* 1:22–45, 1979.
20. Life Sciences Research Office: Guidelines for Use of Dietary Intake Data. Federation of American Societies for Experimental Biology. Bethesda, Md., 1987.
21. Beaton, G. H., Milner, J., Corey, P., McGuire, V., Cousins, M., Stewart, E., de Ramos, M., Hewitt, D., Grambsch, P. V., Kassine, N., and Little, J.A.: Sources of variance in 24-hour dietary recall data: Implications for nutrition study design and interpretation. *Am. J. Clin. Nutr.* 32:2546–2559. Food and Nutrition Board, 1979.
22. Todd, K.S., Hudes, M., and Calloway, D.H.: Food intake measurement: Problems and approaches. *Am. J. Clin. Nutr.* 37:139–146, 1983.
23. Sempos, Christopher, T., Nancy E. Johnson, Evertt L. Smith and Catherine Gilligan: Effects of intraindividual and interindividual variation in repeated dietary records. *Am. J. Epidemiol.* 121(2):120–130, 1985.
24. Willett, Walter C., Laura Sampson, Meir J. Stampfer, Bernard Rosner, Christopher Bain, Jelia Witschi, Charles H. Hennekens, and Frank E. Epiezer: Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am. J. Epidemiol.* 122(2):52–65, 1985.
25. National Research Council: *National Survey Data on Food Consumption: Uses and Recommendations.* Washington, D.C. National Academy Press, 1984.

Chapter 9

Design Issues and Alternatives in Assessing Physical Activity in General Population Surveys

Thomas Stephens, Ph.D.
National Center for Health Statistics

Introduction

Chapter objectives

The relationship of physical activity and health has been a topic of study for almost three centuries (Hyde and Paffenbarger, chapter 26), and some population data on activity prevalence have been collected in the United States since 1961 (1). Questions on exercise first appeared in a National Center for Health Statistics (NCHS) survey in 1971, when the first National Health and Nutrition Examination Survey became operational (2). Despite this record, there remain major questions about physical activity of both an analytic and a descriptive nature. The exercise stimulus for cardiovascular health is just beginning to be understood (3), and major areas of uncertainty persist regarding other health outcomes, whether mental (4) or physical (LaPorte, chapter 24). Even the description of population activity levels is crude, with published estimates of the active population ranging from 15 to 78 percent (5).

These areas of ignorance interact: The absence of a consensus on an adequate exercise dose has made the definition of the active population difficult and highly variable (5); the shortage of detailed data from general population samples has rendered tenuous the inferences from clinical and community studies of activity-health relationships. With improved understanding of the dimensions of activity that are relevant to health outcomes, however, the time has arrived when meaningful data descriptive of populations can be obtained (e.g., see (6)), and when the joint collection of detailed activity and health status measures can be extremely valuable.

The proper assessment of physical activity is not a subject about which universal agreement exists, however, and it is not yet possible to identify a technique suitable for all purposes. Indeed, it is difficult to identify any technique that satisfies all the standard criteria for a good measure, even for narrowly defined purposes. This chapter therefore does not offer a universal remedy for activity assessment in general population surveys. Instead, it provides a framework to help survey designers select (or develop) the technique most appropriate to their purposes and circumstances. To achieve this, several issues relevant to assessing the physical activity of a population, including sample and survey design and operational considerations, are described.

The approach taken is to consider the vast range of techniques currently available for activity assessment and to consider their suitability for different types of population surveys. A detailed critique of individual methods is not given; this has been accomplished in recent reviews (7–9) and in another chapter in this report (chapter 24).

Framework for discussion

Two types of surveys, reflecting the principal intent of survey sponsors or designers, are used for discussion purposes throughout this paper. The types—descriptive and analytic surveys—are not mutually exclusive and in practice often blend into each other. Nevertheless, their pure forms are conceptually distinct and can be useful for identifying significant design issues. The terms are used here in their epidemiological sense: If epidemiology is "the study of the distribution and determinants of health-related states and events in populations" (10), analytic studies test the determinants and descriptive studies provide the distributions.

Examples of important analytic studies that could be effectively conducted by means of a general population survey are found throughout this report. LaPorte (chapter 24) makes a plea for seeking out the dimensions of activity that have health consequences beyond the cardiovascular system. He raises the important question of whether it is activity or fitness that benefits health. Hyde and Paffenbarger (chapter 26) stress the need to describe the natural history of exercise as it relates to cardiovascular and other disease and to include occupational with leisure activity. Blair and Kohl (chapter 23) inquire about the relationship of exercise to other health-related practices and the role of exercise in appetite regulation and weight control. Other contributions to this discussion deal with outcomes as diverse as injuries and mortality (chapter 22). Additional outstanding questions on the health impacts of activity are discussed at length in a recent special issue of *Public Health Reports* (11). All of these examples suggest studies in which activity is the independent variable.

Although more rare in a health context, analytic studies of activity as a dependent variable may also be pursued via general population surveys. For example, reasons for choosing or rejecting specified types and amounts of exercise (chapter 26) would be important intelligence for health education and promotion uses of activity data.

Descriptive studies are fundamental to planning and assessing health promotion programs. For example, the 1990 supplement to the National Health Interview Survey will be the principal test of whether, by that year, "the proportion of adults 18 to 65 participating regularly in vigorous physical exercise (is) greater than 60 percent" (12). However, describing the activity status of the general population or special subgroups is worth while only if two conditions exist, as they now do in the United States and many other developed countries: (a) Physical activity, of specified dimensions, is generally accepted as beneficial to the population and (b) individuals or groups are anxious to use the descriptive data to increase the prevalence of activity.

Because the relationship of physical activity to both its antecedents and consequents remains poorly understood, the description of activity prevalence levels has suffered from the use of inconsistent, and sometimes uninterpretable, definitions. However, recent surveys in which reasonably consistent and rigorous definitions are used and the dimensions of frequency and intensity are incorporated have produced some consensus that 15–25 percent of the adult population regularly engage in physical activity (5, 6).

Special population groups may be effectively described by findings from a general population survey if their proportions are not too small. Examples discussed in this report are children and youths (chapter 10), women (chapter 12) and the elderly (chapter 14). Others, such as persons with cardiac conditions (chapter 13), may be too rare for a general population survey to describe with adequate precision in a cost-effective manner.

One important use of descriptive data is to provide normal values for a population. With such information, a study of activity prevalence in a small area, carried out with standard techniques, can be interpreted against national standards for age and sex groups. Such a rationale, as applied to both activity and fitness parameters, was the principal motive for undertaking a general population survey in Canada in 1981 (chapter 17). Special populations such as the military (chapter 16) will also find merit in population normative data, and international comparisons (chapter 18) may even be possible. All such comparisons, in addition to their immediate value for understanding subgroup data, may lead to an enhanced understanding of the activity-fitness-health relationship. In this fashion are the lines blurred between descriptive and analytic studies. Both types will obviously benefit from the careful selection of a technique to assess physical activity.

Concepts for activity assessment

The choice or development of an assessment strategy should begin with some concept of activity and then proceed to its operationalization. Too often, the sequence of events is reversed or the conceptual stage is omitted altogether.

Three clearly distinct concepts underlie the study of physical activity. These can be labeled (a) participation counts, (b) activity patterns, and (c) energy expenditure. Each has its own uses and limitations.

Participation counts are the simplest form of activity assessment and the most limited in use. They consist of yes-no or equivalent responses to questions about participation in specified activities and are sometimes used for making prevalence estimates (13, 14). Their principal limitation is the difficulty of interpreting results when no other detail exists. For example, What does it mean for health outcomes that, as of 1972, 51 percent of American adults said they "now participate" in 1 or more of 16 listed sports (14)? Adding a frequency dimension assists greatly in interpreting the true meaning of this kind of count.

Such an addition leads to the most modest form of inquiry into activity patterns, the second conceptual category. Frequently, information about average duration is also sought, permitting calculation of total time involved in specified activities. Data on the energy requirements of activities, obtained from controlled studies in exercise physiology laboratories, permit calculation of energy expenditure values for the individual. With sufficiently detailed information, this calculation can produce a value in standardized units, such as kilocalories or equivalents for either total or leisure-time activity (e.g., chapter 17).

A complete inventory of activities, with their associated frequency, duration, and intensity, can serve a multitude of purposes essential to a descriptive study. With the calculation of a physical activity index, subpopulations can be described and readily compared. If the index is in standardized units such as kcal/kg/day, dose-response relationships can be examined, a cutoff point can be established, and the prevalence of beneficial exercise can be estimated (6). Moreover, if the consensus on the energy dose required for cardiovascular benefit is revised, new cutoff points for old data sets can be established and new prevalence rates estimated. Similarly, if the exercise stimulus were established for other health outcomes such as arthritis, osteoporosis, weight and blood pressure control, or mental health, it would be a relatively simple matter to estimate the prevalence of benefit and risk if adequate detail on the appropriate dimensions had been collected in the first place. For such purposes, type of activity is likely to be important (chapter 24) and kilocalories relatively unimportant.

Closely related to this flexibility of definition is the use of standardized units. This permits comparison among surveys and is useful for describing different populations or different points in time. The validity of such comparison rests, of course, on the use of similar questions, data-gathering techniques, sample populations, and energy-cost constants.

Although the calculation of a physical activity index is a useful outcome of obtaining detail on activity patterns, this approach also serves other purposes. Descriptive data on the popularity of individual activities, regularity of participation, and total time commitment are all made available and are frequently of interest for planning health promotion campaigns, fitness facilities, products, and services. Such data are in the spirit of basic market research.

Energy expenditure can be obtained directly rather than by derivation from activity patterns. As is true of the derived value, direct energy expenditure determinations can be useful for descriptive and analytic purposes. However, the detailed description of the exercise stimulus is lacking. This may be an important omission, depending on the nature of the objectives. If they are analytic, the difference between knowing total exercise dose and knowing frequency and magnitude of dose may be crucial. This is the difference, for example, between learning that 3+ kcal/kg/day is beneficial for cardiovascular health and learning that this dose should be in three weekly applications of at least 20 minutes each where the heart rate is 60–70 percent of maximum (15).

The treatment of job-related activity is an issue common to all three of these conceptualizations of physical activity. Theoretically, occupational activity may be combined with leisure-time activity in seeking participation counts, activity patterns, or energy expenditure; its proper description, however, calls for special procedures (chapter 21). Even though employment demands will add little to most individuals' total energy expenditure, overlooking work activity is likely to add considerably to the unexplained variance in health outcomes for entire classes of occupations, such as farmers. Any decision to include occupational activity will not only have to specify the detail required (not necessarily at the same level as for leisure-time activity) but will also have to specify the treatment of students and homemakers. Increasingly, the practice is to treat these as occupations like any other (16, 17).

Of these three concepts of activity assessment, the concept of activity patterns is the one most likely to be useful in a health-related context for either descriptive or analytic purposes. This simply results from the current lack of knowledge about the activity-fitness-health relationship, with the consequence that no large-scale activity assessment can afford to overlook a potentially important dimension of activity. At present, the relevant dimensions appear to be activity type, intensity, frequency, and duration. This is not to deny that less complex questions on activity can serve a valuable purpose in some circumstances.

Activity assessment methods

This section identifies and classifies the options currently available for assessing activity and comments on their suitability for use in a general population survey. However, it is not intended as a detailed critique of methods, as noted earlier.

Options and their suitability

Over 30 methods for activity assessment have been identified (9), a testimony to the complexity of this topic as well as the depth of interest in it. Most of these are readily classified into one of three groups: Self-report, observation and archival records, and physiological measurement. Even within these groups, there is much variation in the conceptual and operational definition of activity. For example, one recent review of eight U.S. and Canadian surveys using self-report methods found definitions of "active" ranging from 1,500 kcal/week of leisure-time expenditure to any participation within 12 months in 1 of 104 listed activities (5).

Because these methods vary in the types and quality of information they yield, they vary in their suitability to descriptive and analytic surveys. They also vary in cost, respondent burden, reactivity, and other practical con-

Table 1. Selected physiological measures of physical activity

Method [1]	Key variables	Practical considerations	Suitability for population surveys
Calorimetry Direct (31)	Heat production during work	Requires special chamber and restricted tasks	Not suitable because of cost, respondent burden, failure to measure normal range of activities
Indirect (32)	Oxygen consumption during work	Requires special apparatus to be worn by subject	Not suitable because of cost, respondent burden, restrictive effect on activities
Heart rate monitoring (33)	Heart rate during work	Requires calorimetry to equate HR to VO_2 for each subject	Not suitable because of calorimetry requirement; may also be reactive and burdensome
Motion sensors (34)	Quantity of gross motor activity; intensity as well for certain types	Easily worn and not restrictive but of generally undemonstrated validity and reliability in general population; may take a considerable length of time for reactivity to dissipate	Practical, but data are of unknown quality and with uncertain relationship to energy expenditure
Doubly labeled water (35)	Total energy expenditure over specified time period estimated from consumption of isotopically labeled oxygen	High cost per subject (9)	Not suitable on large scale because of cost; does not distinguish parameters such as frequency, duration, intensity, type of activity

[1] Reference example in parentheses.

siderations. Finally, there is considerable difference among these methods in the extent to which they have been used in previous population surveys, and thus in the amount of experience accumulated as well as the number of possibilities for comparing results across studies. Tables 1–3 summarize these features for the principal examples of each of the three main classifications of measurement type.

The physiological measures listed in table 1 are restricted to those that provide fairly direct indications of physical activity, defined as "any bodily movement produced by skeletal muscles that results in energy expenditure" (18). Not included are those measures, such as maximum oxygen uptake and weight-height indexes, that are more properly regarded as *outcomes* of activity (but only partial outcomes).

With the exception of motion sensors, all of the physiological techniques can be ruled out at the present time for practical reasons: They are simply too burdensome and/or costly for a general population survey, even one of the scope of the third National Health and Nutrition Examination Survey (NHANES III). However, the development of motion sensors and doubly labeled water bears monitoring; these methods may one day be suitable for wide application.

Of the observational-archival measures listed in table 2, only job classification is feasible for a general population survey. Direct observation, although appealing on conceptual grounds, is practical only for small, captive populations, a restriction that also applies to sports medicine clinical records. These measures suffer from the additional problem that they are available only for a self-selected population. Data on facility use, equipment sales, and the like, although often cited as evidence of a fitness boom, are more useful as social indicators than as measures of individual activity.

In sharp contrast to the physiological and observational methods are the various self-report measures listed in table 3, all of which are potentially suitable. Excluded from this list on conceptual grounds are dietary measures of physical activity (19). Although caloric intake has a demonstrated relationship to physical activity, it is not a suitable substitute for activity assessment in a general population for which energy balance is not established. Moreover, caloric intake is at least as difficult to measure accurately as is physical activity (chapter 8).

Because these self-report measures are generally acceptable on practical grounds, they have proliferated in recent years. This should not be taken as evidence of validity or reliability, however. The quality of data produced by the self-report methods is highly variable and often poor, and should be one of the principal considerations for survey planners considering these methods. The other, logically prior, consideration is the appropriateness of the topics covered by these methods.

A population survey with an objective to describe the *total* energy expenditure of the population could conceivably employ most of these self-report techniques, suitably worded to include all daily activities. However, the typical use of these methods has been to describe leisure-time activities. Total energy expenditure might be most adequately described by a combination of self-report measures and job classification (chapter 21). Describing total energy expenditure requires special procedures for those outside the traditional work force who may have substantial amounts of time that is neither

Table 2. Selected observational and archival measures of physical activity

Method [1]	Key variables	Practical considerations	Suitability for population surveys
Behavioral observation (36)	General activity level at specified time points; potentially, total activity, type, frequency, intensity, duration	Requires trained observers with attendant costs, restricts normal activity, need to sample representative time points	Not suitable because of cost and probable reactivity
Job classification/time and motion studies (chapter 21) (37)	Work-related energy expenditure on ordinal scale, types of activity and time parameters (frequency, duration of tasks)	Requires detailed up-to-date analysis of industry-specific occupational categories, with seasonal factors accounted for	Suitable for this type of activity if analyses available; unobtrusive and inexpensive
Sports medicine clinical records (38)	Nature of injury and associated activity, other variables according to clinician's interest	Access to records problematic, highly selected sample, unknown data quality	Not suitable for all the same reasons that population health status cannot be fully determined from morbidity statistics
Facility use, equipment sales (39)	Aggregate data on activity types, frequency	Not individually based, unknown quality, sales or memberships may not accurately reflect use	Not suitable because data are aggregated but potentially useful for confirming trends, regional differences, and the like

[1] Reference example in parentheses.

occupational nor leisure. At least two large surveys (16, 17) have attempted to resolve this problem for homemakers by treating household chores as activities that can be described in terms of type, frequency, and duration.

Analytic objectives imply more detail on the nature of activity. Because this currently means activity type, frequency, intensity, and duration, many of the self-report methods are potentially suitable. Polls and self-assessments are unlikely to provide the needed detail, but this is easily determined once objectives have been adequately specified. Job classification, although potentially detailed, suffers from a different type of drawback that may limit its usefulness in analytic studies: Intraoccupation variations are not taken into account (chapter 21). Descriptive statistics can tolerate a certain amount of such random error, but analytic studies of activity-disease relationships will be rendered less conclusive by such variation.

In conclusion, it would appear that all of the self-report techniques could be suitable for descriptive surveys and most have potential for analytic studies. Motion sensing and job classification appear to be more useful for descriptive than for analytic purposes, but this will depend on precise objectives.

Considerations for general population surveys

This section deals with the measurement quality of the principal self-report techniques and makes some recommendations for use in general population surveys of health.

The literature indicates that there are not only several identifiable types of self-report measures available for assessing physical activity (table 3 lists five) but there are also variations, sometimes several, of each. Table 4 lists nine examples of self-report measures, focusing on those that are reasonably well documented and/or have been widely used. On the basis of the information summarized in table 4, it is possible neither to eliminate any of these procedures from further consideration nor to identify a single one that is best for use in all population surveys. Each has some flaws and some strengths.

Taken as a whole, the self-report measures suffer from an absence of direct evidence on their psychometric properties, especially reliability. Only two methods, the 3-day diary (20) and the 7-day recall (21), provide any evidence of repeatability. Validity, in the absence of any agreed-upon criterion measure of activity, has been only indirectly demonstrated, but the evidence is more abundant than for reliability. Many techniques show the same patterns for sociodemographic groups despite different definitions and the widely varying prevalence estimates that result (5). This suggests either that some underlying construct is common to all these techniques or that response bias is uniform across varied populations over 10 years of detailed surveying (1972–81). Presumably, the former is the case because self-reported activity is consistently related to health characteristics (some, but by no means all, being also self-reported) in the expected direction. In summary, it may be said that self-report techniques appear generally to have some validity, but no single method is clearly superior to the others. Coupled with the uninspiring evidence on repeatability, a property that is particularly important for collecting good population trend data, this strongly suggests that further developmental work is required in the area of activity assessment.

Practical considerations, principally time and re-

Table 3. Selected self-report measures of physical activity

Method [1]	Key variables	Practical considerations	Suitability for population surveys
Current status			
Polls (1) and open-ended surveys (27, 40)	Vary; tend to emphasize frequency of participation in leisure-time activities, little other detail; reference period is open	Inexpensive, timely; can be telephone, mail, or face to face; validity and reliability variable	Can be very suitable if desired detail is minimal; not generally suitable for energy expenditure; tend to reflect current situation, which may not be typical for the individual
Diary (20)	Vary; can be total energy expenditure, type, frequency, intensity, duration; can include dietary intake	Very demanding of respondent, may be reactive, reliability can be high	Unlikely to be suitable for general population because of burden involved; tend to be restricted to a short period, which may not be typical for individual
Self-assessment (41)	Comparison of own activity level with that of others of same age and sex, generally without specific time reference	Easily administered by mail, telephone, or in person; interpretation problematic because activity level is relative, not absolute; tendency to exaggerate self-assessment	May be suitable if required detail is minimal and comparisons with National Center for Health Statistics surveys dating back to 1975 are desired
Comparison with a standard	Rating of own activity as above or below a defined standard, such as behavioral equivalent of 3 kcal/kg/day	Easily administered by mail, telephone, or in person, but behavioral question difficult to frame for a range of body weights, activity types, intensity, and other uncontrolled variables	Unlikely to be suitable for general population because of uncontrolled factors, as noted
Detailed history			
Short term (16, 21)	Vary, can be total, work, or leisure energy expenditure; type, frequency, intensity, duration, etc. over last 1–4 weeks	Cost and burden depend on detail sought but are generally acceptable; validity and reliability vary and are often unknown, but probably are superior to long-term recall	Can be very suitable if subject matter is appropriate; adaptable in this regard; capable of administration by telephone, mail, or in person; considerable experience and many bases for comparison, but may not reflect habitual activity of individual
Long term (17, 42)	Vary, but similar to short-term history with reporting periods up to 12 months	Cost and burden depend on detail; recall burden and error probably higher than for short term, but seasonal fluctuations and recent changes less a factor; validity and reliability not generally established	Similar to short-term history, with data on longer period substituting for accuracy of recall

[1] Reference example in parentheses.

spondent burden, are bound to influence the selection of a survey technique, particularly when other considerations such as psychometric properties do not dictate a clear choice. The entries in table 4 suggest that time and burden are primarily a function of the detail sought by the investigator and secondarily a function of the level of accuracy demanded. Thus the lengthy procedures (10 or more minutes) are those in which details of frequency, intensity, and duration are sought for a range of activities or precision of response is demanded. Length of recall period, which ranges from 7 days to 12 months in these examples, does not seem to be strongly associated with respondent time. In contrast, the length of the activity list that respondents must consider and whether it extends to occupational activity and household chores will affect the time required for completion. Time and excessive demands on respondent memory will presumably affect quality of response.

Above all, survey objectives should dictate the choice of activity assessment technique and the nature and level of detail sought. Practical considerations cannot be ignored but should be secondary. New developments in this field are commonplace, even if documentation on them is not, and useful variations on the methods in table 4 can be developed. (Some guidelines for this eventuality are provided in the final section of this chapter.)

However, the common characteristic of all the techniques in table 4 is their use in projects of major importance. All have been employed in general population surveys, most of these on a national basis in the United States or Canada. Some have venerable histories, such as the Gallup series, beginning in 1961, or the NCHS self-assessments, dating back to 1975. Given the almost complete absence of time-trend data on physical activity (5), survey designers should give careful thought

Table 4. Properties of selected self-report measures of physical activity

Method	Approach	Validity, reliability	Time-burden	Major use
Current status				
Polls, e.g., Gallup (1)	"Aside from any work you do at home or at a job, do you do anything regularly, that is, on a daily basis, that helps keep you physically fit?"	Adequate construct validity may be inferred from distributions of activity by age, sex, income, etc., but question is susceptible to changing concepts of "physically fit"	Less than 5 minutes	Gallup polls, 1961–84
Open-ended surveys, e.g., NSPHPC[1] (40)	"How often do you—(7 activities listed)?" (Often/sometimes/rarely/never)	No evidence available.	Less than 5 minutes	NCHS,[2] 1979–80
Diaries, e.g., 3-day	Activities recorded in 15-min blocks for 2 weekdays, 1 weekend	Repeatability over 3 days ($r = .96$); significant correlations between energy expenditure and physical work capacity ($r = 0.31$) and body fat ($r = -.08$ to $-.13$) provide indirect validation (20)	288 entries required, each with a code of 1 to 9 corresponding to intensity of activity	Bouchard et al. in Quebec City
Self-assessments, e.g., NHIS[3] (41)	"Compared to others your own age, are you more/less/as active?"	Validity indirectly indicated by correlation with estimated $MVO_2 = .50$, the best of 8 questions on physical activity (43)	Less than 5 minutes	NCHS,[2] 1975, 1977, 1979, 1980, 1985, 1990
Detailed histories				
Short term, e.g., 7 day (21)	Occupational, household, and sports activities are divided into three classes of intensity for the respondent: Moderate (3–5 METS), hard (5.1–6.9 METS), very hard (7+ METS); respondents indicate the hours of participation in each class of activity during the previous 5 weekdays and previous weekend	2-week repeatability for *number* of activities higher than hours or total energy/expenditure; latter higher for women than men, for overweight than normal, for kilocalories per day than kilocalories per kilogram per day. Age-sex distributions consistent with expectations. Correlation with daily diary for activity-hours ranged from .39 to .90, with motion sensor for activity minutes from $-.38$ to .91 for 12 subjects (21)	15–20 minute interview, focusing on only 3–5 hours/day for moderate and more strenuous activities; balance is sleep and light activities, which are obtained by subtraction	Five-City Project
Short term, e.g., 14 day (16)	2 lists are presented: 14 sports and conditioning activities and 9 household chores. Respondent indicates total occasions in last 2 weeks and average duration in 4 categories (1–15/16–30/31—60/more than 60 mins) for each activity	Validity indirectly indicated by age-sex-education-occupation differences in non-work-related activity level and by associations in expected direction between activity and health status (disability days, activity limitation, emotional health, blood pressure, health services use), activity and estimated MVO_2 (16)	5–10 minutes, self-completed	Canada Health Survey
	Interviewer asks about participation during previous 14 days in up to 23 activities, then probes for frequency, intensity and duration (chapter 17, appendix F)	No evidence or data yet	Approximately 7 minutes, including supplementary knowledge questions	NHIS,[3] 1985, 1990
Long term, e.g., 12 month	Respondents indicate which leisure-time activities in a list of 62 they have done in the past year; trained interviewer probes for number of occasions in each month of previous year and average time per occasion (42)	Validity indirectly supported by significant association between activity index and performance on exercise treadmill, e.g., for 150 bpm, $r = 0.409$ (42)	25 minutes	Minnesota Heart Survey (26)
	Respondents complete questionnaire giving activity type (from reference list of 104 activities), average duration, frequency, intensity	Further indirect evidence from positive association with step-test results, interaction in expected fashion with smoking status (chapter 22, table 13)	5–15 minutes, depending on extent of activity; use of 3 reporting periods complex for respondent	Canada Fitness Survey (17) (chapter 17)

---

[1] National Survey of Personal Health Practices and Consequences.

[2] National Center for Health Statistics.

[3] National Health Interview Survey.

Table 5. Classification of self-report methods

Short-term focus		Long-term focus	
Leisure-time activity only	All activity	Leisure-time activity only	All activity
	Kilocalories		
Canada Health Survey	7-day recall 3-day Diary	Taylor LTA	
Canada Fitness Survey National Health Interview Survey 1985, 1990		Canada Fitness Survey	
	Ordinal scale		
Polls Open-ended surveys Self-assessments			

to the merit of repeating old questions, even if less than perfect. The most desirable approach for a new general population survey of physical activity may be to use an established approach (if the content is suitable) and complement it with a new and demonstrably better technique. In this fashion, trends and comparisons are made possible, new and higher quality data are obtained, and the equivalence of the two techniques is established.

The ultimate selection of a self-report method will be made much easier if answers are available for three key questions:

1. Is the activity assessment to result in a kilocalorie value or just an ordinal ranking?
2. Is all activity to be included or just leisure-time activity?
3. Are long-term patterns needed or is current activity sufficient? If it is long term, how long a reporting period is theoretically relevant?

Table 5 assigns the various self-report techniques to the eight categories that result from answering these three questions. This assignment is based on the principal uses to date; adaptations could expand the possibilities. For example, any technique that has been successfully used to measure total activity could be applied just to leisure activity. The converse is not necessarily true.

Implications for survey design

Using the distinction between descriptive and analytic objectives for the study of physical activity in a general population survey, this section discusses the implications for survey methods and content of emphasizing one or the other of these objectives. It is assumed that such a survey will cover the noninstitutionalized population by means of a representative household-based sample. Sample design, regionality, and seasonality are discussed in this section.

Age coverage

Descriptive objectives for assessing physical activity must identify the population of interest at the outset. For an assessment related to the 1990 national objectives, for example, age groups 10–17, 18–65, and 65 years and over have been identified (12, pp. 79–81). If the entire population is to be described, only ages 0–9 need be added. Restrictions in age coverage may be justified if adequate descriptive data exist from other sources, as they do for children ages 10–17 (22) and for adults ages 18 and over (1985 National Health Interview Survey supplement). By 1990, reasonably detailed descriptive data on leisure-time physical activity may be lacking in the United States only for those under 10 years of age.

Analytic objectives for the study of physical activity in any population survey should also make explicit mention of age. Restrictions may be justified for at least two reasons, assuming activity is the independent variable and health status or physical fitness the dependent. First, adolescents or the retired may be excluded in order to reduce the effect of extraneous variables on the relationship. Second, data needed to analyze a relationship may not be available for all ages. For example, children and young adults do not show enough evidence of cardiovascular disease to investigate its relationship to activity. Elderly survey participants may decline to be fitness tested (chapter 17), thereby ruling out an analysis of activity as related to fitness for this age group.

Sample selection and proxy responses

One of the important choices in designing a survey sample has to do with the final selection of the individual: Is one member of a household to participate or are all members included, at least within specified age ranges?

The consequences of sampling several household members are less serious for descriptive than for analytic studies if population estimates are to be produced using poststratified sample weights. This procedure is typical in some large data series and has no adverse impact on the reliability of population estimates.

For analytic studies, the consequences are more profound and require precision of objectives. It is generally true that better analysis will be possible by choosing only one survey participant per household. This is because strong familial influences (genetic and environmental) will reduce the variability in physical fitness, health status, and probably physical activity found among family members in contrast to unrelated individuals (23). This violation of the assumption of a simple random sample, although not uncommon, complicates analysis and should be avoided if possible.

On the other hand, if there is an explicit interest in intrafamily relationships (chapter 19), sampling several members per household may be not only acceptable but essential to the analysis. Often, however, cost considerations dictate that all or most family members are selected for survey participation, but the individual as unit of analysis is treated as if selected at random.

The problem of missing data is endemic to general population surveys and is sometimes successfully handled by acquiring proxy responses. This approach is unlikely to be adequate for data on physical activity, however, unless simple participation counts are all that is sought. The detail required to identify activity patterns or calculate energy expenditure is, by its nature, likely to be of acceptable accuracy only if collected directly from each individual respondent. Although this requirement may create some difficulties for data collection, telephone interviewing or self-administered questionnaires are possible solutions. These are discussed later.

Special subpopulations

The nature of health problems and health care delivery in contemporary Western society is such that certain groups that are proportionately small in size compared with the general population have a disproportionately large share of both risk exposure and frank problems. Thus, obtaining adequate detail on the health of, for example, black persons, Hispanics, the elderly, handicapped, and native people is a design challenge for a general population survey.

If the objective is descriptive, oversampling may be required. This is a relatively straightforward statistical matter that results from identifying the groups of interest; the detail desired for age, sex, and the dependent variable(s); the geographic or political divisions of interest; and the likely frequency of response to the relevant questions. Unfortunately, among some groups of interest, including low-income people and the elderly, positive responses to questions about leisure-time physical activity are likely to be rare (5).

Oversampling certain groups within a limited budget implies undersampling others. Taken to extremes, this can jeopardize the quality of estimates for the whole population. One creative solution, employed in national surveys (16, 17) where the population distribution between political units is highly skewed, is to sample proportional to the square root of these populations.

Analytic objective may lead to a very different conclusion than oversampling: It may be logical to exclude certain groups if it is reasonably suspected that subpopulation attributes will confound the relationship under study. This is the rationale for the numerous community intervention trials that have focused on white employed men.

Whether this is a politically acceptable approach for a national statistical agency is doubtful. Fortunately, there are three alternatives to excluding subpopulations from the survey design. One is to include them in the survey, perhaps for the purposes of descriptive statistics, but exclude them from the analysis of certain relationships. This implies a smaller number of cases for the analytic study within a fixed budget. A second alternative is to ignore the possible confounding effects on the grounds that the proportion of cases from these special populations is small. However, this may lead to Type II errors in analysis. The third possibility, and probably the most desirable, is to minimize a priori exclusions from the survey and use multivariate techniques to account for influences caused by sociocultural factors.

Region and season

For logistical or other reasons, data collection may be scheduled to avoid winter weather; this is the case with the National Health and Nutrition Examination Survey (NHANES) design (2). This approach will introduce bias to the data on physical activity, which has a distinct seasonal pattern (chapter 17, figure 3). This bias will affect both descriptive and analytic studies, but in different ways.

The impact of following the good weather would probably be to increase the estimates of the active population because total activity tends to be higher in warmer weather. However, descriptive statistics are unlikely to be influenced in a uniform fashion. Three factors—level of activity, type of activity, and region—operate to influence some people more than others.

Those likely to be least affected by seasonal factors are the sedentary and the perenially active. However, as these two groups constitute an estimated 40 percent and 20 percent of the population, respectively (5), a sizable remainder would be influenced by seasonal factors. As this moderately active group is the target of many health promotion efforts, estimating its size accurately over time is vitally important.

Most outdoor activities and many indoor ones are seasonal by custom, if not by necessity. Of the seven most popular activities—walking, swimming, calisthenics, bicycling, jogging, bowling, and softball (5)—three or four are done outdoors and are curtailed in cold weather. This fact may interact with personal attributes to complicate descriptive statistics. Gardeners, for example, are more likely to be female and older (24). Avoiding data collection during the winter may thus inflate the activity estimates of certain population groups more than others.

Finally, it should be noted that this is a problem that affects some regions of the United States more than others, and it is not a straightforward matter of temperature. In those northern regions that NHANES avoids because of potential snow, for example, winter activities can be expected to replace summer activities (although more for the young and the affluent than the older and less affluent). The greatest reduction in activities may be in those areas with enough cold to discourage summer-type activities and insufficient snow to encourage true winter sports.

Analytic studies may also be jeopardized by surveys emulating the current NHANES design, depending on precise objectives. A short-term, seasonally induced change in activity pattern, if not accompanied by a corresponding change in some health parameter, may lead to erroneous conclusions about the relationship between activity and health. This may be most pronounced for chronic conditions, but too little is known about the nature of the exercise stimulus and its relation to health states to predict the effect of short-term inactivity. Such knowledge will be more difficult to acquire, however, with a study design that is not sensitive to fluctuations in activity patterns.

This seasonal-regional factor will be a problem for any population survey that focuses on current or recent activity patterns, whether the objectives are descriptive or analytic. If a short reporting period is used, as is sometimes recommended for the sake of accurate recall (7), there are four design alternatives to be considered.

Most satisfactory for good descriptive statistics is to sample in all parts of the country during all seasons. The most prominent example of this approach is the National Health Interview Survey (25). However, this approach will not by itself provide better analytic data on chronic activity. Equally satisfactory for descriptive and analytic study aims would be to combine the current NHANES approach with a followup to produce an average annual value for the individual's energy expenditure or activity patterns.

Two other possibilities exist; neither is completely satisfactory for overcoming this problem. One is to increase the reporting period to a full year. Although this has been tried with some success (17, 26), as described in the previous section, it does increase respondent burden and suffers from inaccurate recall, especially for those with erratic exercise patterns. A final option is to maintain the short-term reporting period and ask the participant how this compares, by his own assessment, with the long term. This approach can identify recent changes and thus potential problem cases for analysis, but it is not clear how to use such data to adjust the short-term reports to more accurately reflect long-term patterns.

Implications for data collection

Data collection requirements normally flow from content specifications and operational definitions rather than directly from objectives. However, because descriptive objectives can normally be satisfied by more modest content than can analytic objectives, it is possible to compare how choice of objectives will affect time requirements, field staff qualifications, data collection site, and context.

Time requirements

A wide range of methods exists for assessing physical activity, as described earlier. Although generalizations about the time requirements of these methods are perilous, some illustrations are possible.

Two minutes are sufficient for the Centers for Disease Control Behavioral Risk Factor Surveillance approach (27) in which an individual is asked to report on his or her most frequent activity. Further detail on frequency, duration and intensity, or more activities requires more time. The 1985 National Health Interview Survey (NHIS) supplement requires an average of 7 minutes to identify type, frequency, average duration, and intensity of up to 23 leisure-time activities. (See chapter 17, appendix F.) Either of these approaches or their equivalent might be suitable for meeting descriptive objectives.

Analytic objectives require more detail and more time. A *minimum* would be 7 minutes for questions, as in the NHIS supplement. Additional questions on work-related activity and household chores could add an equal amount of time. Still other topics and procedures, such as clinical assessments of disease states or fitness testing, could consume an hour or more. (See especially chapter 4 regarding fitness testing.)

Field staff qualifications

The types of activity assessment suitable to either descriptive or analytic objectives are unlikely to pose any great demands on NHANES field staff or other experienced interviewers. The use of diary methods or motion sensors, as well as the various questionnaire and interview approaches, are all capable of being administered without great difficulty, although the importance of adequate training and field instructions should not be minimized. The one exception might be direct observational methods, a technique that seems unlikely to be judged suitable on other grounds. Far more significant is the training required to undertake any related fitness testing, a topic dealt with by Wilmore, Bar-Or, Wenger, and Smith in this report (chapters 4, 10, 13, and 14).

Data collection site

Administration at home or by telephone are the major options for assessing physical activity in most general population surveys. Mailed questionnaires are theoretically possible and sometimes effective (28) but have been rarely used for this topic. NHANES has the additional possibility of gathering data on physical activity in the mobile examination centers. Not all sites are suitable for all methods, but the self-report methods likely to be used for descriptive purposes, being brief, could be conducted at home, in mobile centers, or by telephone. Some physiological procedures require special facilities and are best done away from a group setting to avoid distraction; others are meant to assess the normal range of free movement and could not be limited to one site. Some limited physiological testing has been successfully conducted in the home as part of a general population survey (16, 17); this may have the effect of enhancing participation.

The more detailed data characteristic of analytic studies generally require face-to-face contact for collection. Extensive self-report data would be difficult, but not impossible, to obtain over the telephone. If an analytic study is the goal, with the greater participant demands that this implies, it may be useful to conduct a basic activity assessment in the home at the time of initial enumeration and supplement this with more detail later on. In this fashion, it would be possible to identify bias among those who do not undergo any subsequent examination and testing. A variation of this approach was used in one large survey, revealing that participants who declined fitness testing were more likely to be sedentary than were test subjects (chapter 17, table 6).

Context for data collection

Survey contexts may be established by the identity of the sponsor, the wording of questions, the demeanor of staff, and other such factors. There are at least three possible contexts for assessing physical activity in a general population survey—health, disease, and fitness. Each is likely to have a different and systematic impact on the data that result.

A fitness context is one that implies a positive value for fitness, and it may induce the participant to inflate reports on activity number, frequency, intensity, duration, etc. Observational and physiological measurements of activity are also reactive and may be affected by the context. A disease context, which focuses on illness, disability, and the negative consequences of activity, may produce underreporting or otherwise depress indications of physical activity. A "health" context that is nonjudgmental about the value of physical activity is most likely to have a neutral impact on the data. The earlier a physical activity assessment occurs in the sequence of data collection activities, the more likely is it to be seen in a neutral light.

The consequences of systematic errors in activity assessment that arise from the data collection context will vary depending on whether a descriptive or analytic study is being undertaken.

Systematic error will affect the estimates sought in a descriptive study. The direction of error may be discernible, but its magnitude will probably not be. This is especially serious for comparisons between surveys if their contexts vary. Cross-sectional comparisons within a data set will be less affected unless there is some interaction between reactivity of the context and demographic attributes. Different reactions by different age-sex groups to the disease and fitness contexts are easy to imagine; assessing their extent is not.

Systematic error will affect analytic studies by contributing unexplained variance to an activity-disease association; this will pose a particular problem if it is greater for certain groups. For example, in a fitness context, smokers may feel constrained to report greater activity to compensate for their smoking habit. This would confound any subsequent analysis of the association between activity and cardiorespiratory fitness.

Recommendations

Substantive and operational recommendations directed at general population surveys and NHANES, in particular, are offered here. These are necessarily general, in the absence of specific survey objectives. Indeed, the primary recommendation to all survey designers is to begin with the obvious but often overlooked step—*specify the objectives.*

Recommendations for all general population surveys

Specifying objectives in the context of the current discussion means:

- Select the most appropriate concept of physical activity—participation counts, activity patterns, or energy expenditure indexes.
- Decide whether to include occupational activity and housework.
- Consider the importance of comparability with existing data on physical activity and with the 1985 and 1990 NHIS supplements.
- Specify the levels of detail and precision required, as these have a direct bearing on procedures and associated cost, time, and burden.

Clarity of objectives on these dimensions will lead almost effortlessly to the selection of an existing assessment procedure or to the realization that modifications are required. As the procedure is likely to involve self-report, four cautions are offered:

- Define key terms. Unqualified use of terms such as "regular" and "often" should be avoided. The problems that can arise are illustrated by the results of two otherwise equivalent midwestern surveys: 48 percent of participants claimed to exercise "on a regular basis"; the proportion increased to 65 percent when this was explained as "at least once a week in the last month"(29).

- Use an activity list. In the self-report techniques based on recall, lists of activities are generally used to prompt the respondent's memory and to establish the concept of physical activity of interest to the researcher. The activity lists employed in the major surveys described in this paper range in length from 7 to 104 activities, and even this latter list is not exhaustive. However, for long recall periods, an extensive list seems essential. This is illustrated by the results of one large survey (13) in which contradictory results were reported for several activities, the discrepancy apparently a function of the inconsistent use of activity lists. For example, 13 percent claimed to have bicycled within the last month when the activity was listed, but only 1 percent volunteered that they had bicycled in the last year, when the activity was not listed but had to be written in.
- Make the list complete. Related to this issue and to the researcher's concept of physical activity is the composition of the list. Two activities that are often omitted are gardening and dancing. If the interest is total energy expenditure, however, these activities must be included in any list presented to the respondent. They ranked 5th and 11th in number of participants in the Canada Fitness Survey and were reported by 30 and 13 percent of the population ages 10 and over, respectively (17). Typically, however, these activities are not listed and are not volunteered by the respondent, even though they require more energy than many listed activities.
- Qualify "walking." Walking is another problematic activity. It consistently ranks first in population surveys (5) and probably inflates estimates of the population active in its leisure time, although it is a legitimate part of *total* activity. Walking "for exercise" is a frequent qualification but is not satisfactory: The introduction of motives to the measurement of energy expenditure can only lead to undesirable complications. Preferable would be a descriptive qualification such as "brisk walking." Another approach, if intensity of activity is determined, is to include in the active population only those who walk at moderate or high intensity.

Inherent in this discussion is the assumption that standardization of physical activity assessment techniques will increase the usefulness of data. The difficulty at this stage of development is finding techniques that are good enough to standardize. Fortunately, this qualification does not apply to the questions on sociodemographic attributes, which form part of every survey. Nevertheless, variations in content and wording continue to be found, even in major non-Federal surveys.

- Identify age, sex, marital status, education, occupation (and industry), income, race, ethnicity.
- Use the standard wording for questions on these attributes, as found in recent Federal surveys, and cover the categories listed in figure 1.

Demographic Variables Recommended for Exercise Surveys

Sex

Age
  Preferred: month, day, year of birth; confirm respondent's age
  Acceptable: month and year of birth; confirm respondent's age

Marital status
  Single (never married)
  Married (including common-law)
  Separated or divorced
  Widowed

Education
  Elementary or less
  Some high school
  Completed high school
  Some college
  Completed college or university

Occupation (and industry)
The standard four-part, open-ended question is preferred. Once coded to occupation and industry, it is subject to aggregation in a variety of ways.

1. For whom do you work?
2. What kind of business or industry are you employed in?
3. What kind of work are you doing?
4. What are your most important activities or duties at this job?

Income
Determine total family (household) pre-tax income from all sources, including wages, salaries, retirement benefits, welfare, investments, etc.

Preferred: ask respondent to estimate income to the nearest $1,000, with an optional top category of "greater than $___" (amount to be adjusted with time, but not less than $40,000 presently).

Acceptable: provide categories, with smaller increments below $25,000 and larger above.

Race and ethnicity
For surveys being conducted by or for the U.S. Government, the following format complies with official recommendations:

Race	*Ethnicity*
American Indian or Alaska Native	Hispanic origin
Asian or Pacific Islander	Not of Hispanic origin
Black	
White	

For other surveys we suggest that consistent categories be utilized, while considering the sponsor's requirements and the characteristics of the proposed study population.

Figure 1

- Identify geographic location in as much detail as confidentiality requirements permit, and record this with each case so that data can be aggregated by region in a flexible fashion.

A final consideration for any physical activity survey is the impact of seasonal changes on activity and thus on the quality of data obtained. Because leisure-time physical activity is generally greatest in the summer and least in the winter in terms of both numbers of participants and average time per session and because this variation is probably not uniform across the United States:

- Consider how seasonality might affect the data and its interpretability, and consider how alternative survey designs might alleviate this influence.

In a new survey, whether an existing technique is used in modified form or a new approach to activity assessment is developed, validity and reliability must be tested and documented. Even if a familiar method is adopted intact, its metric properties cannot be taken for granted, because extremely few procedures have been subjected to the appropriate scrutiny more than once. In the absence of an accepted standard that can serve as a basis for comparison:

- Establish the validity of survey instruments by two or more means, such as construct, convergent, and discriminant validity.
- Establish the reliability of survey instruments. Usually this will mean test-retest reliability, and the time interval chosen will have to be long enough that memory is not a factor and short enough that seasonal changes do not influence responses; 2 to 4 weeks is probably appropriate for most situations.

Recommendations for NHANES

Substantive issues

As noted in the introduction to this paper, there are many outstanding questions regarding the association of physical activity to health. Among the most significant are the dimensions (type, frequency, intensity, duration) of activity benefically associated with aspects of mental health (30), osteoporosis, diabetes, weight control, and blood pressure. Because it is the one of very few health examination surveys in the world, and the only continuing one, NHANES is uniquely situated to study these associations. Therefore, the designers of NHANES are urged to:

- Collect detailed data on leisure-time *activity patterns* and occupational activity.
- Consider both current activity and its natural history for the individual to permit distinctions between acute and chronic effects of activity.
- Collect appropriately detailed data on a wide range of health outcomes, including cardiovascular disease and those listed earlier, for cross-sectional and retrospective analyses.
- Give serious consideration to an epidemiological followup of all or part of the NHANES sample to identify (a) changes in activity patterns and health status, through reinterview and reexamination, and (b) the timing and cause of death, through use of the National Death Index.

Operational issues

In the absence of specific objectives for NHANES III, it is not possible to identify all of the operational issues that the designers of that survey will face. Certainly, all those enumerated earlier as generic problems for activity surveys will affect the NHANES design. Two deserve emphasis here.

- The core of the activity assessment should be the same as that used for the NHIS supplement in 1985. This may be supplemented with additional questions, e.g., on occupational activity and lifetime patterns, but comparability with the NHIS questions is highly desirable. (This applies only to the questions shown in appendix F of chapter 17, not the additional questions on knowledge and related topics.)
- Seasonality of physical activity patterns could be a problem for NHANES if the designers maintain the past practice of "following the sun." There are three remedies to consider: (a) Change this design, and sample in all parts of the country during all times of the year; (b) return to survey participants after 3, 6, and 9 months to establish the true annual patterns; or (c) use a recall period of 12 months.

Conclusion

Activity assessment is barely beyond its infancy, both conceptually and operationally. Nevertheless, it is clear from the extensive literature of varied disciplines that activity is both relevant to health and measurable with some validity. Improvements in activity assessment techniques are achievable and likely, and the NHANES III experience should both profit from and contribute to these developments.

References

1. The Gallup Poll: Six of ten adults exercise regularly. The Los Angeles Times Syndicate, May 1984.
2. National Center for Health Statistics: Plan and operation of the second National Health and Nutrition Examination Survey, 1976–1980. *Vital and Health Statistics,* Series 1, Number 15, July 1981.
3. Haskell WL, Montoye HJ, and Orenstein D: Physical activity and exercise to achieve health-related components of physical fitness. *Public Health Rep* 100:202–212 (1985).
4. Taylor CB, Sallis JF, and Needle R: The relationship of physical activity and exercise and mental health. *Public Health Rep* 100:195–202 (1985).

5. Stephens T, Jacobs DR Jr, and White CC: A descriptive epidemiology of leisure-time physical activity. *Public Health Rep* 100:147–158 (1985).
6. Stephens T, Craig CL, and Ferris BF: Adult physical activity in Canada: Findings from the Canada Fitness Survey I. *Can J Public Health* 77:281–285 (1986).
7. Blair SN: How to assess exercise habits and physical fitness. In *Behavioral health: a handbook of health enhancement and disease prevention,* edited by J.D. Matarazzo et al., John Wiley and Sons, New York, 1983.
8. Montoye HJ and Taylor HL: Measurement of physical activity in population studies: A review. *Hum Biol* 56:195–216 (1984).
9. LaPorte RE, Montoye HJ, and Caspersen CJ: Assessment of physical activity in epidemiologic research: Problems and prospects. *Public Health Rep* 100:131–146 (1985).
10. Last JM, ed.: *A dictionary of epidemiology.* Oxford University Press, New York, 1983.
11. Public health aspects of physical activity and exercise (special section). *Public Health Rep* 100:118–224 (1985).
12. Department of Health and Human Services: *Promoting health/preventing disease: Objectives for the Nation.* U.S. Government Printing Office, Washington, D.C., fall 1980 (a) pp.79–81.
13. Statistics Canada: *Culture statistics/recreation activities 1976.* Minister of Supply and Services, Ottawa, 1980.
14. President's Council on Physical Fitness and Sports: National Adult Physical Fitness Study. *Phys Fitness Res Digest* 4:1–27 (1974).
15. American College of Sports Medicine: The recommended quantity and quality of exercise for developing and maintaining fitness in healthy adults. *Med Sci Sports Exerc* 10:vii–ix (1978).
16. Health and Welfare Canada and Statistics Canada: *The health of Canadians: Report of the Canada Health Survey.* Minister of Supply and Services, Ottawa, 1981 (a) Chapter 3, pp.71–86.
17. Stephens T: *Fitness and lifestyle in Canada.* Canadian Fitness and Lifestyle Research Institute, Ottawa, 1983, p.32.
18. Caspersen CJ, Powell KE, and Christenson GM: Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Rep* 100:126–131 (1985).
19. Sopko G, Jacobs DR Jr, and Taylor HA: Dietary measures of physical activity. *Am J Epidemiol* 120:900–911 (1984).
20. Bouchard C, Tremblay A, Leblanc C, Lortie G, et al.: A method to assess energy expenditure in children and adults. *Am J Clin Nutr* 37:461–467 (1983).
21. Sallis JF, Haskell WL, Wood PD, et al.: Physical activity assessment methodology in the Five City Project. *Am J Epidemiol* 121:91–106 (1985).
22. Ross JG and Gilbert GG: The national children and youth fitness study: A summary of findings. *Phys Educ Recr Dance,* 45–49 (January 1985).
23. Bouchard C and Malina RM: Genetics of physiological fitness and motor performance. *Exerc Sports Sci Rev II:* 306–339 (1983).
24. Canada Fitness Survey: Gardening—growing for fitness. HIGHLIGHTS 31 (July 1984).
25. National Center for Health Statistics. Health Interview Survey procedure 1971–1974. *Vital and Health Statistics.* Series 1, Number 11, 1975.
26. Folsom AR, Caspersen CJ, Taylor HL, et al.: Leisure-time physical activity and its relationship to coronary risk factors in a population-based sample: The Minnesota Heart Survey. *Am J Epidemiol* 121:570–579 (1985).
27. Bradstock MK et al.: Behavioral Risk Factor Surveillance, 1981–1983. *CDC Surveillance Summaries* 33:155–455 (1984).
28. Siemiatyki J, Campbell S, Richardson L, et al.: Quality of response in different population groups in mail and telephone surveys: *Am J Epidemiol* 120:302–314 (1984).
29. Jacobs DR Jr: Personal communication, 1984.
30. Stephens T: Physical activity and mental health in the United States and Canada: Evidence from four population surveys. *Prev Med* (in press).
31. Orsini D and Passmore R: The energy expended carrying loads up and down stairs: Experiments using the Kofranyi-Michaelis calorimeter. *J Physiol* 115:95–100 (1951).
32. Passmore R: The regulation of body weight in man. *Proc Nutr Soc* 30:122–127 (1971).
33. Shephard RJ. Normal levels of activity in Canadian city dwellers. *Can Med Assoc J* 97:313–318 (1967).
34. LaPorte RE, Kuller LH, Kupfer DJ, et al.: An objective measure of physical activity for epidemiologic research. *Am J Epidemiol* 109:158–168 (1979).
35. Schoeller DA and van Santen E: Measurement of energy expenditure in humans by doubly-labeled water method. *J Appl Physiol* 53:955–959 (1982).
36. Ary D and Suen HK: The use of momentary time sampling to assess both frequency and duration of behavior. *J Behav Assess* 5:143–150 (1983).
37. Taylor FW: *Principles of scientific management.* Harper and Bros., New York, 1911.
38. Mueller FO and Blyth CS: Annual survey of catastrophic football injuries: 1977 to 1983. *Physician and Spts Med* 13(3):75–81, 1985.
39. Statistics Canada: Recreation equipment ownership, 1974–1982. *Travel-log* 1(4), 1982.
40. National Center for Health Statistics: Highlights from Wave I of the National Survey of Personal Health Practices and Consequences: United States, 1979. *Vital and Health Statistics.* Series 15, No. 1, June 1981.
41. National Center for Health Statistics: Health practices among adults: United States, 1977. *Vital and Health Statistics, Advance Data* 64 (November 1980).
42. Taylor HL, Jacobs DR Jr, Schucker B, et al.: A questionnaire for the assessment of leisure-time activities. *J Chron Dis* 31:741–755 (1978).
43. Thacker JD: The validation of selected physical activity questions using the Canadian Home Fitness Test (CHFT). *Can J Pub Health* 72:455–458 (1981).

Part IV

Special Subpopulation Issues

Chapter 10

Fitness and Activity Assessment of Children and Adolescents

Oded Bar-Or, M.D.
CHEDOKE-McMaster Hospitals, Hamilton, Ontario

Introduction

This paper is meant to present design issues and alternatives related to assessing physical fitness of children and adolescents within the NHANES protocol. It is assumed that the definition of physical fitness; the relevance of fitness to health; and the design, logistical considerations, and data analysis in a fitness survey are only marginally dependent upon the age of the subjects. This paper will therefore focus on *special considerations* in the testing of children and adolescents, rather than discuss the general issues of fitness surveys. The latter have been presented in chapter 4 and in other chapters.

Age range

Adolescence is commonly defined as the period of life between puberty and adulthood. There is no definite stage at which a child becomes an adolescent. Some consider the period during which pubertal changes take place as part of childhood, and others consider it the start of adolescence. In this document, we shall not address separately the fitness issues of childhood and of adolescence, but rather look at the preadulthood years as one continuum. For the sake of brevity we shall use "children" to represent this whole continuum. A distinction between children, prepubescents, pubescents, and adolescents will be made only when specifically warranted.

Only little information is available on physical fitness and its assessment in children younger than 6 years. Most studies are related to the school-age child. This paper will therefore limit its scope to school-age or, broadly speaking, to the age range 6–18 years.

Health of subjects

This paper is the only one within this NHANES document that suggests protocols for the child population. Thus, although focusing on subjects with no apparent disease, it includes also some considerations on the testing of children with a disease or an illness.

Fitness in the health context

The fitness characteristics needed by an elite young gymnast are quite different from those needed by a recreational swimmer and further different from those expected from a wheelchair-bound child with spina bifida. Before attempting to construct a testing battery that reflects the physical fitness of a population, one must therefore answer the questions: Fitness for what? and Whose fitness?. In keeping with the health orientation of the NHANES, the fitness components selected in this paper are meant to reflect the health, rather than athletic prowess, of the general child and adolescent U.S. population.

Physical activity in the health context

Spontaneous physical activity (hence, activity) seems to be age- and maturation-dependent. Children are habitually more active than adolescents who, in turn, are more active than adults (1–3). Although sedentary adults are not considered a priori ill, a sedentary lifestyle in children often connotes ill health. Such children are often physically or emotionally handicapped, or they experience social isolation and maladjustment. Moreover, a child who is hypoactive (i.e., is less active than most children in his society) often enters a vicious circle in which hypoactivity induces detraining, reduced physical function, loss of self-confidence, increased social isolation, and further hypoactivity.

Even though definitive data are not available on "tracking" of activity patterns throughout adult years, one may assume that a hypoactive child will likely turn into a hypoactive adult. Conversely, an individual who during the formative years acquires certain sports skills and activity habits is more likely to remain active in later years.

It is therefore of the utmost importance to include in any health-oriented fitness survey of children a measure of their habitual activity. One should attempt to describe the time devoted to exercise and sports, the intensity of the activities, the type of activities that are liked/disliked by the child, attitudes toward exercise and sports as displayed by the child and by other family members, and any logistical or other factors that may influence the activity of the child.

Testing of physical fitness

Characteristics of children and adolescents with implications for fitness testing

When planning a testing battery for children and adolescents, one should not assume that they are merely scaled-down adults. There are morphologic, physiologic, and behavioral growth- and maturation-related differences that must be taken into consideration. These are listed in table 1.

Morphologic characteristics

Small body size

Although children's body proportions differ from those of adults, it is primarily the difference in body size that dictates the need for equipment modification. Little information is available on the optimal dimensions of cycle ergometers for people who differ in body size (4, 5). It is commonly agreed, however, that adult-size ergometers can be used, without modification, for children whose body height is 130 cm or more. For smaller children, modification is needed in seat height, pedal-shaft length, and the distance between the seat and the handlebar. Klimt and Voigt (5) have found shaft lengths of 13 and 15 cm to be most efficient (i.e., causing the lowest O_2 uptake) for 6- and 8-year-old children, respectively.

Table 1. Morphologic, physiologic, and behavioral characteristics of children and adolescents with implication to testing of physical fitness

Characteristic	Implication to testing
Morphologic	
Small body size	Lower ergometer seat, shorten pedal shaft
Small knee extensors	Use cycle instead of treadmill in preschoolers
Large surface area per mass	Strict climatic control
Low density of lean mass	Modify density-to-fat equation
Physiologic	
High O_2 cost of locomotion	Modify prediction equations for treadmill
Rare leveling off of $\dot{V}O_2$	Modify criteria for $\dot{V}O_2$max
Short metabolic transients(?)	Shorten stage duration in a multistage test
Low sweating capacity	Keep tests short and a strict climate control
Behavioral	
Short attention span	Keep tests short; avoid distraction
Undeveloped eye-leg coordination	Modify pacing metronomes; use treadmill in young children
Lower intrinsic motivation	Seek extrinsic motivations

Figure 1 is a schema of a cycle ergometer, used in the author's laboratory, that has been modified for use with children and adults of various body sizes. The seat may be raised or lowered as well as moved forward and backward, to allow for a proper knee-ankle-pedal relationship. Both horizontal and vertical seat extension bars are marked with gradations of 2 cm, to allow for replication of seat position. Similar markings are available for the handlebar, which can be moved forward and backward. Pedal-shaft length can be set at one of three positions (9.5, 13.5, 17.5 cm). These have been selected to suit leg or arm lengths of the entire preschooler-to-adult range. In this particular ergometer (Fleisch-Metabo, Switzerland), resistance increments can be as small as 60 g.

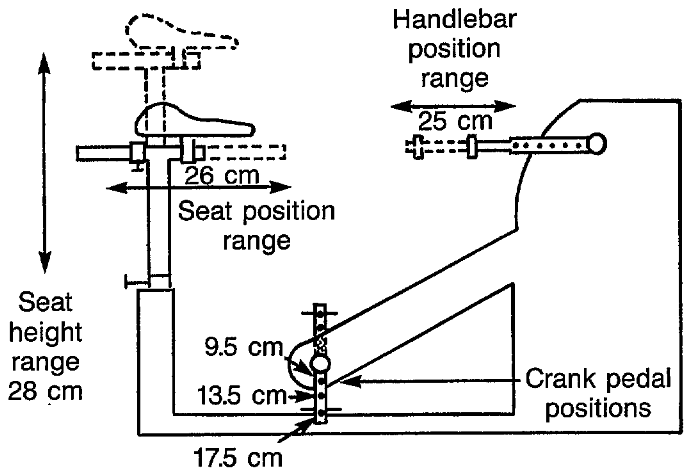


Figure 1. Modifications of a cycle ergometer for use with children

A commercially made Von Dobeln-type pediatric ergometer is available (Monark, Sweden). Its seat is small and it can be lowered more than in an adult-type ergometer. Pedal-shaft length can be set at 12, 14, and 16 cm, and the handlebar position can be somewhat modified. The resistance pendulum is lighter than in an adult-size Monark ergometer, such that a resolution of 250 g can be achieved.

Small thigh muscle mass

The above modifications allow young children to use a cycle ergometer with relative comfort, which is adequate for submaximal tasks. The ability of these children, however, to reach maximal aerobic power on the cycle ergometer seems quite limited (6). We often encounter a 6-year-old child who cannot exceed heart rates of 160 beat/min during all-out cycling, compared with 195–210 beat/min during a treadmill run. Although the cause for this difference is not clear, one can assume that it reflects, to some extent, the small mass of children's knee extensors, which take the brunt of load during cycling.

Large body surface area per body mass

Children's body surface area, relative to body mass, is greater than in adults. It is, for example, 380 cm^2/kg in a child who weighs 25 kg and is 128-cm tall and only 280 cm^2/kg in a 177-cm adult who weighs 64 kg. Such a large surface area allows for a faster heat exchange between skin and environment. The implication to the investigator is that ambient temperature should be strictly controlled (a 2° dry bulb temperature range is reasonable) when children are exercise tested.

Low density of lean body mass

In children, the density of such lean mass components as protein and bone is lower than in the "chemically mature" person (7, 8). This invalidates the use of equations, originally derived for adults, for calculating %fat from total body density. Correction factors for such equations have been suggested. These are discussed in the section on body composition.

Physiologic characteristics

High O_2 cost of locomotion

When walking or running at a given velocity, children consume more O_2 per kg body mass than do adults. Such a difference is greater the younger, or smaller, the child (9–12). A 6-year-old child, for example, requires some 8–10 ml/kg/min more oxygen than a 17-year-old adolescent, both running at the same submaximal speed. The cause of such a difference is not fully understood. It may represent a biomechanically more "wasteful" gait style in the child that can be improved with training (13). It may also reflect a lower degree of absorption of kinetic energy in the elastic components of the leg muscles due to the child's small body mass (14). The implication to exercise testing is that the metabolic cost of locomotion during any given treadmill speed and gradient or during a walk or run test "in the field" cannot be assumed based on equations derived for adults. Nor can prediction of maximal aerobic power be made based on these equations. An example is the Bruce treadmill test (10, 15), in which $\dot{V}O_2max$ is assumed from the final velocity and gradient completed by the subject. Recently, age-related equations were calculated for the prediction of $\dot{V}O_2max$ from performance in the 12-min run test (11).

Inconsistent leveling-off of O_2 uptake

In adults, leveling off of $\dot{V}O_2$ during a progressive exercise test is one of the criteria for assumption of $\dot{V}O_2max$. Although some children achieve a plateau in $\dot{V}O_2$, others may become exhausted before such leveling off has taken place (16–18). The cause for such a difference is not clear. It could result from the low ability of children to utilize anaerobic glycolytic pathways. To increase the likelihood of reaching $\dot{V}O_2max$, it is recommended that a progressive exercise test not be terminated as long as the child can still exert.

Short metabolic transients

Some reports (19) but not others (20) suggest that metabolic transients at the start of exercise or on a rise in power load are shorter in children than in adults. The implication is that to assess submaximal performance of children one can keep short the stages of a progressive test. More research is indicated, however, to define the time period needed by different age groups to reach a metabolic steady state.

Low sweating capacity

The sweating rate of prepubescents rarely exceeds 350 $ml/m^2/hr$, compared with 500 $ml/m^2/hr$ or more in adults (21–23). Furthermore, the threshold for sweating (i.e., the increase in core temperature at which sweating starts during exercise) seems higher in children (21, 24). There are only few data on the changes in core temperature of children during prolonged exercise (25). Until more information is obtained, we recommend that to prevent any excessive rise in body core temperature, one should attempt to keep short (10–12 min) the total duration of an aerobic exercise protocol. The low evaporative capacity of children is another reason for a strict standardization of ambient conditions in the exercise laboratory.

Undeveloped eye-leg coordination

Although a sense of rhythm is highly individual, young or mentally retarded children often find it hard to pace their cycling cadence according to a metronome. This is particularly apparent when the mental age is less than 6 years. If one uses an ergometer in which power is pedaling-rate dependent, one may find inadequate the commonly used musicians' metronome. To the best of this author's knowledge, this topic has not been studied systematically.

A possible improvement is the use of a pacing system with feedback capability, in which the subject can continuously compare his rate with the required one. The use of two concentric light circles, one serving as a reference and the other activated by the pedals, has been found useful with some young children. This system is available in the Swiss-made Fleisch-Metabo mechanical ergometer. We are planning to develop an alternative feedback pacing system, using the technology of computer games.

An alternative approach is the use of an electromagnetic cycle ergometer with a mode in which power output is rate independent. This, however, will have to be preceded by a study that will assess the degree to which metabolic cost in young children is dependent on pedaling rate. Another approach is the use of a treadmill instead of a cycle ergometer for young children. This is the recommended solution for NHANES.

Behavioral characteristics

Short attention span

Young children have a shorter attention span than adults (26). We have found that those with a mental age of less than 5 years seldom concentrate on an exercise task for more than 2–3 min. They have to be reminded repeatedly about the task at hand. This is another reason why test protocols for young children must be kept as short as possible. It is also important to keep distractions—visual or auditory—to the minimum in the exercise laboratory.

Low intrinsic motivation

In general, children are motivated to perform well in fitness tests. Some, however, particularly the young or those with a motor or a cognition disability, do not have this intrinsic motivation. In a study performed in the author's laboratory at the Wingate Institute, various motivational factors were compared regarding their effect on performance in the Wingate Anaerobic Test (27). Subjects were healthy adolescents with reasonable motivation to perform well. This author is not aware of studies that have assessed the effect of motivation in preschoolers or children whose a priori motivation for exercise is low. One pragmatic approach to increasing the cooperation and motivation of such children is the use of extrinsic means of motivation. Our technicians have been successful in using stickers or similar prizes that the young child wins on completion of each minute of the test. We often find it useful to suggest to a young boy that he should perform like a Wayne Gretzky, an astronaut, or a famous baseball player.

Fitness components

Maximal aerobic power

Protocols for the testing of maximal aerobic power of children include either a direct measurement of $\dot{V}O_2max$ or its indirect assessment. The latter can be either submaximal or maximal, performed in the field or in the laboratory. Some specific protocols, with selected references, are listed in table 2. The rationale for selecting a test protocol from the above options is similar for children and adults, but some specific morphologic, physiologic, and behavioral characteristics must be taken into consideration.

Table 2. Test protocols for assessing maximal aerobic power in children

Type of test	Protocol examples	Selected references
	Indirect	
Submaximal, cycle	W_{170} Astrand/Ryhming nomogram	Rutenfranz (36) Hermansen and Oseid (28)
All-out field run/walk	12-min run/walk 1-mile run	Roche (45) Krahenbuhl et al. (42)
All-out, cycle/tread.	Bruce, treadmill McMaster, cycle	Cumming et al. (10) Bar-Or (39)
Step test	Canada Home	Jetté et al. (38)
	Direct	
All-out, treadmill	Progressive continuous	Skinner et al. (12) Cunningham et al. (17)
All-out, cycle	Progressive continuous	Andersen et al. (55) Woynarowska (33)

Indirect assessment: Submaximal cycle tests

An indirect assessment of $\dot{V}O_2max$ is based on either submaximal heart rate, postexercise heart rate, peak mechanical power, or running performance in a field test. Various studies with children have shown that submaximal, indirect estimates have low validity for assessing *individual* values of $\dot{V}O_2max$ (28–33). The popular Astrand-Ryhming nomogram yields results that, in children, underestimate $\dot{V}O_2max$ by 10–25% (18, 28, 31, 33). Indirect submaximal tests are useful, however, for group estimates (34–37).

Because of their low validity, such submaximal laboratory protocols as the W_{170} or the Astrand-Ryhming nomogram should not be considered for the NHANES. Nor can such step tests as the Canada Home Fitness Test (38) or the Hanne Progressive Step Test (39) be considered, because of the inherently low standardization of postexercise heart rate monitoring and the interindividual variability in economy of stepping.

Indirect assessment: All-out field tests

Numerous authors have used run or runwalk field tests to assess maximal aerobic power. Distances included 600 yards (40), 800 m (41), a mile (42, 43), and 2,400 m (41). In other protocols, the distance covered during a 9-min (44) or a 12-min run (11, 45) was measured. Such field tests are useful for assessing, within a short time, performance of large groups of children.

Some authors have validated these field tests against direct measurement of $\dot{V}O_2max$ and found correlation coefficients ranging from 0.26 (42) to 0.92 (43). Factors that increase the variance of performance in such tests include improper pacing, inefficient running style, insufficient motivation, boredom, and changes in ambient temperature, humidity, and wind velocity. It has been shown (45a) that the correlation between long-distance-running performance and $\dot{V}O_2$max is higher among fit individuals than among unfit ones. Because of the inherently low standardization of such field tests, they are not recommended for inclusion in the NHANES.

Indirect assessment: All-out treadmill tests

Various all-out laboratory tests have been suggested for children to indirectly assess their $\dot{V}O_2$max. Treadmill protocols are available in which the speed and/or gradient is progressively increased. Examples are the Bruce test (10) and the modified Balke test (46) in which the total time to exhaustion is the criterion. In the Bruce test, correlation between the total time and the directly measured $\dot{V}O_2$max was 0.88 for girls (n = 53) and 0.85 for boys (n = 24) (10). Interpretation of individual scores in any indirect treadmill test must, however, take into account the age (and, possibly, the body size) of each child. This is because of the age-related differences in running economy, as discussed in the section above, high O_2 cost of locomotion. Correction equations are not yet available to account for these differences.

Indirect assessment: All-out cycle ergometer tests

The advantage of cycle tests over the indirect treadmill ones is that mechanical power output, submaximal or maximal, is readily measured. Furthermore, performance on the cycle ergometer is only marginally dependent on interindividual differences in mechanical economy, skill, or style. On the other hand, walking or running uses a larger muscle mass than cycling, which is an important asset, particularly for the young or small child (see section on body size).

Most all-out cycle ergometer protocols for children are "progressive" (i.e., with a gradual increase in load). Some are intermittent and others continuous. Performance criteria include peak mechanical power (37, 39, 47, 48), total work (49, 50), or time to exhaustion (50).

One major difference among these protocols is the duration of each stage. These range from an almost continuous increase in load to 6 min per stage. The latter are usually done with rest periods between stages. It is the contention of this author that, because the total time of a test for maximal aerobic power should not exceed 10–12 min, each stage must last not more than 2 min. An exception is a test that is meant to also yield data on *submaximal* functions, thus requiring the attainment of a steady state at each stage. There is insufficient information on the time needed for attaining a steady state in children of various ages. One can assume, however, that stages for such a test must last at least 3–4 min.

Peak mechanical power varies markedly as a function of body size (a 16-year-old boy may reach an output that is two- to threefold that of a 6-year-old one). To construct a test that lasts 8–12 min, the *initial load* as well as the *increments* must therefore be determined according to body height or mass. Table 3 is an example for a protocol that takes the gender and body height into account. Using this protocol (the McMaster All-Out Progressive Continuous Test), we usually complete the testing of healthy children and adolescents within 8–11 min.

Table 3. An all-out cycling protocol (the McMaster Test), by body height and gender

Body height (cm)	Gender	Initial load (Watt)	Load increments (Watt)
≤119.9	Female	12.5	12.5
≤119.9	Male	12.5	12.5
120–139.9	Female	12.5	25.0
120–139.9	Male	12.5	25.0
140–159.9	Female	25.0	25.0
140–159.9	Male	25.0	25.0
≥160	Female	25.0	25.0
≥160	Male	25.0	50.0

In the above tests, peak power is defined as that power which the child has sustained for the prescribed duration of the final stage. If that stage is terminated prematurely, a credit is given on a prorated basis.

Direct measurement of $\dot{V}O_2$max: Treadmill tests

As in the indirect tests, protocols are progressive, with stages lasting between 1 and 4 min. Even though both speed and slope can be increased, many authors recommend keeping the speed constant while gradually increasing the slope (12, 31, 43, 51). As shown by Skinner et al. (12), the same $\dot{V}O_2$max is achieved by 2-min continuous, 3-min continuous, and 4-min intermittent walking protocols. In that study, the average total duration of these protocols was 8.3 min, 10.1 min, and 23.5 min, respectively, in 6- to 15-year-old boys (52). For the sake of expediency, and bearing in mind the short attention span of young children, we have adopted the 2-min protocol. As shown for adults, running yields a higher $\dot{V}O_2$max than does walking (53, 54). Direct measurement of $\dot{V}O_2$max using a running test is therefore a procedure highly suitable for NHANES.

Direct measurement of $\dot{V}O_2$max: Cycle ergometer tests

These have been popular in many laboratories, particularly in Europe (33, 55–58). The protocols are similar to those described in the section on indirect assessment. Although simpler to administer than the treadmill tests (in the latter, one investigator is committed throughout the test to safeguarding the child), cycling yields a lower $\dot{V}O_2$max than running or walking does, especially in the young child (12, 30, 59). Though no systematic comparison is available to this author, preschoolers seem to master the skill of treadmill walking or running better than that of cycling at an even pace (60).

Recommendation for a test of maximal aerobic power

It is recommended that $\dot{V}O_2max$ be directly measured during a treadmill run. An all-out progressive continuous protocol can be used in which each stage lasts 2 min. An alternative method is a direct measurement of $\dot{V}O_2max$ during cycle ergometry. The latter is logistically simpler, and it requires fewer personnel to administer, but treadmill running will likely yield higher $\dot{V}O_2$ than will cycling.

Peak anaerobic power and muscle endurance

No studies are available to this author that have compared the *pattern* of spontaneous activities in children, adolescents, and adults. It has been a common observation, however, that a child often adopts a spurt-like pattern of exertion rather than a continuous, prolonged one. The ability to perform short-term exercise of high intensity is dependent on local muscle performance rather than on the 0_2 transport system. The former includes strength, peak anaerobic power and muscle endurance.

Within fitness surveys, field tests of muscle function have included such items as the standing long jump or the vertical jump-and-reach test that assess muscle power and others (situps, pushups, flexed arm hang) that assess muscle endurance (40, 61, 62). In contrast, laboratory-based tests of muscle function have traditionally focused on muscle strength but ignored the measurement of muscle power and endurance. The main reason for such an omission is that, until recently, there have been few laboratory tests suitable for the measurement of muscle power and, especially, muscle endurance in children.

With the recent increase in interest in these fitness characteristics equipment and laboratory-based protocols have become available for such measurements. It is strongly recommended that NHANES include a laboratory test of muscle power and muscle endurance. There are three laboratory tests that assess both muscle power and endurance and one that assesses muscle power only.

Isokinetic muscle test

This test comprises 50 repeated "all-out" contractions of the knee extensors for about 60 sec, at a predetermined angular velocity (63). One of the very first contractions yields the peak power of the respective muscle group, and the total work (or the power drop) reflects muscle endurance. Three systems are available commercially that can be used for these measurements: the Cybex II (Lumex, New York), Ariel 4000 (Ariel Dynamics, Amherst), and Kin Com (Med-Ex Diagnostics, Vancouver). Among these, the Cybex system has been used extensively in research, mostly with adults, but also with adolescents (64), such that protocols are available for various muscle groups. Both the Ariel 4000 and the Kin Com have a more sophisticated and flexible software than does the Cybex II, but their reliability and validity, as well as accepted protocols, have yet to pass the scrutiny of research.

Constant velocity cycle test

Some 4 years ago a cycle ergometer was described (65) in which pedaling speed was kept constant by a motor that drove the pedal crank. Strain gauges at the pedal shaft enable the continuous measurement of torque. This ergometer was further developed at the McMaster University (66, 67) such that the motor, rather than drive the pedals at a constant speed, presets the *upper limit* of speed that the subject can reach. These ergometers have enabled the studying of isokinetic torque-velocity and power-velocity relationships during an "all-out" short-term cycling test (20–30 sec).

Sargeant (68) has recently studied, in 13- to 14-year-old girls and boys, the peak power, muscle endurance, and optimal angular velocity for reaching peak power. The suitability of this "force bike" for younger children has yet to be determined. Although they yield important information on the contractile characteristics during "anaerobic" cycling, these ergometers are still in the prototype stage. They cannot therefore be recommended for use in a large-scale national survey.

Wingate anaerobic test

This test has been developed in the last decade at the Wingate Institute, Israel. It consists of pedaling or arm cranking for 30 sec at "all-out" speed, against a constant resistance. Peak and mean mechanical power are determined; the latter is taken to reflect muscle endurance (69). One can use both a mechanical or an electric ergometer that has the rate-dependent mode. This author has had much experience with the Fleisch-Metabo mechanical ergometer, but the Von Dobeln type (Monark) can also be used. The former is readily adjustable for use with arms or legs and maintains a constant resistance better than the Monark. A schema for adapting it for pediatric use is described above (see figure 1).

The Wingate test is highly reliable when used with young adults and healthy or sick children (yielding test-retest r of 0.95–0.98) (70–72). Performance of children is unaffected by high air temperature or humidity (73), but it is improved following an intermittent warm-up period (74). The effect of various motivational means on the performance of adolescents has been described (27). The Wingate test is valid as a measure of anaerobic characteristics (4, 45, 69, 75–77). Optimal resistance that yields the highest power has been determined for young adults (71, 78) and for 13- to 14-year-old girls and boys (71), and the optimal length of pedal shaft has been determined for people who vary in their leg length (4).

Since its development, this test has been used in various laboratories to determine fitness of specific groups (79, 80) or for research related to short-term, supramaximal exercise (81–85).

Allowing for a warmup and a cool-down period, the test lasts 10–12 min. If warming up and cooling down

Table 4. Comparison of a Cybex II Test and the Wingate Anaerobic Test for measuring muscle power and endurance in children

Characteristic	Cybex II	Wingate
Suitable for upper and lower limbs	Yes	Yes
Number of muscle groups in each test	One	Multiple
Measuring peak power	Yes	Yes
Measuring muscle endurance	Yes	Yes
Measuring contractile characteristics	Yes	No
Suitable for young children	Undetermined	Yes
Cost of equipment, U.S. dollars	\$25,000–\$30,000	\$4,000–\$5,000

are done off-line, some 15–20 children can be tested within 1 hour. If logistically indicated, the arm and the leg tests can be performed in one session, some 30–40 min apart. If both aerobic and anaerobic testing is to done in one session, the anaerobic test must be done first, followed by a 30–40 min rest period.

Margaria step running test

This test, originally introduced in Italy (86), measures the peak mechanical power that is produced during running up stairs at maximal speed. The measured variable is the time needed to perform the task. Knowing the vertical elevation and the body weight, one can calculate the peak power. This test has been used successfully with children of different ages (87, 88) although its reliability in these subjects has not been reported. Although inexpensive and simple to perform, the Margaria test does not measure local muscle endurance.

Recommended anaerobic test

The force-bike test, though yielding much information, is not recommended for NHANES because only prototype apparatus is currently available. Among the isokinetic tests, the ones using the Ariel 4000 and the Kin Com will require extensive pilot studies to determine their suitability for a large-scale survey, particularly with the pediatric age group.

Table 4 summarizes the characteristics of the Cybex-based test with those of the Wingate test. The Cybex II can be used in NHANES with adults and adolescents, but its technical specifications may not allow for measuring small children. This issue will require some pilot work. In contrast, the Wingate test has been used extensively with children, healthy or sick, of the whole school-age spectrum. Another limitation of the Cybex system is that only one muscle group of one limb is measured in each test. This gives this test an edge over the Wingate test for studying contractile characteristics of muscles. It may, however, be less valid as a test that reflects *performance* of the whole limb (bilaterally). Within a health-oriented national survey, such performance is of more relevance than the characteristics of a single-muscle group. Because only one muscle group is tested at each Cybex test, one must take special precaution to isolate all other muscle groups that otherwise may contribute to the motion. Such a precaution is less critical in the Wingate test, although positioning of the seat and pedals is important. The cost of a Cybex II system is some \$25,000–\$30,000, compared with \$4,000–\$5,000 for the Wingate test. The latter includes a Fleisch ergometer, modified for use with children and a microcomputer for counting of pedal revolutions and online analysis of peak power and muscle endurance.

Based on the above considerations, it is recommended that, for pediatric populations, the Wingate Anaerobic Test be used in NHANES to assess muscle power and endurance of the arms and the legs.

Muscle strength

In the growing individual, muscle strength seems to depend primarily on body size and age (89, 90). Although strength per se seldom reflects a child's state of health, it nevertheless may be relevant to present or future health. Weak muscles in a child with physical disability are often the limiting factor in the ability to ambulate or perform other daily activities. It is likely, although not demonstrated, that, even in a healthy child, muscle weakness may dictate a sedentary lifestyle. This, in turn, may affect this individual's present or future health. It is therefore recommended that measurement of the strength of various muscle groups in children be included in NHANES.

Muscle strength, which is the maximal force or torque that the muscle can exert voluntarily, has been measured during isometric, concentric, and eccentric contraction. The latter two can be measured while the muscle contracts against a constant external force (isotonic contraction) or at a constant angular velocity (isokinetic contraction). All the above procedures yield information on the maximal voluntary contraction.

It is also possible to measure the forces that are generated following an electrical stimulation of the muscle or its nerve. Such a method, which requires highly specialized equipment and strict safety measures, has been used to analyze contractile characteristics in healthy and sick children (91–93), but it is not suitable for determining the maximal *voluntary* contraction.

Static strength

Various systems are available for the testing of maximal isometric contraction of children. These include a spring-based device that has been used extensively for hand grip (58), but also for other muscle groups; a cable tensiometer (94–96), primarily used with the Clarke table to assess limb and trunk strength; and the more sophisticated strain gauge dynamometer (90, 97, 98). Two instruments have been developed for a quantitative clinical assessment of the static strength of young patients. In both, the observer exerts by hand through the apparatus an increasing force against the patient's limb. The highest force that the patient can resist is defined as his static strength. One such device is the modified sphygmomanometer (99) and the other is a hand-held myometer (93, 100).

In isometric testing, one must standardize the joint angle at which the contraction is performed, as well as the position of the limb. Reliability and intercorrelations of scores have been determined. The strain gauge-based systems yield the highest test-retest correlations (r = 0.91–0.96 in 48 7–17-year-old girls and boys) (97). In the general population there is a fair correlation between static strength of one muscle group and that of other groups (97). Particularly high correlations (r = 0.80–0.83) have been found between identical muscles of right and left. Also related are the agonists and antagonists and the muscles of the same limb (97).

Dynamic strength

Children are seldom required to exert static force during daily activities or in most athletic events. Dynamic strength is therefore more relevant to their physical performance. There are two main approaches to testing dynamic strength: the pulling or pushing of weights and contraction against hydraulically controlled resistance. The former is used for isotonic testing and the latter for either isotonic or isokinetic testing.

With the advent of the Cybex apparatus and, more recently, the Ariel 4000 and the Kin Com (for more details see the section isokinetic muscle test), testing of isokinetic strength has gained popularity. It is preferable to the weights principle, because maximal force can be exerted in each try, without the need to search first for the weights that will yield maximal force. By continuously recording the torque during an isokinetic test, one can identify the joint angle at which the maximal torque has been reached. Most importantly for children, the isokinetic machines have a built-in safety factor because the child, rather than the machine, determines the force exerted. Using the Cybex system, test-retest reliability among 100 7–15-year-old healthy girls and boys was high. Maximal score deviation on three trials was 0–8% for various muscle groups. Mean deviation ranged between 5% and 6% (101).

Recommended strength test

Dynamic strength is more relevant to the daily functioning and physical performance of children than is static strength. Thus, within a health-oriented fitness survey, one should prefer to test dynamic strength. Among the available methods, the isokinetic approach is recommended. Even though no definitive comparison is available, this method is most likely safer for children than is a method based on lifting weights. It requires less trials to ascertain the maximal force exerted by any given muscle group, and it can yield basic information regarding force-velocity and force-length characteristics.

Muscle groups to be tested should include the upper limb (e.g., elbow flexors and elbow extensors), the lower limb (e.g., knee extensors and knee flexors), and, if technically feasible, a measurement of trunk strength. Testing can be done unilaterally (e.g., the dominant side or the right/left side).

If the isokinetic principle is adopted for NHANES, pilot studies will be needed to determine the optimal angular velocity to be used at each age group for the various muscle groups and the apparatus that is best suited for children (particularly the minimal torque required to activate the resistance and the resolution at low torque ranges). Additional evaluation is needed of protocols to measure trunk strength.

Flexibility

Joint flexibility has been included within some fitness surveys of children and adolescents (40, 102–104). It has also been considered for health-oriented fitness testing of adults (Wilmore, chapter 4). Children are inherently more flexible than adults. Although range of motion and joint flexibility may be limiting factors in the daily activities of the aged, they are less relevent to the general child population. A main drawback of the measurement of flexibility is its low repeatability. Stretching exercises and limb manipulations, which diminish the stretch reflex just prior to testing, can *markedly* modify the score. This author therefore does not recommend including flexibility in the testing of children and adolescents within NHANES.

Body composition

Although body composition is not a performance item in itself, it is an important determinant of a child's fitness. Furthermore, body composition is highly relevant to health, both when the child is undernourished or overnourished. Assessment of body composition is therefore part and parcel of any health-oriented fitness survey.

At least two other chapters in this publication (Blair et al., chapter 3; Wilmore, chapter 4) discuss the methodology of body composition assessment. This section will therefore focus on factors that merit special attention when planning such measurement in children and adolescents. Two recent reviews (105, 106) have served this author in summarizing these factors.

Available methods

To assess fat mass and fat-free mass of children, many authors have been using equations derived from data on adults. The use of such equations is based on the *assumption of constancy,* throughout growth, in the water and solids content of the fat and fat-free compartments. This assumption has been shown to be wrong, which has implications to such methods as total body water, ^{40}K spectrometry, underwater weighing, and skinfold thickness. The consequences of applying these equations to children are summarized in table 5.

Table 5. Consequences of using adult criteria for the assessment of percent fat in children

Method	Implication to % fat	
	Underestimated	Overestimated
Underwater densitometry		X
Total body water	X	
^{40}K		X
Skinfolds	X	

Total body water. In adults, water content of fat-free tissue is taken as 73%. It is assumed to be negligible in fat. Thus, fat-free mass can be calculated by dividing total body water by 0.73. This equation is not applicable for children, because the water content of their fat-free mass is as high as 75–77% (105). The use of adult criteria in this case will therefore *underestimate* percent fat.

^{40}Potassium spectrometry. In adult males, 1 kg of fat-free mass is assumed to contain 2.66 g potassium (107). Recent data (105, 108) suggest that the respective value is lower in children (e.g. 2.50 g/kg in a 10-year-old boy). Thus, the use of adult criteria for the ^{40}K method will *overestimate* percent fat.

Underwater densitometry. Estimation of percent fat from body density is based on constant density values for fat (e.g., 0.900 g/cm^3) and fat-free body (e.g., 1.100 g/cm^3). As shown recently (105, 108) the density of fat-free body in prepubescents is considerably lower than in adults. Using adult equations therefore will considerably *overestimate* percent fat of a child. The error may be as high as 7% of body weight (8).

Skinfold thickness. Measurement of the subcutaneous thickness over specific skin sites has been the most common method for assessing percent fat of adults and children alike. Equations are available to transform the sum of two or more folds into body density. Validation studies have shown that, by using adult equations, body density of children is overestimated (Lohman, 1982) and percent fat is underestimated.

Recommended method

Total body water, ^{40}K spectrometry, and underwater weighing are logistically demanding and may not be suitable for a large-scale fitness survey. The skinfold measurement is fast, inexpensive, socially acceptable, and educational to the young subject. It is, however, less objective than the other methods, and its reliability is questionable when used with obese subjects. If adopted for NHANES, it will require thorough training of teams and strict quality control. With such precautions, the skinfold technique can predict percent fat with a SEE that is not different from that achieved with any of the more sophisticated methods (Wilmore, chapter 4).

The issue of site selection for skinfold measurement is unresolved. The number of sites used with children has ranged from 1 to 12. Recent studies (e.g., 104, 105) have used two sites—triceps and subscapula. Others have felt that the lower limb must be represented in any site combination (e.g., 109). Until this issue is resolved, we recommend that for children and adolescents NHANES data base will include at least seven sites (triceps, subscapular, chest, suprailiac, abdomen, anterior thigh, and medial calf). This will allow for future statistical experimentation.

Scores can be described either in a raw form (i.e., the sum of folds) or by using an equation to convert skinfold thickness to percent fat. One such equation that, tentatively, can be adopted is that recently offered by Boileau et al. (105).

$$\text{For girls \% fat} = 1.35(\text{sum TR} + \text{SUBSC})^2 - 0.012(\text{sum TR} + \text{SUBSC}) - 2.4$$

$$\text{For boys \% fat} = 1.35(\text{sum TRI} + \text{SUBSC}) - 0.012(\text{sum TRI} + \text{SUBSC})^2 - 4.4$$

in which TRI = triceps and SUBSC = subscapular folds in millimeters.

The above discussion indicates the potential pitfalls in borrowing adult criteria for assessment of body composition. More research is needed to establish criteria for children of both sexes, in whichever method is adopted for NHANES. It is likely that different equations are needed for prepubescents, pubescents, and postpubescents. There is a special need for *cross-validating* these equations in populations outside those used in any specific study.

Safety considerations

No epidemiologic data are available on the inherent risk in exercise testing of children and adolescents. It seems, however, that such risk is considerably lower than that for adults (110). The two main exercise-induced complications encountered in adults—myocardial ischemia and ventricular fibrillation—are not expected in children. Possible exceptions are such rare congenital heart defects as aortic stenosis, coarctation of the aorta, or tetralogy of Fallot (following surgical correction).

Safety precautions taken in the NHANES for the testing of adults are therefore more than adequate for school-age children. It is important, however, that personnel engaged in the testing of children become familiar with the specific criteria for exclusion from testing and for termination of an exercise test, as summarized in tables 6 and 7, respectively.

Assessment of habitual activity: Available methods to assess physical activity in children

A detailed survey of methods that assess habitual activity in adults is given in chapter 24. With some modifications, such methods are also suitable for children. The need for such modifications stems from differences in the daily routine (e.g., a child spends many hours at school), the ability to report reliably (parents rather than the young child may need to answer

Table 6. Contraindications for exercise testing in pediatric patients

Acute febrile condition
Acute inflammatory cardiac disease, e.g., pericarditis, myocarditis, acute rheumatic heart disease
Congestive heart failure—uncontrolled
Bronchial asthma—child dyspneic at rest, or whose resting FEV$_{1.0}$ or PEF is less than 60% of height-predicted value
Acute renal disease, e.g., acute glomerulonephritis
Acute hepatitis, during 3 months since onset
Insulin-dependent diabetes mellitus—child has not taken his prescribed insulin, or is ketoacidotic
Drug overdose affecting cardiorespiratory response to exercise, e.g., digitalis or quinidine toxicity, salicylism, antidepressants
Seizure disorder—uncontrolled

FEV$_{1.0}$ = Forced Expired Volume in the first second.
PEF = Peak Expiratory Flow.
Source: Adapted from American Heart Association (1982). Reproduced from Bar-Or (1983).

questionnaires or be interviewed), and body size (some monitoring devices that can be carried by adults are too heavy for children).

Each method for assessing activity has merits and weaknesses. Those that are simple and logistically suitable for mass testing (e.g., activity questionnaires) lack objectivity, reliability, and validity. Conversely, those that are objective and valid may be too restrictive or socially not acceptable to the child (measurement of $\dot{V}O_2$ and to a lesser extent of heart rate by a portable device). Although objective, methods such as time-and-motion studies or heart rate monitoring are interventional and may induce changes in the spontaneous activity of the child. Following is a brief discussion of available methods, with emphasis on their use with children.

Recall questionnaire/interview

This method is the most commonly used, particularly within large-scale surveys, for assessing activity habits (1, 2, 111–114). For use with children, questions must cover activities at school and after school hours: participation in physical education classes, during recess; in teams (intramural, school, or club); means of transportation to and from school; recreational activities after school and during weekends with/without other family members; seasonal changes (including separation into school-year vs. summer vacation); and exercise-induced complaints. Information must be sought on the intensity, duration, frequency, and type of activity, as well as the *attitudes* of the child and family toward physical activity and its relevance to health; the relationship between the child's level of activity and that of other family members; most/least favorable sports; and factors that may limit the child's activity (disease or illness; distance from, or nonavailability of, sport facilities; lack of free time; nonsupportive parents; cultural or religious constraints).

Table 7. Criteria for termination of an exercise test in the pediatric age group

Clinical	Symptoms: chest pain, severe headache, dizziness, chills, sustained nausea, inappropriate dyspnea
	Signs: sustained pallor, clammy skin, disorientation, inappropriate tachypnea, inappropriate affect
Electrocardiographic	Ventricular tachycardia
	Supraventricular tachycardia
	ST segmental depression, or elevation, of more than 3 mm
	Intracardiac block—triggered by exercise
	Premature ventricular contractions with increasing frequency
Blood pressure	Excessive levels: systolic BP above 240 mmHg diastolic BP above 120 mmHg
	Progressive fall in BP

Source: Reproduced from Bar-Or (39).

The validity of any recall questionnaire depends on how informed the respondent is about the subject matter, his/her memory, and objectivity. Semantics may also determine the outcome (exercise, activity, and disease mean different things to different people). It is therefore important that parents answer the questionnaires for their young children (tentatively those in their first decade of life) and that terminology and examples are compatible with the jargon and culture of any given subpopulation. If questionnaires are resorted to in the NHANES, it is important that they be investigator administrated. Whenever feasible, both child and parent should be interviewed.

Although the recall questionnaire is not the method of choice for assessing the habitual activity of individuals, it is still a useful and logistically feasible means of studying populations (112, 115).

Self-keeping diary

This technique obviates the reliance on memory, and it has been found useful by various authors (116–119). A major drawback of this method is its interventional nature, which may be more manifested in a less disciplined child than in an adult. Its test-retest reliability is high (r = 0.91) among 10–18-year-old girls and boys when categories of activity are tabulated each 15 min (116). No data are available on the reliability of diary taking among children younger than 10 years. Nor are there data on the validity of this method among children. One can assume, however, that both reliability and validity will strongly depend on the cooperation and intelligence of the child.

Observation by an investigator

This is based on monitoring and recording activities by an observer or through photographic techniques. Some authors have used descriptive criteria in their analysis (120–122). Others have attempted the time-and-motion principle, in which the child's activities are broken down to such elements as sitting, walking, running, climbing, or carrying loads; and the duration of each is calculated (123–126). Such observations, although objective and quantifiable, are time consuming and may induce some loss of spontaneity in the child's activities. The photographic techniques are limited to confined spaces and should not be considered for epidemiologic studies.

Movement counters

These devices record mechanical movement of the body or part of the body.

Pedometers. Usually hung on the waist, pedometers are meant to record the number of walking or running steps. They have been used in studies with healthy and sick children (127–130). Although inexpensive, simple to use, and noninterventional, pedometry lacks in validity and has a low reliability (131). In children, pedometers underestimate the number of steps during slow walking and overestimate it during fast running (127, 128). They can neither differentiate walking from running nor detect the speed of locomotion. Activities performed by the upper limbs and the trunk are usually not detected.

Actometers. Worn on the wrist or the ankle. They have a watchlike mechanism that records acceleration and deceleration of the limb. Actometers were first introduced in 1959 for the study of children's activity (132) and have since been used in pediatric populations by various authors (133–136).

Unlike the pedometer, an actometer registers also the *intensity* of the movement and not only the number of movements. This may explain the higher correlation between actometer testing (fixed to the ankle) and treadmill speed than between pedometer testing and treadmill speed (129). Actometers seem also to be better predictors of habitual activities (128). Test-retest correlation was 0.80 in 33 mentally retarded children (134). However, marked differences were recorded when two supposedly identical units were attached in tandem to a moving device (137).

Some new types of movement monitors have been developed in recent years (for review see (112)). Their reliability and validity has yet to be tested with children. One (138) is a counter where an electric circuit is closed through the movement of a mercury column, which is triggered by body movement (Large-Scale Integrated Motor Activity Monitor). This device has been used since in various adult age groups for epidemiologic purposes (118). Its possible suitability for children has not been reported.

Heart rate integrators

The use of these devices is based on the linear relationship between submaximal heart rate and $\dot{V}O_2$. The most common means of transduction of myocardial depolarization in children to a monitoring device has been by a magnetic tape, using chest electrodes as sensors (139, 140). Other means include telemetry (126, 141, 142) and, more recently, registration of the R–R intervals in solid-state memories (143). The use of electrochemical cells to record heart beats has been attempted with children (144, 145) but with little following.

Because the heart rate-$\dot{V}O_2$ regression varies among persons, one has to determine the "calibration line" for each subject. This can then be used to transform the heart rate data to energy expenditure. One major weakness of this concept is the *intraindividual* variation in the heart rate-$\dot{V}O_2$ relationship, especially at low metabolic levels (e.g., 146). It is not possible, for example, to separate the contribution of emotional excitement from that of increased metabolism to the rise in heart rate at any given time. Climatic changes also alter the heart rate-$\dot{V}O_2$ relationship. Systematic attempts are yet to be made to partial out these factors, such that heart rate can indeed reflect the metabolic level. The above regression line depends also on the muscle group (e.g., arms vs. legs) that performs the activity. As shown among children with spastic cerebral palsy (141), different heart rate-$\dot{V}O_2$ regression lines may be obtained at different exercise intensities. Ideally, one would like to establish individual calibration lines for activities that are representative of those anticipated from each child. This, however, is impractical in a large-scale survey.

Validation of the heart rate integration method has been done in the laboratory, using indirect (146, 147) or direct (148) calorimetry as a criterion. Validation for *spontaneous* activities is harder to perform for the lack of a "gold standard." One attempt has been the comparison with food intake among 8-year-old children (147).

In spite of the need for further validation, the heart rate integration method may be a useful method for NHANES, because of its objectivity and social acceptability in children.

Doubly labeled water

This technique has recently been attempted with humans, including children (149). Its principle is described in chapter 24. The merits of this method are its social acceptability, objectivity, lack of interference in spontaneous activity, and the long duration (5–10 days) that can be covered by one observation. The major disadvantage is that only the total energy expenditure is assessed over a time period. At present, the method is prohibitively expensive for use in large-scale surveys; and its validity and reliability are still undetermined. If further technological breakthroughs are made in the near future, this technique may be considered for inclusion in NHANES.

Calorimetry

Whether done directly in a calorimeter or indirectly by collection of expired gas, calorimetry is the most valid method for determining energy expenditure. (For a recent review, see (150)). This method requires that subjects stay in confined spaces or carry the gas collection system with them. Portable spirometers have been used with adults in industry and for analysis of sports activities. They are, however, too heavy and restrictive to be carried by children. Until further technologies are developed, calorimetry cannot be considered for studies that are meant to observe the spontaneous activity of children. It still has an important role for the calibration and validation of other techniques.

Energy intake

During a state of calorie balance, energy intake equals expenditure. The food intake of an individual or of populations can therefore reflect their physical activities. As in methods that assess physical activity, the food intake techniques vary in sophistication, reliability, validity, and cost. This topic is reviewed in chapter 8.

Recommended methods to assess habitual activity

Ideally, an instrument used for assessing the habitual physical activity of children and adolescents within a large-scale study should be informative (encompassing in- and off-school activities), objective, reliable, valid, noninterventional, and socially acceptable by the child. None of the above techniques satisfy all these criteria. One must therefore resort to a combination of two or more modes of measurement that will complement each other.

It is recommended for NHANES that an interviewer-administered questionnaire be used in conjunction with heart rate integration. The latter must cover 3 days, including 1 weekend day. One may need to consider using alternative questionnaires for different parts of the United States or for specific subpopulations, to account for semantic and cultural differences. Another approach to consider is the addition of a self-keeping diary for a 3-day period. Some pilot work is needed to compare the characteristics of various heart rate integrators that are commercially available, particularly regarding their suitability for children. An additional pilot study is indicated to select the exercise conditions at which heart rate-$\dot{V}O_2$ regression lines will be determined.

The sick child

NHANES is meant to study the "general population" of the United States. Although the great majority of school-age participants will be free of disease, a representative sample will include also children with a disease or an illness. One can expect to encounter fair numbers of children and adolescents with such common conditions as obesity (10–20% of the total sample), bronchial asthma (3–6%), mental retardation (3–5%), seizure disorder (1–2%), cerebral palsy (0.1–0.5%), and diabetes mellitus (0.2–0.4%).

The planners of NHANES need to reach a policy decision regarding the inclusion of such patients. Their inclusion will have implications to the preparation of teams (need for some understanding of the diseases and their possible effects on exercise performance, means of motivating the physically and cognitively disabled child, thorough knowledge of safety precautions), the test protocols (need for arm or leg cycle ergometry instead of treadmill, shorter protocols for the mentally retarded child), and the data analysis (Should one establish separate norms for specific patient groups?).

Some review articles (151–154) and books (39, 47) are available that discuss the specific considerations in testing of sick children. These can be consulted if such children are included in NHANES.

Summary and recommendations

Outlined in this chapter are the rationale, alternative strategies, recommended test items, safety precautions, and logistic constraints of assessing physical fitness and habitual activity of school-age children and adolescents, within a nationwide health survey.

Children have certain morphologic, physiologic, and behavioral characteristics that may affect their attitude to and performance in fitness testing. These must be recognized before a testing battery is constructed.

The following fitness components are recommended for inclusion in NHANES: maximal aerobic power; peak anaerobic power and muscle endurance of the upper and lower limbs; muscle strength of the upper and lower limbs; and the trunk, body height, mass, and composition.

Maximal aerobic power will be assessed by a direct measurement of maximal O_2 uptake ($\dot{V}O_2$max). A progressive, continuous "all-out" protocol of running on a treadmill is the method of choice. A logistically simpler approach is the use of a cycle ergometer modified for children in which both $\dot{V}O_2$max and peak mechanical power can be determined. Irrespective of the ergometer, the duration of the test should not exceed 12 min.

Peak anaerobic power and muscle endurance of the upper and lower limbs will be tested by the Wingate Anaerobic Test. This test is highly reliable among children and valid as a task that taxes the anaerobic system.

Muscle strength will be determined, measuring the knee extensors, elbow extensors, and possibly a major muscle group of the trunk. The measuring device of choice (e.g., Cybex II, or Kin Com) must yield isometric and isokinetic data. A system of strain gauges can be used as an alternative, less expensive device. This system will yield only isometric data. For isokinetic measurements a pilot study is indicated that will determine the optimal angular velocity for each muscle group.

Flexibility has been suggested as a health-related fitness component for adults and children. This author is not convinced that it should be included within a test battery for children. Moreover, with some manipulation, the range of motion around many joints can be changed *within minutes*. This detracts from the reliability and validity of flexibility testing.

Body composition, although not a performance item, is highly related to children's fitness and may affect their state of health and, possibly, their health in future years. Various sophisticated methods are available for partitioning of the body into fat and fat-free mass. For logistical and ethical reasons it is recommended that, within a nationwide survey, one use the skinfold method. Some pilot work is needed to establish equations that relate skinfold thickness to body density throughout the whole school-age range.

Habitual activity will be assessed as a variable that affects physical fitness. It should also be considered a variable which by itself is an indicator of a child's present and, possibly, future health.

No single method is available that can determine both the overall calorie expenditure and the nature of the activities throughout the day. Two methods will therefore be combined to describe habitual activity: a detailed recall questionnaire (interviewer administered) and a 48–72-hour integration of heart rate. The latter will be done in conjunction with determination of the heart rate-to-$\dot{V}O_2$ relationship in each subject. It is possible that, by the time NHANES III is launched, the doubly labeled water ($^2H^1H^{18}O$) technique will be suitable for use in mass surveys. This method may then be considered for inclusion in the study. Some pilot work is indicated for construction of a child-oriented activity questionnaire (more than one version may be needed to account for the diverse cultural and ethnic groups in the United States. Some refinements are also needed to increase the social acceptability of devices currently used for heart rate monitoring.

Although NHANES III is directed towards the general population, special considerations are warranted regarding the sick child. These are related to methodology of testing, exclusion criteria, and safety during a test. This chapter also outlines safety measures in testing the general child population.

References

1. Engström, L-M: Physical activity of children and youth. *Acta Paediatr. Scand. Suppl.* 283:101–105, 1980.
2. Ilmarinen, J., and J. Rutenfranz: Longitudinal studies of the changes in habitual physical activity of school children and working adolescents. In: Berg, K., and B.O. Eriksson (eds.) *Children and Exercise IX.* Baltimore, University Press, 1980. pp. 149–159.
3. Kemper, H.C.G., H.J.P. Dekker, M.G. Ootjers, B. Post, J. Snel, P.G. Splinter, L. Storm-van-Essen, and R. Verschuur: Growth and health of teenagers in the Netherlands: survey of multidisciplinary longitudinal studies and comparison to recent results of a Dutch study. *Int. J. Sports Med.* 4:202–214, 1983.
4. Inbar, O., R. Dotan, T. Trousil, and Z. Dvir: The effect of bicycle crank-length variation upon power performance. *Ergonomics* 26:1139–1146, 1983.
5. Klimt, F., and G.B. Voigt: Investigations on the standardization of ergometry in children. *Acta Paediatr. Scand. Suppl.* 217:35–36, 1971.
6. Pařízková, J., A. Adamec, J. Berdychová, J. Čermák, J. Horná, and Z. Teplý: *Growth, Fitness and Nutrition in Preschool Children.* Prague, Charles University, 1984.
7. Boileau, R.A., T.G. Lohmann, and M.H. Slaughter: Exercise and body composition in children and youth. *Scand. J. Sports Sci.* (In press).
8. Lohman, T.G., R.A. Boileau, and M.H. Slaughter: Body composition in children and youth. In: Boileau, R.A. (ed.) *Advances in Pediatric Sport Sciences. Vol I.* Champaign, Human Kinetics, 1984. pp. 29–57.
9. Åstrand, P.O.: Experimental studies of physical working capacity, in relation to sex and age. Copenhagen, Munksgaard, 1952.
10. Cumming, G.R., D. Everatt, and L. Hastman: Bruce treadmill test in children: normal values in a clinic population. *Am. J. Cardiol.* 4:69–75, 1978.
11. MacDougall, J.D., P.D. Roche, O. Bar-Or, and J.R. Moroz: Maximal aerobic capacity of Canadian schoolchildren. Prediction based on age-related oxygen cost of running. *Int. J. Sports Med.* 4:194–198, 1983.
12. Skinner, J.S., O. Bar-Or, V. Bergsteinová, C.W. Bell, D. Royer, and E.R. Buskirk: Comparison of continuous and intermittent tests for determining maximal oxygen uptake in children. *Acta Paediatr. Scand. Suppl.* 217:24–28, 1971.
13. Daniels, J., and N. Oldridge: Changes in oxygen consumption of young boys during growth and running training. *Med. Sci. Sports* 3:161–165, 1971.
14. Davies, C.T.M., and K. Young: Effects of external loading on short-term power output in children and young male adults. *Eur. J. Appl. Physiol.* 52:351–354, 1984.
15. Bruce, R.A., and J.R. McDonough: Stress testing in screening for cardiovascular disease. *Bull. N.Y. Acad. Med.* 45:1288–1305, 1969.
16. Cumming, G., and W. Friesen: Bicycle ergometer measurement of maximal oxygen uptake in children. *Can. J. Physiol. Pharmcol.* 45:937–946, 1967.
17. Cunningham, D.A., B. MacFarlane van Waterschoot, D.H. Paterson, M. Lefcoe, and S.P. Sangal: Reliability and reproducibility of maximal oxygen uptake measurement in children. *Med. Sci. Sports* 9:104–108, 1977.
18. Shephard, R.J., C. Allen, O. Bar-Or, C.T.M. Davies, S. Degré, R. Hedman, K. Ishii, M. Kaneko, J.R. Lacour, P.E. di Prampero, and V. Seliger: The working capacity of Toronto schoolchildren. Part I, II. *Can. Med. Assoc. J.* 100:560–566, 705–714, 1969.
19. Máček, M., and J. Vávra: The adjustment of oxygen uptake at the onset of exercise: A comparison between prepubertal boys and young adults. *Int. J. Sports Med.* 1:70–72, 1980.
20. Cooper, D.M., D. Weiler-Ravell, B.J. Whipp, and K. Wasserman: Growth-related changes in oxygen uptake and heart rate during progressive exercise in children. *Paediatr. Res.* 18:845–851, 1984.
21. Araki, T., Y. Toda, K. Matsushita, and A. Tsujino: Age differences in sweating during muscular exercise. *Jap. J. Phys. Fitness Sports Med.* 28:239–248, 1979.
22. Bar-Or, O.: Climate and the exercising child: A review. *Int. J. Sports Med.* 1:53–65, 1980.
23. Inbar, O.: Acclimatization to dry and hot environment in young adults and children 8–10 years old. Ed. D. Dissertation. Columbia University, 1978.
24. Bar-Or, O.: Physiologische Gesetzmassigkeiten sportlicher Activitat beim Kind. In: Howald, H., and E. Hahn (eds.) *Kinder im Leistungssport.* Basel, Birkhauser, 1982. pp. 18–30.
25. Macová, J., M. Sturmová, and M. Máček: Prolonged exercise in prepubertal boys in warm and cold environments. In: Ilmarinen, Y., and I. Valimaki. (eds.) *Pediatric Work Physiology X.* Berlin, Springer-Verlag, 1983.
26. Corbin, C.B.A.: *Textbook of Motor Development.* Dubuque, Iowa, W.C. Brown, 1973.
27. Geron, E., and O. Inbar: Motivation and anaerobic performance. Proceedings, International Seminar on the Art & Science of Coaching. Natanya, Wingate Institute, 1980. pp. 107–117.
28. Hermansen, L., and S. Oseid: Direct and indirect estimation of maximal oxygen uptake in prepubertal boys. *Acta Paediatr. Scand. Suppl.* 217:18–23, 1971.

29. Cunningham, D.A., and R.B. Exnon: The working capacity of young competitive swimmers, 10–16 years of age. *Med. Sci. Sports.* 5:227–231, 1973.
30. Mocellin, R., H. Lindermann, J. Rutenfranz, and W. Stresny: Determination of W170 and maximal oxygen uptake in children by different methods. *Acta Paediatr. Scand. Suppl.* 217:13–17, 1971.
31. Stewart, K.J., and B. Gutin: The prediction of maximal oxygen uptake before and after physical training in children. *J. Human Ergol.* 4:153–162, 1975.
32. Wilmore, J.H., and P.O. Sigerseth: Physical work capacity of young girls 7–13 years of age. *J. Appl. Physiol.* 22:923–928, 1967.
33. Woynarowska, B.: The validity of indirect estimations of maximal oxygen uptake in children 11–12 years of age. *Eur. J. Appl. Physiol.* 43:19–23, 1980.
34. Adams, F.H., E. Bengtsson, H. Berven, M. Börjeson, I. Engström, D. Ikkos, B. Jonsson, P. Karlberg, and S. Kraepelin: Determination by means of a bicycle ergometer of the physical working capacity of children. *Acta Paediatr. Scand. Suppl.* 118:120–122, 1959.
35. Bengtsson, E.: The working capacity in normal children, evaluated by submaximal exercise on the bicycle ergometer and compared with adults. *Acta Med. Scand.* 154:91–109, 1956.
36. Rutenfranz, J.: Entwicklung und Beurteilung der korperlichen leistungs Fahigkeit bei Kindern und Jugendlichen. Basel, Karger, 1964.
37. Strong, W.B., D. Spencer, M.D. Miller, and M. Salehbhai: The physical working capacity of healthy black children. *Am. J. Dis. Child.* 132:244–248, 1978.
38. Jetté, M., N.J. Ashton, M.T. Sharratt: Development of cardiorespiratory step-test of fitness for children 7–14 years of age. *Can. J. Publ. Health* 75:212–217, 1984.
39. Bar-Or, O.: *Pediatric Sports Medicine for the Practitioner: From Physiologic Principles to Clinical Applications.* New York, Springer-Verlag, 1983.
40. American Alliance Health Physical Education and Recreation. *AAHPER Youth Fitness Test Manual.* Washington D.C., 1976 (revised).
41. Conger, P.R., H.A. Quinney, and R. Gauthier, D. Massicotte: The CAHPER Fitness Performance Test. 1966–1979. *CAHPER J.* 48:3–10, 1981.
42. Krahenbuhl, G.S., R.P. Pangrazi, L.N. Burkett, M.J. Schneider, and G. Peterson: Field estimation of $\dot{V}$O2 max in children eight years of age. *Med. Sci. Sports,* 9:37–40, 1977.
43. Krahenbuhl, G.S., R.P. Pangrazi, G.W. Petersen, L.N. Burkett, and H.J. Schneider: Field testing of cardiorespiratory fitness in primary school children. *Med. Sci. Sports,* 10:208–213, 1978.
44. Jackson, A.A., A.E. Coleman: Validation of distance run tests for elementary school children. *Res. Q.* 47:88–94, 1976.
45. Roche, P.D.: The development of norms for run-walk tests for children aged 7–17. *CAPHER J.* 6–13, 1980.

45.a Bar-Or, O., and O. Inbar: Relationships among anaerobic capacity, sprint and middle distance running of school children. In: Shephard, R.J., and H. Lavallée (eds.) *Physical Fitness Assessment.* Springfield, C.C. Thomas, 1978. pp. 142–147.

46. Riopel, D.A., A.B. Taylor, and A.R. Hohn: Blood pressure, heart rate, pressure rate product and electrocardiographic changes in healthy children during treadmill exercise. *Am. J. Cardiol.* 44:697–704, 1979.
47. Godfrey, S.: Exercise testing in children. *Applications in Health and Disease.* Philadelphia, W.B. Saunders, 1974.
48. James, F.W.: Exercise testing in children and young adults: an overview. *Cardiovasc. Clin.* 9:187–203, 1978.
49. Goldberg, S.J., R. Weiss, and F.H. Adams: A comparison of the maximal endurance of normal children and patients with congenital cardiac disease. *J. Paediatrics* 69:46–55, 1966.
50. Petäjoki, M.L., M. Arstila, and I. Välimäki: Pulse-conducted exercise test in children. *Acta Paediatr. Belg.* 28 Suppl.:40–47, 1974.
51. Shephard, R.J., H. Lavallée, J.C. Jéquier, R. LaBarre, M. Rajic, and C. Beaucage: Seasonal differences in aerobic power. In: Shephard, R.J., and H. Lavallée (eds.) *Physical Fitness Assessment.* Springfield, C.C. Thomas, 1978. pp. 194–210.
52. Bell, C.W.: A comparison of three methods for obtaining maximal oxygen consumption in children of low-normal and below normal intelligence. Unpublished MSc. Thesis. Penn. State University, 1971.
53. Pollock, M.L., R.L. Bohannon, K.H. Cooper, J.J. Ayres, R. Ward, S.R. White, and A.C. Linnerud: A comparative analysis of four protocols for maximal stress testing. *Am. Heart J.* 92:39–46, 1967.
54. Stamford, B.: Maximal oxygen uptake during treadmill walking and running at various speeds. *J. Appl. Physiol.* 39:386–389, 1975.
55. Andersen, K.L., V. Seliger, J. Rutenfranz, and R. Mocellin: Physical performance capacity of children in Norway. Part I. Population parameters in rural inland community with regard to maximal aerobic power. *Eur. J. Appl. Physiol.* 33:177–195, 1974.
56. Iliev, I.B.: Maximal aerobic power in girls and boys aged 9 to 16 years who participate regularly in sport. In: Shephard, R.J., and H. Lavallée (eds.) *Physical Fitness Assessment.* Springfield, C.C. Thomas, 1978. pp. 172–176.
57. Rutenfranz, J., K.L. Andersen, V. Seliger, F. Klimmer, I. Berndt, and M. Ruppel: Maximal aerobic power and body composition during the pubertal growth period: similarities and differences between children of two European countries. *Europ. J. Paediatr.* 136:123–133, 1981.
58. Seliger, V., and Z. Bartunek: Mean values of various indices of physical fitness in the investigation of Czechoslovak population aged 12–55 years. CSTV Praha (CSSR), 1976.
59. Silverman, M., and S.D. Anderson: Metabolic cost of treadmill exercise in children. *J. Appl. Physiol.* 33:696–698, 1972.
60. Mrzena, B., and M. Máček: Use of treadmill and working capacity assessment in pre-school children. In: Borms, J., and H. Hebbelinck. (eds.) *Pediatric Work Physiology.* Basel, Karger, 1978. pp. 29–31.
61. Jéquier, J.C., R. Labarre, R.J. Shephard, H. Lavallée, C. Beaucage, and M. Rajic: Seasonal variations of CAPHER performance tests. In: Shephard, R.J., and H. Lavallée (eds.) *Physical Fitness Assessment.* Springfield, C.C. Thomas, 1978. pp. 85–93.
62. Ruskin, H.: Physical performance of school children in Israel. In: Shephard, R.J., and H. Lavallée (eds.) *Physical Fitness Assessment.* Springfield, C.C. Thomas, 1978. pp. 273–320.
63. Thorstensson, A., G. Grimby, and J. Karlsson: Fatiguability and fiber composition of human skeletal muscle. *Acta Physiol. Scand.* 98:318–322, 1976.
64. Sjödin, B.: The relationships among running economy, aerobic power, muscle power, and onset of blood lactate accumulation in young boys (11–15 years). In: Komi, P.V. (ed.) *Exercise and Sport Biology* Champaign, Ill., Human Kinetics Pub., 1982. pp. 57–60.
65. Sargeant, A.J., E. Hoinville and A. Young. Maximal leg force and power outout during short-term dynamic exercise. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 51:1175–1182, 1981.
66. McCartney, N., G.J.F. Heigenhauser, and N.L. Jones: Power output and fatigue of human muscle in maximal cycling exercise. *J. Appl. Physiol.: Respirat. Environ. Exerc. Physiol.* 55:218–224, 1983.
67. McCartney, N., G.J.F. Heigenhauser, A.J. Sargeant, and N.L. Jones: A constant velocity cycle ergometer for the study of dynamic muscle function. *J. Appl. Physiol.: Respirat. Environ. Exerc. Physiol.* 55:212–217, 1983 [a].
68. Sargeant, A.J., P. Dolan, and A. Thorne: Isokinetic measurement of maximal leg force and anaerobic power output in children. In: Ilmarinen, J., and I. Välimäki (eds.) *Children and Sport.* Berlin, Springer-Verlag, 1984. pp. 93–98.
69. Bar-Or, O.: The Wingate Anaerobic Test. Characteristics and applications (in French). *Symbioses* 13:157–172, 1981.

70. Bar-Or, O., R. Dotan, and O. Inbar: A 30-second all-out ergometric test—its reliability and validity for anaerobic capacity (abstract). *Isr. J. Med. Sci.* 13:326, 1977.

71. Evans, J.A., and H.A. Quinney: Determination of resistance settings for anaerobic power testing. *Can. J. Appl. Sport Sci.* 6:53–56, 1981.

72. Tirosh, E., P. Rosenbaum, D. Moroz, and O. Bar-Or: The reliability and feasibility of the Wingate anaerobic test among children with cerebral palsy and muscular dystrophy. In preparation.

73. Dotan, R., and O. Bar-Or: Climatic heat stress and performance in the Wingate anaerobic test. *Eur. J. Appl. Physiol.* 44:237–243, 1980.

74. Inbar, O., and O. Bar-Or: The effects of intermittent warm-up on 7–9 year-old boys. *Eur. J. Appl. Physiol.* 34:81–89, 1975.

75. Ayalon, A., O. Inbar, and O. Bar-Or: Relationship between two measurements of explosive strength and three measurements of anaerobic power. In: Nelson, E., and C.A. Morehouse (eds.) *International Series on Sports Sciences, Biomechanics IV.* Baltimore, University Press, 1974. pp. 143–151.

76. Bar-Or, O., R. Dotan, O. Inbar, A. Rotshtein, J. Karlsson, and P. Tesch: Anaerobic capacity and muscle fiber type distribution in man. *Int. J. Sports Med.* 1:82–85, 1980.

77. Inbar, O., and O. Bar-Or: Relationships of anaerobic and aerobic arm and leg capacities to swimming performance of 8–12 year old children. In: Shephard, R.J., and H. Lavallée (eds.) *Frontiers of Activity and Child Health.* Quebec, Pelican, 1977. pp. 283–292.

78. Dotan, R., and O. Bar-Or: Load optimization for the Wingate anaerobic test. *Eur. J. Appl. Physiol.* 51:409–417, 1983.

79. Murphy, M.M., J.F. Patton, and F.A. Frederick: A comparison of anaerobic power capacity in males and females accounting for the differences in thigh volume, body weight and lean body mass (abstract). *Med. Sci. Sports Exerc.* 16:108, 1984.

80. Thorland, W., G. Johnson, C. Cisar, T. Housh, and G. Tharp: Strength and anaerobic influences on running ability in young female athletes (abstract). *Med. Sci. Sports Exerc.* 16:158, 1984.

81. Bondi, K.R., B.L. Bennett, and D.V. Tappan: Physiological variables and dynamics of 30 sec. all out, high intensity bicycle ergometer exercise and training (abstract). *Med. Sci. Sports Exercise* 16:109, 1984.

82. Inbar, O., A. Rotstein, I. Jacobs, P. Kaiser, R. Dlin, and R. Dotan: The effects of alkaline treatment on short-term maximal exercise. *J. Sports Sci.* 1:95–104, 1983a.

83. Jacobs, I.: The effects of thermal dehydration on performance of the Wingate anaerobic test. *Int. J. Sports Med.* 1:21–24, 1980.

84. Jacobs, I., P.A. Tesch, O. Bar-Or, J. Karlsson, and R. Dotan: Lactate in human skeletal muscle after 10 and 30 sec. of supramaximal exercise. *J. Appl. Physiol.: Respirat. Environ. Exerc. Physiol.* 55:365–367, 1983.

85. Tharp, G.D., R.D. Newhouse, L. Uffelman, W.G. Thorland, and G.O. Johnson: Comparison of sprint and run times with performance on the Wingate anaerobic test. *Res. Q.* (in print).

86. Margaria, R., P. Aghemo, and E. Rovelli: Measurement of muscular power (anaerobic) in man. *J. Appl. Physiol.* 21:1662–1664, 1966.

87. di Prampero, P.E., and P. Cerretelli: Maximal muscular power (aerobic and anaerobic) in African natives. *Ergonomics* 12:51–59, 1969.

88. Kurowski, T.T.: Anaerobic power of children from ages 9 through 15 years. M.Sc. Thesis, Florida State University, 1977.

89. Alexander, J., and G.E. Molnar: Objective quantitative muscle testing in children: Pilot study. *Arch. Phys. Med. Rehab.* 54:224–228, 1973.

90. Asmussen, E.: Growth in muscular strength and power. In: Rarick, L. (ed.) *Physical Activity, Human Growth and Development.* New York, Academic Press, 1973. pp. 60–79.

91. Davies, C.T.M.: Strength and mechanical properties of muscle in children and young adults. *Scand. J. Appl. Sports Sci.* (in print).

92. Davies, C.T.M., M.J. White, and K. Young: Muscle function in children. *Eur. J. Appl. Physiol.* 52:111–114, 1983.

93. Edwards, R.H.T., and M. McDonnell: Hand-held dynamometer for evaluating voluntary muscle function. *Lancet* 2:757–758, 1974.

94. Carron, A.V., and D.A. Bailey: Strength development in boys from 10 through 16 years. *Mon. Soc. Res. Child Develop.* 39 (serial No. 157): 1–37, 1974.

95. Clarke, H.H., R.N. Groing, and B.H. Heath: Relation of maturity, structural and strength measures to the somatotypes of boys 9 through 15 years of age. *Res. Q.* 32:449–458, 1961.

96. Fowler, W.M. Jr., and G.W. Gardner: Quantitative strength measurements in muscular dystrophy. *Arch. Phys. Med. Rehab.* 48:629–644, 1967.

97. Heebøll-Nielsen, K.: Muscle strength of boys and girls, 1981 compared to 1956. *Scand. J. Sports Sci.* 4:37–43, 1982.

98. Nielsen, B., K. Nielsen, M. Behrendt Hansen, and E. Asmussen: Training of "functional muscular strength" in girls 7–19 years of old. In: Berg, K., and B. Eriksson. (eds.) *Paediatric Work Physiology, IX.* Baltimore, University Park Press, 1980. pp. 690–78.

99. Helewa, A., C.H. Goldsmith, and H.A. Smyth: The modified sphygmomanometer—an instrument to measure muscle strength: a validation study. Presented at the 14th International Conference Rheumatology, San Francisco, 1977.

100. Hosking, G.P., U.S. Bhat, V. Dubowitz, and R.H.T. Edwards: Measurements of muscle strength and performance in children with normal and diseased muscle. *Arch. Dis. Child.* 51:957–963, 1976.

101. Molnar, G.E., J. Alexander, and N. Gutfield: Reliability of quantitative strength measurements in children. *Arch. Phys. Med. Rehabil.* 60:218–221, 1979.

102. Campbell, W.C.: A high school physical fitness kit. *CAHPER J.* 43:3–5, 1977.

103. Leger, L., J. Stothart, W. Stewart, and A.W. Taylor: Physical fitness evaluation at the Edmonton YMCA. *CAHPER J.* 41:18–20, 1974.

104. National Children and Youth Fitness Study. Summary of findings. *J.O.P.E.R.D.* 56:44–90, 1985.

105. Boileau, R.A., T.G. Lohman, M.H. Slaughter, T.E. Ball, S.B. Going, and M.K. Hendrix: Hydration of the fat-free body in children during matuation. *Human Biol.* In Press.

106. Lohman, T.G.: Measurements of body composition in children. *J. Phys. Ed. & Rec.* 53:67–70, 1982.

107. Forbes, G.B., J. Gallup, and J.B. Hursh: Estimation of total body fat from potassium-40 count. *Sci.* 133:101–102, 1961.

108. Fomon, S.J., F. Haschke, E.E. Ziegler, and S.E. Nelson: Body composition of reference children from birth to age 10 years. *Am. J. Clin. Nutr.* 35:1169–1175, 1982.

109. Ross, W.D., D.T. Drinkwater, N.O. Whittingham, and R.A. Faulkner: Anthropometric prototypes: age six to eighteen years. In: Berg, K., and B.O. Eriksson. (eds.) *Children and Exercise IX.* Baltimore, University Park Press, 1980. pp. 3–12.

110. Cumming, G.R.: Exercise studies in clinical pediatric cardiology. In: Lavalée, H., and R.J. Shephard (eds.) *Frontiers of Activity and Child Health.* Quebec, Pelican, 1977. pp. 17–45.

111. Montoye, H.J.: Estimation of habitual physical activity by questionnaire and interview. *Am. J. Clin. Nutr.* 24:1113–1118, 1971.

112. Montoye, H.J., and H.L. Taylor: Measurement of physical activity in population studies: A review. *Human Biol.* 56:195–216, 1984.

113. Saris, W.H.M., R.A. Binkhorst, A.B. Craminckel, A.M. van der Veen-Hezemans, and F. van Waesberghe: Evaluation of somatic effects of a health education program for school children. *Bibl. Nutr. Diet.* 27:77–84, 1979.

114. Taylor, H.L., D.R. Jacobs, Jr., B. Schucker, J. Knudsen, A.S. Leon, and G. Debacker: A questionnaire for the assessment of leisure time activities. *J. Chronic Dis.* 31:741–755, 1978.

115. Andersen, K.L., R. Masironi, J. Rutenfranz, and V. Seliger: Habitual Physical Activity and Health. WHO regional publication. Europ. Series No. 6 Copenhagen, World Health Organization, 1978.

116. Bouchard, C., A. Tremblay, C. Leblanc, G. Lortie, R. Savard, and G. Thériault: A method to assess energy expenditure in children and adults. *Am. J. Clin. Nutr.* 37:461–467, 1983.

117. Durnin, J.V.G.A., and R. Passmore: *Energy, Work and Leisure.* London, Heinemann, 1967.

118. LaPorte, R., J.A. Cauley, C.M. Kinsey, W. Corbett, R. Robertson, R. Black-Sandler, L.H. Kuller, and J. Falkel: The epidemiology of physical activity in children, college students, middle-aged men, menopausal females and monkeys. *J. Chron. Dis.* 35:787–795, 1982.

119. Rutenfranz, J., I. Berndt, and P. Knauth: Daily physical activity investigated by time budget studies and physical performance capacity of school boys. *Acta Paediatr. Belg.* 28(Suppl.): 79–86, 1974.

120. Bullen, B.A., R.B. Reed, and J. Mayer: Physical activity of obese and non-obese adolescent girls appraised by motion picture sampling. *Am. J. Clin. Nutr.* 14:211–223, 1964.

121. Bruch, H.: Obesity in childhood. IV Energy expenditure of obese children. *Am. J. Dis. Child.* 60:1082–1109, 1940.

122. Corbin, C.B., and P. Pletcher: Diet and physical activity patterns of obese and non-obese elementary school children. *Res. Q.* 39:922–928, 1968.

123. Durnin, J.V.G.A.: Physical activity by adolescents. *Acta Scand. Suppl.* 217:133–135, 1971.

124. Ellis, M.J., and G.J.L. Scholtz: *Activity and Play of Children.* Englewood Cliffs (N.J.), Prentice Hall, 1978.

125. Viteri, F.E., and B. Torún: Nutrition, physical activity and growth. In: Ritzen, M., A. Aperia, K. Hall, A Larsson, A. Zetterberg, and R. Zeterstrom (eds.) *The Biology of Normal Human Growth.* New York, Raven Press, 1981. pp. 265.

126. Wade, M.G., and M.J. Ellis: Measurement of free-range activity in children as modified by social and environmental complexity. *Am. J. Clin. Nutr.* 24:1457–1460, 1971.

127. Kemper, H.C., and R. Verschuur: Validity and reliability of pedometers in habitual activity research. *Eur. J. Appl. Physiol.* 37:71–82, 1977.

128. Saris, W.H.M., and R.A. Binkhorst: The use of pedometer and actometer in studying daily physical activity in man. Part I: Reliability of pedometer and actometer. *Eur. J. Appl. Physiol.* 37:219–228, 1977.

129. Saris, W.H.M., and R.A. Binkhorst: The use of pedometer and actometer in studying daily physical activity in man. Part II: Validity of pedometer and actometer measuring the daily physical activity. *Eur. J. Appl. Physiol.* 37:229–235, 1977a.

130. Stefanik, P.A., F.P. Heald, and J. Mayer: Caloric intake in relation to energy output of obese and non-obese adolescent boys. *Am. J. Clin. Nutr.* 7:55–62, 1959.

131. Washburn, R., M.K. Chin, and H.J. Montoye: Accuracy of pedometer in walking and running. *Res. Q. Exerc. Sports* 51:695–702, 1981.

132. Schulman, J.L., and J.M. Reisman: An objective measure of hyperactivity. *Am. J. Mental Def.* 64:455–456, 1959.

133. Mack, R.W., and M.E. Kleinhenz: Growth, caloric intake and activity levels in early infancy: A preliminary report. *Human Biol.* 46:345–354, 1974.

134. Massey, P.S., A. Lieberman, and G. Batarsch: Measure of activity level in mentally retarded children and adolescents. *Am. J. Mental Def.* 76:259–261, 1971.

135. Porrino, L.J., J.L. Rapoport, D. Behar, D.R. Ismond, and W.E. Bunney: A naturalistic assessment of the motor activity of hyperactive boys. *Arch. Gen. Psychiatry* 40:688–693, 1983.

136. Rose, H.E., and J. Mayer: Activity, calorie intake, fat storage and the energy balance of infants. *Pediatrics* 41:18–29, 1968.

137. Johnson, C.F.: Hyperactivity and the machine: the actometer. *Child Dev.* 42:2105–2110, 1971.

138. LaPorte, R.E., L.H. Kuller, D.J. Kupfer, R.J. McPartland, G. Matthews, and C. Caspersen: An objective measure of physical activity for epidemiological research. *Am. J. Epid.* 109–158–168, 1979.

139. Gilliam, T.B., P. S. Freedson, D.L. Greenen, and B. Shahraray: Physical activity patterns determined by heartrate monitoring in 6–7 year-old children. *Med. Sci Sports Exerc.* 13:65–67, 1981.

140. Rutenfranz, J., V. Seliger, K.L. Andersen, et al.: Erfahungen mit einem transportablen Gerat zur kontinuierlichen Registrierung der Herzfrekwenz. *Eur. J. Appl. Physiol* 36:171–185, 1977.

141. Berg, K.: Heart rate telemetry for evaluation of the energy expenditure of children with cerebral palsy. *Am. J. Clin. Nutr.* 24:1438–1445, 1971.

142. Berg, K., and T. Olsson: Energy of school children with cerebral palsy as determined from indirect calorimetry. *Acta Paediatr. Scand. Suppl.* 204:71–80, 1970.

143. Saris, W.H.M., P. Snel, and R. A. Binkhorst: A portable heart rate distribution recorder for studying daily physical activity. *Eur. J. Appl. Physiol.* 37:17–25, 1977.

144. Bradfield, R.B., H. Chan, N.E. Bradfield, and P.R. Payne: Energy expenditure and heart rates of Cambridge boys at school. *Am. J. Clin. Nutr.* 24:1461–1466, 1971.

145. Bradfield, R.B., J. Paulos, and L. Grossman: Energy expenditure and heart rate of obese high school girls. *Am. J. Clin. Nutr.* 24:1482–1488, 1971.

146. Christensen, C.C., H. Frey, E. Foenstelien, E. Aadland, and H. Refsum: A critical evaluation of energy expenditure estimates based on individual O2 consumption/heart rate curves and average daily heart rate. *Am. J. Clin. Nutr.* 37:468–472, 1983.

147. Saris, W.H.M.: Aerobic power and daily physical activity in children. With special reference to methods and cardiovascular risk indicators. Ph.D. Dissertation, Catholic University, Nijmegen, Krips Repro. Meppel., 1982.

148. Dauncey, M.J., and W.P.T. James: Assessment of the heart rate method for determining energy expenditure in man using a whole body calorimeter. *Br. J. Nutr.* 42:1–13, 1979.

149. Schoeller, D.A.: Energy expenditure from doubly labelled water: some fundamental consideration in humans. *Am. J. Clin. Nutr.* 38:999–1005, 1983.

150. Horton, E.S.: Appropriate methodology for assessing physical activity under laboratory conditions in studies of energy balance in adults. In: E. Pollitt, and P. Amante (eds.) *Energy Intake and Activity.* New York, Alan R. Liss, Inc., 1984. pp. 115–129.

151. American Heart Association Council on Cardiovascular Disease in the Young: Standards for exercise testing in the pediatric age group. *Circulation* 66:1377A–1397A, 1982.

152. Bar-Or, O.: Exercise in pediatric assessment and diagnosis. *Scand. J. Sports Sci.* In press.

153. Kulangara, R.J., and W.B.: Strong. Exercise stress testing in children. *Comprehens. Ther.* 5:51–61, 1979.

154. Thorén, C.: Exercise testing in children. *Pediatrician* 7:100–115, 1978.

Chapter 11

Evaluating Fitness and Activity Assessments From the National Children and Youth Fitness Studies I and II

James G. Ross
Macro Systems, Inc.
Silver Spring, Maryland

The two National Children and Youth Fitness Studies (NCYFS I and II) are among several nationwide studies of physical fitness and habitual physical activity that are thought likely to yield information for decisions as to the design and implementation of related components in general population surveys. The objectives of NCYFS I and II were to develop normative data on the health-related physical fitness status of children and adolescents 6–18 years, to describe their physical activity habits, and to define the relationships between physical activity habits and measured fitness (1, 2). The purpose of this chapter is to share some of the lessons learned in NCYFS I and II for the conduct of related national surveys, especially those with an interest in using a constant set of measures across the life cycle.

The chapter begins with a review of the structure and major decision points in the design of NCYFS I and II, with special attention to meeting the physical fitness and exercise (PFE) objectives found in *Promoting Health/Preventing Disease: Objectives for the Nation* (3), which provides the impetus and framework for any federally sponsored activity in this area. We then examine the implications of these studies for construction of a fitness test battery and design of a physical activities survey.

NCYFS I and II: An overview

NCYFS I, funded by the Office of Disease Prevention and Health Promotion (ODPHP) of the U.S. Public Health Service, was the first study of the physical fitness of American schoolchildren to be performed on a national probability sample in nearly a decade and the first such study supported by a Federal health agency. NCYFS I, conducted in 1982–85, was limited to ages 10–18 years (1); NCYFS II, conducted in 1985–87, extended the original study to ages 6–9 years (2). Compared with the previous national probability studies in 1958, 1965, and 1975 (4), NCYFS I and II broke precedent in at least four ways:

- *Definition of physical fitness*—NCYFS I and II adopted a health-related definition of fitness, which downplayed certain aspects of physical performance (5). Their test batteries incorporated the four tests from the AAHPERD Health Related Physical Fitness

Test (6), released in 1980, with one additional performance test and several additional body composition measures. National probability studies prior to NCYFS I used the older American Association of Health, Physical Education, Recreation, and Dance (AAHPERD) Youth Fitness Test (7), released in 1957, which consisted of six items heavily weighted in the direction of motor performance or athletic fitness, e.g., a standing broad jump, shuttle run, and 50-yard dash. Both AAHPERD tests contain measures of abdominal strength-endurance and cardiorespiratory endurance. NCYFS I and II produced the first norms for AAHPERD's health-related test on a national probability sample (8, 9).

- *Scope of the universe*—The universe for NCYFS I and II encompassed all students attending regular public and private schools. Prior studies excluded students in private schools. Students attending special schools (e.g., schools for the orthopedically disabled or for the moderately retarded) were excluded from NCYFS I and II, as from prior studies; however, disabled students mainstreamed in regular schools were eligible for inclusion in NCYFS I and II (10, 11).
- *Field data collection methods*—Small, suitably trained field staffs roved among 25 primary sampling units, independently selected for NCYFS I and II, to work with teachers in collecting the fitness and physical activity data (12, 13). In NCYFS I, the field staff assumed exclusive responsibility for collecting certain data and remained at a school to help teachers to collect all other data (12). In NCYFS II, field staff collected all data, with the exception of the distance walk-runs, in one-on-one testing sessions (13). Prior national studies of American schoolchildren relied on hundreds of teachers to test their own students after minimal direct training (4).
- *Linking fitness to habitual physical activity*—NCYFS I participants completed a detailed survey of their physical activity habits. In NCYFS II, parents and teachers provided information on the activity habits of the young children. The physical activity profiles describe a significant aspect of the health behavior of American youngsters (14–18). Linked to the fitness data, they also permit estimation of concurrent relationships between physical activity habits and measured fitness (19, 20). Prior normative studies of the fitness of national probability samples of American schoolchildren did not combine fitness assessment with estimation of physical activity habits (4, 21).

Meeting the health promotion objectives for the nation

Fitness and activity assessments planned for any federally sponsored general population survey must accommodate the 1990 PFE health promotion objectives (3) or their successors (22, 23). The 11 1990 PFE objectives call for reduction in specific risk factors, increased public-professional awareness, improved services, and improved surveillance-evaluation systems. The objectives apply to all ages, targeting two special settings, the school and the workplace. The PFE objectives justify two types of research activities: (1) Collecting baseline data on fitness behavior and knowledge and (2) implementing improved systems for monitoring trends of participation in and for assessing the health effects of habitual physical activity (3). NCYFS I and II fell into both categories.

Table 1. 1990 health objectives for the Nation related to the fitness and physical activity habits of youth

By 1990, the proportion of children and adolescents participating regularly in appropriate physical activities, particularly cardiorespiratory fitness programs which can be carried into adulthood, should be greater than 90 percent.

By 1990, the proportion of children and adolescents ages 10 to 17 participating in daily school physical education programs should be greater than 60 percent.

By 1990, a methodology for systematically assessing the physical fitness of children should be established, with at least 70 percent of children and adolescents ages 10 to 17 participating in such an assessment.

By 1990, data should be available with which to evaluate the short- and long-term health effects of participation in programs of appropriate physical activity.

By 1990, data should be available for regular monitoring of national trends and patterns of participation in physical activity, including participation in public recreation programs in community facilities.

Planning research to address the PFE objectives has been impeded by difficulties in interpreting and operationalizing the objectives, which were set in 1980, generally without benefit of baseline data. Five of the 11 PFE objectives directly affected design of NCYFS I and provided a general framework for planning NCYFS II. (See table 1.) Three of them were targeted to behavioral changes in children and adolescents ages 10–17; the other two involved improvement of surveillance systems without regard for age. The difficulties and questions that arose in translating the PFE objectives into the NCYFS I design may be instructive in planning future studies.

Participation in appropriate physical activity

There are separate objectives on participation in appropriate physical activity (APA) for three age groups: 10–17, 18–65, and over 65 years of age. Each of the objectives uses the term "activity" somewhat differently, with "appropriate physical activity" replaced by "vigorous physical activity" for the middle group. For the group of relevance to NCYFS I, children and adolescents 10–17 years, the objective is that, by 1990, 90 percent participate in "appropriate physical activities, particularly cardiorespiratory programs which can be carried into adulthood." So stated, the objective raised a number of issues. The narrative supporting the PFE objectives included a definition of APA, but, we asked, should this definition be accepted or made more stringent (e.g., duration increased from 20 to 30 minutes; intensity increased from 60 to 70 percent of cardiorespiratory capacity)? How should the word "especially" be interpreted? Does the 90-percent goal refer to participation in APA, or must some portion of APA come from activities with carry-over value into adulthood? What activities should be considered to carry into adulthood? Operationally, can one expect to capture participation in APA through a paper-and-pencil survey? Is such behavior best inferred from a detailed description of the duration, frequency, and intensity of physical activities? Can we really expect children to report these variables reliably? Would we be better off assessing participation in APA by obtaining the student's estimate of the gross frequency of experiencing the behavioral signs of exercise stress (i.e., sweating, breathing hard, and rapid heartbeat)? Should the assessment of APA among children and adolescents 10–17 years be limited to school and leisure-time activity, or should work and household chores be included? If we decide to exclude work and chores, should we also exclude the physical education class, on the grounds that content from class to class is too variable to assess contribution to APA?

Another line of questioning arose. Why are we asking only about APA? What about other aspects of physical activity that may contribute to health, including activities to develop strength and flexibility? From the standpoint of habit formation and the prevention of a wide range of degenerative diseases, is it not important to find out what percentage of youngsters participate in regular physical activity at intensities lower than prescribed?

Participation in daily school physical education programs

One of the PFE objectives targeted to the age group 10–17 years is that 60 percent participate in daily school physical education programs. At first glance, the objective seems clear; still, there were problems in defining "participation," "physical education," and "daily." Does a child who receives no formal instruction in physical education but who receives physical education credit for monitored recess time participate in physical education? Similarly, if a varsity football player receives physical education credit but is exempted from class attendance, does he participate in physical education? What about the star tuba player who gets credit for physical education while attending band practice? What about students who obtain physical education credit for attending ROTC drills? Even if there is real instruction and students actually attend class, should they be considered daily participants if they spend most of the year in the bleachers, recuperating from an illness, or fending off a fit of "gymophobia"? Should students be regarded as daily participants in physical education if the teacher lectures on health education 1 or more days per week or if students are left to fend for themselves with no real instruction?

This initial line of questions leads logically to a second. Why do the objectives specify nothing about the *conduct* of the physical education program? Would it not be consistent with the interest in APA to determine how much time is spent during each class meeting in moderate to vigorous physical activity? Similarly, the narrative supporting the PFE objectives stresses the importance of lifetime physical activity. Why is there no specific interest in measuring the portion of class time spent on lifetime activities? If we are willing to acknowledge the importance of measuring the lifetime portion of physical education time, how should we define "lifetime" activity? Should the lifetime aspect of activity be determined empirically (e.g., activities that adults generally report on national surveys) or is there something intrinsic about an activity (e.g., that it typically requires no more than two people and can readily be performed by adults) that earns the "lifetime" designation?

Participation in school fitness testing

One of the more confusing objectives is that "a methodology for systematically assessing the physical fitness of children should be established, with at least 70 percent of children and adolescents ages 10–17 participating in such an assessment." The disconcerting aspect of this objective is that methodologies for assessing the fitness of schoolchildren already exist and have been reasonably well established for decades. Does the objective mean that 70 percent of children and adolescents 10–17 years will participate in *any* such testing program? Or is something special intended, e.g., participation in a test of the health-related components of fitness or in the "National Fitness Test," were a single test ever to be recognized? Given that the *Objectives for the Nation* (3) were published in parallel with the AAHPERD Health Related Physical Fitness Test (6), it seems unlikely that widespread participation in this or another specific health-related test was specifically envisioned in the objective.

Recognizing that the objective is listed under "improved surveillance/evaluation systems," does establishment of an assessment methodology also entail the generation of population-based norms and criterion-referenced standards for each age-sex segment of the population? Although NCYFS I and II could (and did) produce normative data (8, 9), we questioned whether the studies could validly assess whether students participate in school-based fitness testing programs and, if so, of what type. We suspected that involvement in the study as a subject would confound data collection activities with the variables we sought to measure. In NCYFS I, we abandoned any attempt to measure whether students *would have* participated in fitness testing had they not been recruited for the study and, if so, of what type, because students were our only data source. In NCYFS II, these and related questions were addressed to the teacher.

Data on health effects of physical activity

Another PFE objective affecting the NCYFS design was that "data should be available with which to evaluate the short- and long-term health effects of participation in programs of appropriate physical activity." The objective seemed limited to evidence on the effects of APA on cardiovascular functioning and reduction of coronary heart disease (CHD) risk, which are discussed in the supporting narrative. Our initial concern was that there was no interest in the content of the program, when it was begun, by what biological mechanism APA affects CHD risk, or related issues.

We were also concerned that the objective did not inquire into the health effects of aspects of habitual physical activity other than its appropriateness. For example, there is a growing body of evidence that weight-bearing activities, such as weight lifting and walking, increase bone density and, therefore, may be expected to have a measurable impact on the risk and timing of onset of osteoporosis. It appeared that the objective and our attempts to measure behavior should be broadened to encompass such health effects. We were equally concerned that the objective did not address the potential benefits of APA other than reduction of CHD risk, including effects on noninsulin-dependent diabetes, control of depression and anxiety, weight control, and cessation of addictive behaviors. It also seemed appropriate that, without downplaying the health benefits of exercise, we also pay some attention to the adverse health effects of physical activity. It was clear that this PFE objective, as stated, was unnecessarily restrictive; however, in designing the study, we could only hope to examine point-in-time estimates of relationships between activity habits and measured fitness. Most of these concerns amounted to an academic exercise.

Data for monitoring trends of participation in physical activity

One of the vaguest PFE objectives was that "data should be available for regular monitoring of national trends and patterns of participation in physical activity, including participation in public recreation programs in community facilities." How should we define "public recreation programs," "community facilities," and "participate"? Does this objective require more than an assessment of the utilization of community facilities for physical activity? How important is it to estimate the contribution of community facilities to participation in APA? Should "community" be defined broadly to include all organized settings, other than schools, in which children are likely to participate in physical activity? Is it important to determine the portion of total activity time spent in community facilities, so defined? In planning for development of community programs, is it important to partition the time spent in various physical activities by source (e.g., school physical education, other school programs, community facilities, and home-neighborhood)? Can we expect children to report the source of an activity accurately, given complex arrangements in providing physical activity (e.g., an afterschool program in a school operated by the parks and recreation department)? Even if children could make such differentiations, would not attempting to do so cause double reporting and, consequently, exaggerate physical activity levels? These issues were left for resolution in the NCYFS I planning period.

Resolution of problems with PFE objectives

These questions about the five listed PFE objectives were resolved with varied degrees of comfort in the NCYFS I design. Four of the five objectives were operationalized with reasonable success. In general, we allowed ourselves to cover different interpretations of the objectives (e.g., alternative definitions of APA; lifetime characterization of physical education activities). However, we decided that it was not feasible in NCYFS I to collect valid data related to participation in school fitness testing because the participation of the student and the school in the study would either change the school's usual testing procedure or alter the student's perception of it. Details on resolution of questions related to the PFE objectives are given later in this chapter.

In hindsight, we realize that, in NCYFS I, other PFE objectives could easily have been adapted to gather additional information to meet the spirit of the objectives. For example, one PFE objective is that "the percentage of adults who can accurately identify the variety and duration of exercise thought to promote most effectively cardiovascular fitness should be greater than 70 percent." There is no reason why this objective, incorporated into the 1985 National Health Interview Survey (NHIS) supplement as a three-part question, could not have been extended to children in the NCYFS I survey.

Midcourse review of the 1990 PFE objectives, conducted in 1985, subsequent to publication of NCYFS I, summarized progress toward the objectives and explored many of the conceptual issues that impeded the NCYFS I design (22). Subsequently, recommendations were drafted for a set of revised objectives to be achieved by the year 2000 (23). The draft 2000 PFE objectives affect conceptualization of research about school-age populations by lowering expectations about participation in APA by children and adolescents; introducing the importance of exercise at lower levels of intensity; lowering expectations regarding the percentage of students who can reasonably be expected to participate in daily physical education; specifying the need for a health-related approach to fitness testing supported by population-based norms and criterion-referenced standards; calling for research on the relationships of many forms of exercise to disease prevention and health promotion; and adding the determinants of participation in exercise programs as an interest area. Most important from the vantage point of justifying NCYFS II, the draft recommendations for the year 2000 extended all objectives originally targeted to children and adolescents 10–17 years down to the age of 6. In addition, one objective specifically called for research on children 6–9 years: "By 2000, population-based descriptive information about the levels of physical fitness and physical activity patterns of children ages 6 to 9 will be available" (23).

History of NCYFS I

NCYFS I was initially conceived by ODPHP as a 12-month study requiring no formal piloting of measurement techniques. The study was restricted to the age group 10–17 years because the *Objectives for the Nation* (3) and prior population-based normative studies (4) were limited to this age group. No formal piloting was anticipated on the assumption that the prior normative studies provided sufficient working experience. In the Government's original plan, approximately 2 months were allotted to complete the initial design and submit a clearance package to the U.S. Office of Management and Budget (OMB), which must approve all federally sponsored data collections; 3 months for field data collection; and 4 months for data preparation, data analysis, and report writing. In fact, the actual study took over twice as long, 28 months, by the time final versions of all reports and computer deliverables were accepted by the Government. Slippage in schedule resulted from the addition of a formal pilot study of all data collection procedures (which delayed the national study by a year) and from longer than expected Government reviews. A brief NCYFS I history is offered to illustrate the decision-making framework within which NCYFS I was designed and implemented (24).

Instrument design panels

Independent of but anticipating NCYFS I, ODPHP convened three panels to recommend content for the NCYFS I instruments. Two panels, convened before NCYFS I began, were to recommend measures for a fitness test battery. A third panel, convened shortly after the start of NCYFS I, was to recommend content for the physical activities survey (25).

Panel I: Alternative measures

Panel I was given the task of identifying the dimensions of fitness related to health and, for each dimension, coming to agreement on measures appropriate to three contexts: The clinical laboratory (Level 1), the less equipped clinical setting of a physician's or fitness specialist's office (Level 2), and field conditions in which mass testing would occur with minimal equipment (Level 3). Panel I came to agreement on five components of fitness: Cardiorespiratory endurance, muscular strength, muscular endurance, body composition, and flexibility. Clinical measures were identified for each component. For three of the five components, there was little or no distinction between measures appropriate for a physician's office versus field conditions (25).

Panel II: School mass-testing measures

Panel II inherited the recommendations of Panel I and was tasked with recommending a test battery for field use in the schools. This panel concurred with Panel I on field measures for four fitness components, disagreeing only on muscular strength. The resulting six-item test battery consisted of a mile walk-run; triceps and subscapular skinfolds; the Wells and Dillon sit-and-reach test; 60-second, bent-knee situps (arms folded across chest); the bent-arm hang; and pullups. The first four items were drawn from the AAHPERD Health Related Physical Fitness Test (6). The last two, measuring the muscular strength and endurance of the upper arm-shoulder girdle, were taken from the older AAHPERD Youth Fitness Test (7). Panel II rejected more sophisticated measures of muscular strength and endurance for four reasons: (a) Needed equipment would not be available in most schools for subsequent replication; (b) standardized equipment for women and children are generally not available; (c) the tests tended to be hazardous for an unprepared subject; and (d) strength-testing equipment needs frequent recalibration. The panel considered recommending that all students attempt the pullup test, with those students failing to complete a single pullup doing the bent-arm hand on a subsequent day. However, the panel ultimately decided that all students, regardless of sex, should attempt both tests (25).

Panel III: Physical activities survey

Panel III was convened to recommend content and structure for a survey "that could be used to monitor trends and patterns of participation in physical activities by school children and youth, including participation in school physical education programs, participation in public recreation programs in community facilities, and participation in other physical activities" (25). Panel III reached several decisions:

- Physical activity should be examined in four contexts: School physical education programs, other school programs, community organizations, and home-neighborhood. Distinguishing public and private community organizations may be useful to planners but is beyond the grasp of most children.
- Many physical education class activities take place for a short period of time. Therefore, detailed information about physical education class activities should be collected in a context that allows reporting of

activities by number of weeks, with general information about the physical education program collected once.

- Children's games and other physical activities should be added to the more traditional list of competitive sports and adult leisure pursuits.
- Valid estimates of the frequency and duration of physical activities can probably be obtained from 5th through 12th graders, but it might not be feasible to collect valid intensity data, especially from children under age 13.
- Penetration or seasonality is important, but collecting activity data retrospectively by month would strain recall, add excessively to respondent burden, and impair the validity of data. Therefore, simpler alternatives should be considered (e.g., by season).
- Both instructions and questions must be read out loud to students. It should not be assumed that even high school students can read.
- Consideration should be given to the potential need for two different surveys for the younger versus older children.
- It would be highly desirable to examine the relationships between other health behaviors (e.g., smoking, alcohol and drug use, and nutrition) and both fitness status and physical activity habits. However, doing so was beyond contract scope and could impede timely receipt of OMB clearance.

Testing of NCYFS I procedures

NCYFS I procedures were preliminarily field tested on a limited basis to complete the initial design. Subsequently, the design was formally piloted in three States to test assumptions and refine procedures. The formal piloting resulted in several noteworthy changes in the design. Some of these design adjustments are specific to the age group under study, to mass testing situations, or to the school setting; others apply more generally to fitness testing or surveys of physical activity habits (24).

Limited field testing

The fitness battery and physical activities survey were field tested on a limited basis in parallel with preparation of the OMB clearance package. The Government agreed to add height, weight, and waist girth measurements to permit the computation of body mass and ponderal indexes but declined to use a 1½ mile walk-run for the older students (14–18 years), which the *AAHPERD Health Related Physical Fitness Technical Manual* (26) suggests is a more valid measure of cardiorespiratory endurance, especially for fitter students. Three alternative versions of the physical activities survey were also prepared. Both the test battery and survey were field tested in the Harford County, Maryland, public school system. The test battery was administered to nine eighth graders on the assumption that a pubescent group represented the toughest test of the alleged sensitivity of certain study measures. (OMB limits such tests to a maximum of nine subjects.) Each of the three initial versions of the survey was tested on nine fifth graders, on the assumption that the youngest students would be the best test of the instrument's overall feasibility.

Based on the field tests and after debriefings with the students and teachers, one version of the survey was selected, revised, retested, and finalized. Regarding the test battery, we decided to: (a) Drop the bent-arm hang to control burden and in recognition of its documented lack of reliability; (b) use only female data collectors to reduce the potential sensitivity of the skinfold and waist measurements; (c) design and use two-person sit-and-reach boxes (shown in figure 1) to economize on time and make the test more fun without loss of accuracy; and (d) allow at least 2 days to complete the fitness tests. Within 2 months of the start of NCYFS I, the basic design and instrumentation were completed and an OMB clearance package was submitted to the Government on schedule.

Formal pilot testing

Close scrutiny of the OMB clearance package by various agencies within the U.S. Public Health Service produced numerous recommendations for improvement of protocols for taking the anthropometric measurements and for administration of the survey; it also forced clarification of the sampling design. However, the most significant outcome of these reviews was acknowledgment that the study explored new ground and that, therefore, a formal, OMB-sanctioned pilot test of study procedures should precede the national study. The OMB clearance request was revised to reflect the plan for a pilot test and resubmitted, with clearance to conduct the pilot finally being granted in mid-May 1983, 3 days before the projected abort date. A special feature of conducting data collection in schools is accommodating the school calendar, which includes not being able to

Figure 1. 2-person sit-and-reach box.

test students during the summer, when schools are not in session.

All study procedures were pilot tested on 1,006 students representing 18 schools in three Eastern States (Maryland, New Jersey, and Pennsylvania). In each State, four public schools, one Catholic school, and one non-Catholic private school were selected judgmentally within a single county to represent a wide range of conditions. This yielded four complete grade ranges in the public school systems and one each in the Catholic and non-Catholic private schools. Teachers played a major role in administering the test battery in the pilot test. Two weeks were allotted to test approximately 400 students in each county. The field staff member's responsibilities were to train the teacher, assure proper equipment setup, collect the skinfold data, intermittently monitor the teacher's data collection activities, and return to review and gather all completed surveys and fitness test cards.

Pilot test results

Based on the pilot test, we arrived at several conclusions resulting in major design changes:

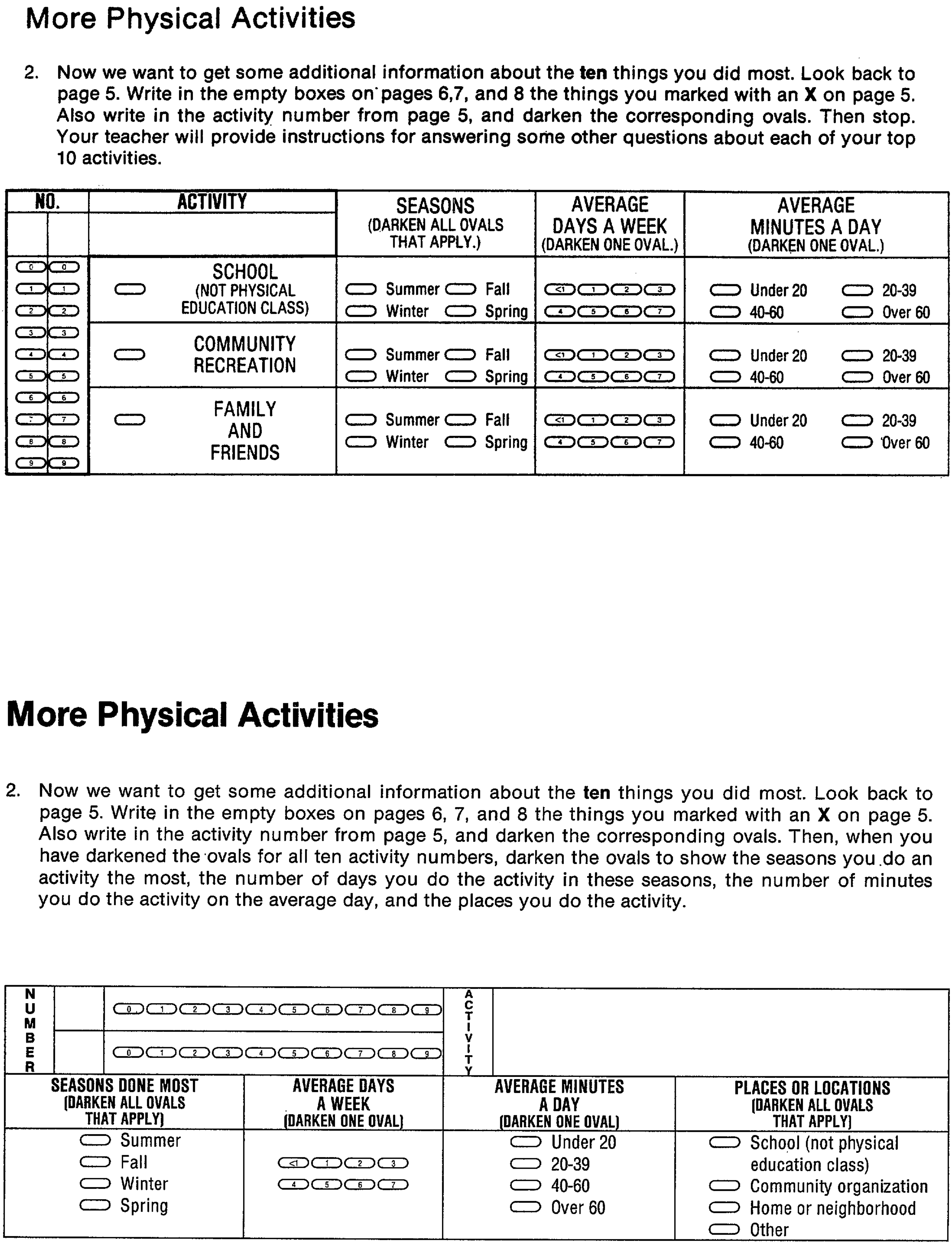
- Many teachers could not assume the data collector's role effectively, even with support. Therefore, it was decided that the role of the field staff would be expanded in the national study. The field staff member would take the skinfold, waist, and sit-and-reach test measurements; remain at a school for all data collection activities; and work closely with teachers in all other aspects of data collection.
- Two weeks were not enough time to test 400 students per primary sampling unit. The data collection period was expanded to 4 weeks.
- The testing protocols were not sufficiently clear to yield accurate and consistent performance on the tests. Protocols were revised on the basis of specific problems observed in the field.
- Many schools did not have needed equipment. Nearly three-quarters did not have a track and a third lacked a chinning bar or suitable substitute. Few schools had a measuring wheel or other calibrated instrument for measuring a running course on an open field. Most teachers would have been content to walk off a mile course or measure it with the school bus. As a result, field staff were equipped with measuring wheels; schools without a chinning bar were given one.
- The waist measurement proved to be much more sensitive than the skinfolds, which posed virtually no problems. The landmarks for the waist measurement were changed in an acceptable manner approved by the National Center for Health Statistics. The new definition of waist raised the measurement from the iliac crests to the midpoint between the iliac crests and the inferior border of the ribs.
- The survey was too complex and, therefore, was simplified significantly. Figure 2 illustrates the area in which the greatest single change occurred: Abandonment of efforts to collect detailed data *by source of exercise.* We arrived at this decision because it appeared that students were double-reporting activities, which exaggerated activity levels. In addition, two sets of survey instructions were developed for grades 5–8 and 9–12, with the primary differences between the two being redundancy of instruction and control over pacing. We concluded that separate *surveys* were not necessary, however.
- The pullup test provided poor differentiation in measurement of upper body strength-endurance, especially for females. The mean pullup score for females was less than 1.0, regardless of age. A large percentage of the boys in grades 5–8 could not perform one pullup. Therefore, the pullup test was replaced with the chinup test on the assumption that the easier hand position would produce fewer zero scores and provide greater differentiation in the measurement of strength for both sexes.

National administration of NCYFS I

NCYFS I was administered nationally to 8,800 5th through 12th graders in February through May 1984. The overall participation rate was 85.6 percent, but participation tended to decline with age, with the lowest participation (73.3 percent) among 12th grade females. Item nonresponse was a minor problem except for the mile walk-run, on which 11 percent of the participants refused, terminated, or could not be rescheduled.

The three major problems experienced in NCYFS I were motivating students and teachers to participate, maintaining sample integrity, and dealing with continuously inclement weather. Our greatest challenge was to maintain the motivation of students and teachers to complete their roles in the study. Securing consent was often difficult because (1) the teachers repeatedly forgot to distribute the consent forms, or (2) the students did not take them home (in some schools, taking home parental consent forms violated an unwritten rule), or (3) parents could not understand the form (e.g., non-English speaker, poor reader) or did not want the child to participate. Most teachers were highly supportive, but a small percentage of them were not for any of several reasons (e.g., felt forced into the study, objected to the loss of class time, disliked the school staff member assigned to coordinate the project).

There were also a variety of problems in maintaining the integrity of the sample. Teachers and students alike occasionally tried to "hide" students who would not perform well and to replace them with "ringers." It was sometimes difficult to obtain a complete and accurate class list against which to check the integrity of the sample. And, as in any field study, the vicissitudes of the weather wrought havoc with attendance rates, with conditions on a playing field or track and, as a result, with scheduling of outdoor events in several locations.



More Physical Activities

2. Now we want to get some additional information about the **ten** things you did most. Look back to page 5. Write in the empty boxes on pages 6,7, and 8 the things you marked with an **X** on page 5. Also write in the activity number from page 5, and darken the corresponding ovals. Then stop. Your teacher will provide instructions for answering some other questions about each of your top 10 activities.

NO.		ACTIVITY	SEASONS (DARKEN ALL OVALS THAT APPLY.)	AVERAGE DAYS A WEEK (DARKEN ONE OVAL.)	AVERAGE MINUTES A DAY (DARKEN ONE OVAL.)
○0 ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9	○0 ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9	○ SCHOOL (NOT PHYSICAL EDUCATION CLASS)	○ Summer ○ Fall ○ Winter ○ Spring	○<1 ○1 ○2 ○3 ○4 ○5 ○6 ○7	○ Under 20 ○ 20-39 ○ 40-60 ○ Over 60
		○ COMMUNITY RECREATION	○ Summer ○ Fall ○ Winter ○ Spring	○<1 ○1 ○2 ○3 ○4 ○5 ○6 ○7	○ Under 20 ○ 20-39 ○ 40-60 ○ Over 60
		○ FAMILY AND FRIENDS	○ Summer ○ Fall ○ Winter ○ Spring	○<1 ○1 ○2 ○3 ○4 ○5 ○6 ○7	○ Under 20 ○ 20-39 ○ 40-60 ○ Over 60

More Physical Activities

2. Now we want to get some additional information about the **ten** things you did most. Look back to page 5. Write in the empty boxes on pages 6, 7, and 8 the things you marked with an **X** on page 5. Also write in the activity number from page 5, and darken the corresponding ovals. Then, when you have darkened the ovals for all ten activity numbers, darken the ovals to show the seasons you do an activity the most, the number of days you do the activity in these seasons, the number of minutes you do the activity on the average day, and the places you do the activity.

NUMBER		○0 ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9	ACTIVITY	
		○0 ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9		

SEASONS DONE MOST (DARKEN ALL OVALS THAT APPLY)	AVERAGE DAYS A WEEK (DARKEN ONE OVAL)	AVERAGE MINUTES A DAY (DARKEN ONE OVAL)	PLACES OR LOCATIONS (DARKEN ALL OVALS THAT APPLY)
○ Summer ○ Fall ○ Winter ○ Spring	○<1 ○1 ○2 ○3 ○4 ○5 ○6 ○7	○ Under 20 ○ 20-39 ○ 40-60 ○ Over 60	○ School (not physical education class) ○ Community organization ○ Home or neighborhood ○ Other

Figure 2. Pilot test and final version of "Top 10" Section of National Children and Youth Fitness Study I physical activities survey.

Many NCYFS I problems have limited application to general population surveys that involve other age groups, utilize laboratory measures of fitness, and test subjects in a noninstitutional setting. However, any general population survey of fitness status and physical activity habits will, like NCYFS II, encounter continuing motivational problems. The single most pervasive NCYFS I problem was motivating students (and their parents) to agree to participation; getting students there for the survey and fitness tests (typically administered over three or more sessions); and convincing students to complete the entire survey accurately and to perform all of the fitness tests to the best of their abilities. Negative peer pressure often impeded the motivational task. Motivational problems were apparently not specific to students with low activity levels either. Field staff reported the impression that motivational problems disproportionately affected those at both the upper and lower ends of the able-bodied activity spectrum: the "cocky jock" and the physically underdeveloped child. Continuous efforts were made in NCYFS II to motivate students. The motivational task in any general population survey will be significant (a) if the test battery includes a direct test of VO_2max or other tests requiring maximal effort or (b) if subjects must routinely return for a second testing session.

History of NCYFS II

NCYFS II was initiated in response to widespread interest in extending the earlier study to younger children, ages 6–9 years (27). Doing so was justified by the midcourse review of the 1990 *Objectives for the Nation* (22, 23), general public sentiment, and Bennett's *First Lessons* (28), which accomplished at the elementary school level what the landmark *Nation at Risk* did for secondary education. Like NCYFS I, NCYFS II was designed as a 12-month study; in fact, after a lengthy delay in obtaining OMB clearance, the study was completed in 21 months. With the experience of the earlier study behind us, NCYFS II was approached with several clearly defined assumptions about what was feasible and necessary in studying the younger children (27).

Instrument design panels

NCYFS II was guided by two panels. Panel I was responsible for updating and adjusting the test battery for use with children 6–9 years. Panel II had the responsibility of suggesting design elements and data sources for a survey of physical activity habits.

Panel I: Physical fitness test battery

Panel I approached the task of designing a test battery for use with children 6–9 years with the explicit assumption that, if feasible and appropriate, NCYFS II should use the NCYFS I fitness test battery so measures would be comparable. However, the panel was encouraged to suggest changes in the test battery to accommodate the developmental capabilities of younger students or incorporate developments in the state of the art.

The panel agreed to maintain the sit-and-reach test (with apparatus); the 60-second, bent-knee situp test; and height, weight, and waist measurements. For each of the other three tests, the panel suggested changes or additions:

- The medial calf was added as a third skinfold site to provide an alternative to the subscapular, on the assumption that the medial calf would be a less "sensitive" site without sacrificing validity (29). By adding a site on the lower body, it was also presumed that the validity of the overall assessment of adiposity would be enhanced.
- The modified (Vermont) pullup was viewed as a welcome replacement for customary tests of upper body strength-endurance because it is a gender-free test on which nearly all youngsters earn a nonzero score (30, 31).
- For children under age 8, the distance for the walk-run was set at a half mile rather than a mile.

The panel's position on the distance run was that, although there is no physiological reason why 6- or 7-year-old children cannot run a mile, most children in the general population have not been trained to do so. The young child's inexperience with endurance activity is compounded by a relatively short attention span. Therefore, a mile walk-run would produce data of questionable validity. Moreover, if participation in fitness tests is intended to serve an educational purpose, a half mile distance on a walk-run would be sufficient to introduce children to the experience of paced, long-distance running.

Panel I reached three other conclusions regarding administration of the test battery. First, all tests should be administered one on one by a trained specialist, with teachers present only to serve as recorders and/or to help in maintaining discipline. Second, to maintain comparability with NCYFS I, all testers should be female. Third, some schools will lack the equipment needed to administer the tests; therefore, all needed equipment should be carried by the testers or, in the case of bulky items, donated to the schools.

Table 2 compares the fitness tests used in the NCYFS I pilot, the NCYFS I national study, and NCYFS II.

Panel II: Physical activities survey

Panel II faced the challenge of attempting to design a protocol for gathering data from an age group of children viewed by many survey researchers as not quite old enough to serve as reliable informants regarding their own behavior. Major studies that might serve as models excluded younger children. For example, the Canada Fitness Survey tested the fitness of children as young as age 7 but excluded those under age 10 from the physical activities survey. (See Stephens and Craig, chap-

Table 2. Comparison of fitness tests: NCYFS I pilot, NCYFS I national study, and NCYFS II

Dimension	NCYFS I pilot	NCYFS I national	NCYFS II
Cardiorespiratory endurance	Mile walk-run	Mile walk-run	Mile walk-run (age 8 and over) Half-mile walk-run (under age 8)
Muscular strength-endurance	Timed, bent-knee situps	Timed, bent-knee situps	Timed, bent-knee situps
	Pullups [1] Flexed-arm hang	Chinups	Modified (Vermont) pullups
Flexibility	Sit-and-reach test (with apparatus)	Sit-and-reach test (with apparatus)	Sit-and-reach test (with apparatus)
Body composition	Height Weight Waist (iliac crests) Triceps and subscapular skinfolds	Height Weight Waist (between iliac crests and inferior border of ribs) Triceps and subscapular skinfolds	Height Weight Waist (between iliac crests and inferior border of ribs) Triceps, subscapular, and medial calf skinfolds

[1] Dropped after preliminary field test.
Note: NCYFS = National Children and Youth Fitness Study.

ter 17.) The panel arrived at the following base assumptions to guide the survey design:

- Parents and teachers must be involved as informants because the children cannot be expected to speak for themselves; however, some information should also be gathered from the children.
- Measurement of participation in APA by children ages 6–9 years is not an important public health issue and, based on research by Baranowski and others in the Galveston group, appears to be a very rare event. However, parents and children are likely to exaggerate greatly the frequency of vigorous, continuous exercise. Therefore, no attempt should be made to measure participation in APA by children 6–9 years.
- To categorize children and estimate energy expenditure, it would seem useful to obtain data on the distribution of time spent in a day engaging in different levels of activity (i.e., sleep, standing and sitting, lower level movement, moderate to vigorous movement). Parents should be the primary source of data, but teachers and children can provide supplementary information on school-based activity time. Distinctions should be made between weekdays and weekend days.
- Data related to the determinants of both physical fitness and physical activity patterns of the child should also be collected.
- Too much emphasis has been placed on the impact of school programs on the fitness status of children. Children under age 10 are still strongly influenced by parental behavior and guidance. Therefore, data should be collected on the exercise habits and attitudes toward exercise of parents.
- The primary emphasis in collecting data from students should be measurement of attitudes toward physical education and exercise.

Testing of NCYFS II procedures

NCYFS II procedures underwent a more limited testing on the assumption that many aspects of the study would be very similar to NCYFS I. The primary interest in the tests of procedures were the new physical fitness tests and the physical activities surveys.

Testing of the physical activities survey

The physical activities surveys were preliminarily field tested in the Howard County, Maryland, public school system, at a Maryland private school, and at a YMCA winter camp, in accordance with OMB guidelines. The main objectives of these early tests were to assess the feasibility of the design and to assess respondent burden.

Parent survey. The parent survey did not fare well in the early tests. Figure 3 shows how we attempted to gather data on the distribution of the child's time at various activity levels. It became immediately apparent that most parents did not approach the task of retrospectively partitioning their child's time reflectively. It was a rare event for parent and child to sit together in responding to the survey to reach consensus on an appropriate response. From our perspective, it was very important that most parents seemed to "take the fifth" in describing the child's activities between the time of arriving home from school and sitting down to dinner. All that many parents could report with certainty was that the child spent a certain amount of time "playing outside"; working parents were hard pressed to tell us even that. It was not surprising, however, that parents nonetheless reported a very high amount of time spent in moderate to vigorous physical activity, defined by the first category on our survey. We decided, therefore, that we were unlikely to gather valid and reliable data of this type from parents. Ultimately, we settled on collecting data on the child's television-watching time per weekday and weekend day as one indicator of activity level.

Two other questions were meant to elicit information about the nature of the physical activities the child performed in two settings: (a) Organized sports or afterschool classes at school and (b) community organizations (checklist provided). It appeared that parents were double reporting activities, in some cases because a community organization used a school for an afterschool program. However, the overall frequency of participation in school programs was low. Therefore, we decided to obtain all information on school programs from teachers.

Several other aspects of the survey were changed. A question on which parents were asked to rate the child's overall activity level on a three-point scale did not produce adequate differentiation; therefore, it was changed to a five-point scale. A series of questions geared to tap parental attitudes toward exercise among children did not differentiate among respondents; therefore, the entire series was dropped. A question on factors affecting the child's ability to be more active physically did not yield useful information and was also dropped.

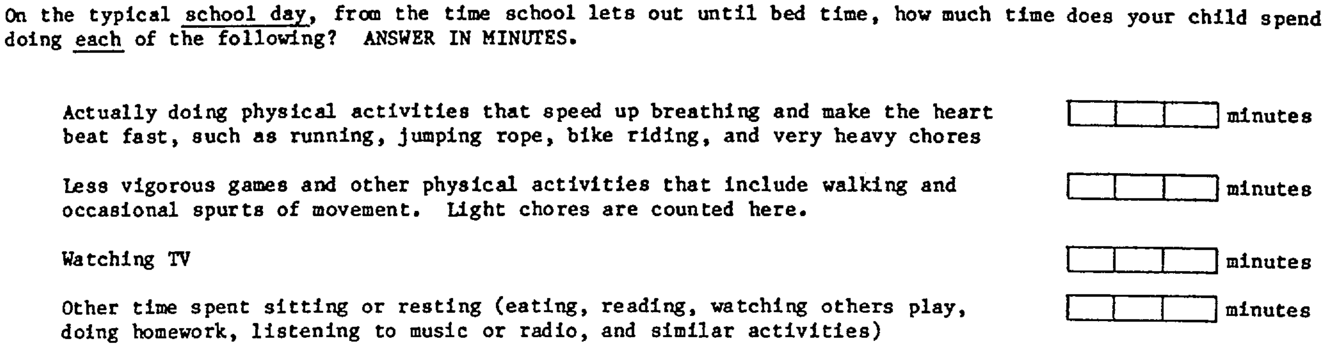
Some questions were retained as is or expanded. A question on the types of community organizations through which the child received exercise at least three times in the past year was retained without change. Similarly, a request to list the five activities in which the child participates most frequently through the listed community organizations was retained without change. In a question on parental participation in APA, the duration was changed from 20 minutes to 30 *continuous* minutes. A question on each parent's activity level was changed from a three- to a five-point scale to be comparable to the child rating question. A question was added on the frequency with which each parent exercises with the child for 20 minutes or more on the assumption that such behavior might represent a major determinant of exercise habits and fitness status.

Child survey. We quickly learned that our initial design for the child survey required more than the developmental level of a 6-year-old would permit. As a result, the child survey was tested over a series of five progressively more successful iterations.

A few questions were clearly not worth asking the children. One question required the child to partition activity time during recess. If parents could not do this, then it certainly made no sense to ask a 6-year-old, who is just obtaining a conventional sense of time, to do so. In two other questions, the child was asked whether he or she liked or disliked physical education class and how much. Because virtually all early elementary school children rate physical education as their favorite subject, this question did not differentiate among respondents and was dropped.

Our rudest shock in testing the child survey was how difficult it was to collect attitudinal data from the child. Our first attempt was to ask the children what they liked most about physical education class, followed by a question about what they disliked most. Students were to be probed for up to three responses per question. We expected the children to tell us that they liked physical education for such reasons as getting outdoors or out of class, learning new activities, having a chance to compete, or getting a good workout. We expected the children to tell us that they disliked physical education for such reasons as having to change clothes or shower, boredom, being in a gym that was too hot or cold, and dislike of getting sweaty. These response categories and others were precoded, but the child was not specifically probed for them. In fact, most children could find little that they disliked and, as far as likes were concerned, could do no more than report which activities they liked better than others.

After several iterations, we revised the child survey in a true-false format that allowed for a "don't know" response. This format forced the child to confront a variety of specific issues regarding likes and dislikes regarding exercise in general and physical education class in particular. In general, this format worked well, although some of the youngest children were prone to answer "make believe" instead of "false" or "not true." When we realized that many could not even pronounce the word "false," we switched to a yes-no format. The



On the typical school day, from the time school lets out until bed time, how much time does your child spend doing each of the following? ANSWER IN MINUTES.

Actually doing physical activities that speed up breathing and make the heart beat fast, such as running, jumping rope, bike riding, and very heavy chores ☐☐☐ minutes

Less vigorous games and other physical activities that include walking and occasional spurts of movement. Light chores are counted here. ☐☐☐ minutes

Watching TV ☐☐☐ minutes

Other time spent sitting or resting (eating, reading, watching others play, doing homework, listening to music or radio, and similar activities) ☐☐☐ minutes

Figure 3. Questions from original National Children and Youth Fitness Study II parent survey to estimate energy expenditure.

final version of the child survey was almost exclusively geared to assess attitudes.

Teacher survey. We encountered no major problems with the teacher survey during pilot testing but did receive numerous suggestions for refinements or additional questions. The final teacher survey included questions on number of days of physical education with a specialist per week, number of days of physical education with a classroom teacher per week, type of certification held by the teacher responsible for physical education, location in which physical education is held, whether children have the opportunity to change clothes for physical education or to shower afterward, types of sports teams available in the school, types of fitness tests administered, whether feedback on fitness test results are sent home to parents, percentage of children recognized for participation in fitness tests, number of and average length of recess periods, and reasons why elementary teachers do not use fitness tests. In addition, in the teacher survey, data were collected for each child on enrollment in physical education and an overall activity-level rating.

Testing of the physical fitness test battery

The physical fitness test battery for children 6–9 years was tested in a private school in Prince George's County, Maryland, and at a YMCA summer camp, in accordance with OBM requirements. The primary objective of the test was to determine whether test protocols required revision for use with this age group.

The test did not result in any major design changes. A nonreactive test sequence that also used time efficiently was developed. Preparatory activities for the distance run were added so that children had a better understanding of pacing. In addition, children were routinely asked to tie their shoes before the run. (Assistance was provided when needed.) Running numbers were put on the shirts of the children to facilitate keeping count of laps completed.

The single most significant change was in clarification of procedures for the modified pullup, which has seen only limited use outside Vermont. The student begins in a supine position and then lifts himself or herself as many times as possible to a level 6 inches below an adjustable bar set just beyond the fingertips. The back of the heels remain in contact with the floor; however, no other part of the body may touch down once the test has begun. The lower back should stay straight throughout the test. We quickly learned that some students did not grasp what was meant by keeping a "stiff body." As a result, children prematurely allowed their buttocks to rest on the floor, thus terminating the test. To correct this, we took two steps. The first was to modify instructions so students were told to "keep as straight as an arrow" and "be like a torpedo." In addition, just before the start, each student was asked to get into a reverse pushup position, providing them with the feeling of a straight back. These steps virtually eliminated problems with this test.

National administration of NCYFS II

NCYFS II was administered nationally to 4,678 first through fourth graders in September through November 1986. The student participation rate was 96.4 percent, with 94.6 percent of the parents of participating students completing a parent survey. Item nonresponse was not a major problem. However, the child survey was not implemented because OMB denied clearance on the grounds that collection of the data would serve no useful Federal purpose.

Some of the administrative problems in NCYFS II were reminiscent of the first study. Getting consent forms home to parents, signed, back to school, and into the hands of the teacher was an even greater challenge than in NCYFS I. Extensive calls home to parents overcame this problem. There were very few instances in which field staff had problems in maintaining the integrity of the sample. Perhaps because the participation rate was so high, teachers were not tempted to add volunteers to the study sample. Administration of all tests (except the distance run) in a single, one-on-one session overcame problems associated with session-to-session attrition. Maintaining motivation among the children 6–9 years was not a problem, at least not in the same way as it was in NCYFS I. It appears that young children are much more likely to stop at the first experience of discomfort than are children over age 10.

The most important lessons in testing the younger children are to: (a) Provide the children with prior experience in pacing; (b) allow the children to experience all of the tests at least once prior to the actual testing; (c) administer the distance run in very small groups, preferably when no other students are in the area; (d) take the time to dispel fears about the skinfold tests before initiating testing; (e) prepare the students to perform a maximal effort; and (f) make sure that all students bring proper clothing and footwear.

Design assumptions for general population surveys

To design and implement protocols for testing physical fitness and physical activity patterns in a general population survey, it is important that the specific objectives are first identified and clearly prioritized before extensive consideration is given to instrumentation. This is important because a study's purposes dictate the rigor with which each aspect must be conducted. For example, a study of the prevalence of certain health behaviors must receive a high participation rate and, therefore, cannot afford to include test items that will increase already marginal levels of case nonresponse or cause high item nonresponse. However, a study of the relationships between physical activity habits and concurrent or future health status does not pose rigorous participation requirements, unless biases in nonresponse truncate the range for either the dependent or independent variables. Similarly, in a study to design a system for assessing the health implications of physical activity habits, one may have the luxury of measuring the same dimensions in several different ways, but on a relatively small population that need not be selected in a nationally probabilistic manner. However, if a study's purpose is to generate incidence and prevalence estimates that may be generalized to the population, then a relatively large population, probabilistically drawn, is essential, and there is little or no room for apparently duplicative measures. In federally sponsored studies, data collection activities should also be streamlined to comply with OMB requirements.

The NCYFS experiences in instrument design, pilot testing, receipt of OMB clearance, and field implementation suggest that several assumptions should guide design of related general population surveys:

- Clearly state the purpose of the study as both descriptive and evaluative.
- Seek to incorporate the most valid and reliable data collection methods into the design within the constraints of cost, feasibility, and respondent burden.
- Consider adoption of data collection techniques that may not currently be cost feasible for widespread use if there is strong reason to believe that the technology will become generally available.
- Calculate respondent burden carefully and include no test items that appear duplicative.
- In weighing the merits of including an item, consider the educational value of public access to normative data on that item.
- Do not include an item that will create major problems in case or item nonresponse.
- Attempt to use the same or similar core tests and survey questions across the age span to permit cross-age comparisons and tracing of development. Ensure that all core measures are feasible and of generally comparable validity across the age span.
- Expand on the core tests and survey questions to examine aspects of physical activity or fitness of particular relevance to specific age groups (e.g., anaerobic power among young children; functional flexibility among the elderly).
- Try to be minimally intrusive. Minimize need for home preparation for the test-survey. Test the fitness of subjects through means that closely resemble day-to-day activities, if feasible.

Developing a physical fitness protocol

Physical fitness has been defined in many ways to express different philosophies of human growth and various situational needs. (See Wilmore, chapter 4, for further discussion of definitions of fitness.) This chapter is limited to a discussion of health-related fitness, which appears appropriate in a general population survey. Pate defined health-related fitness as those fitness components directly related to disease prevention and/or health promotion (5). AAHPERD (26) has identified the following components of health-related fitness:

- *Cardiorespiratory endurance*—The ability to perform high-intensity activity for a prolonged period of time without experiencing fatigue or exhaustion.
- *Muscular strength*—The ability of the muscle to exert or apply force against an unyielding resistance in a single, maximal exertion.
- *Muscular endurance*—Ability of the muscle to sustain repeated, high-intensity contractions or to sustain a fixed contraction for an extended time period.
- *Flexibility*—Range of motion around a joint, as determined by the interrelationships among muscles, tendons, ligaments, and the joint.
- *Body composition*—Body fatness relative to lean body mass.

Each of these components of fitness was measured under field conditions in NCYFS I and II. We strongly urge that field measures be considered for general population surveys if they offer reasonably valid and reliable data and do not pose an unacceptable risk of injury. A fundamental issue affecting test selection, however, is the relative importance of generating clinical data that can be used in defining the health implications of habitual physical activity versus fitness norms that can be used widely in a variety of settings and have significant educational impact. Rigor tends to be exchanged for ease of use and educational value. NCYFS I and II measures and some alternatives are considered here. Our discussion assumes that most measures will be used across a wide age span, starting at age 5 or 6 and including the elderly.

Cardiorespiratory endurance

In NCYFS I and II, norms related to cardiorespiratory endurance were established using an all-out, indirect measure of VO_2max, the 1-mile walk-run (ages 8–18) and the half-mile walk-run (under age 8). Distance runs can be useful tests and readily lend themselves to testing large groups of children and some other able-bodied, moderately trained populations. However, the validity of the distance run as an indirect measure of VO_2max depends strongly on pacing, motivation, efficiency of running style, weather conditions, environmental temperature, and the general fitness of the subject (32). Correlations between estimated VO_2max from a distance run and measured VO_2max have varied from 0.26 to 0.92 (33). In NCYFS I, the mile walk-run was administered over 900 times to groups averaging 10 students with few administrative problems and a high degree of standardization. No attempt was made to extrapolate from running performance to VO_2max. Though appropriate for NCYFS I, a distance run is inappropriate for a general population survey of a broader age span for several reasons:

- A high degree of cooperation and motivation are needed for valid results.
- Careful medical screening is necessary.

- Risk of injury is relatively high.
- Changes in weather and problems in access to a standardized running course complicate scheduling and other aspects of logistics.
- The concern of subjects over having to run a mile may increase case nonresponse and produce a high level of item nonresponse and terminations.

To achieve an acceptable participation rate and maintain tolerable risk levels, it may not be appropriate to use all-out tests of $\dot{V}O_2$max in general population surveys across the lifespan. Aside from the need for more careful medical screening, the principal argument against an all-out test is that it may drive down the participation rate to an unacceptably low level. In NCYFS I, despite an overall participation rate of 85.6 percent for ages 10–18, only 60 percent of 18-year-old females participated in the mile walk-run. Other authors (Montoye, chapter 15; Stephens and Craig, chapter 17) report similar participation problems with females over age 14. McDowell (chapter 2) stresses the need to maintain the current, hard-won participation rates in major general population surveys.

As an alternative to a maximal test, a submaximal indirect measure of $\dot{V}O_2$max is recommended. A treadmill is recommended over a cycle ergometer or a step test for a variety of reasons. Treadmill walking can be learned rather easily and, because walking is a natural movement, the treadmill will probably produce fewer refusals than the cycle ergometer or step test. The treadmill is also the method of choice among physicians (34) and can readily produce valid measurements among almost all age groups. Because the subject carries his or her own weight, no weight adjustments are needed, unlike the cycle. The treadmill also accommodates heavy subjects more readily than a bicycle. Unlike the cycle ergometer or step test, there is no cadence or "rate" problem with the treadmill. Local muscle fatigue is less a problem with the treadmill than with the cycle because large muscle groups are involved. Exercise intensities can be fixed and a steady state can be reached quickly on a treadmill provided the grade increments are small.

The treadmill is not without its problems, however. There is a definite risk of slipping and sliding off the treadmill, especially for young children, the elderly, and others with coordination problems. The treadmill is physically much bulkier than the stepping device or the cycle. If portability is considered desirable, as would be the case in a household survey, then a treadmill would not be the first choice. Aside from sheer bulkiness, getting adequate electrical power could also present a problem. Risk of slippage can be minimized by proper footwear, which could be loaned to subjects. The most serious problem with the treadmill is that arm movements, heel strike, and noise all potentially interfere with measurements. Pilot tests should be performed to refine treadmill safety and measurement procedures as well as to assess the differential effects on participation rates of the treadmill, cycle ergometer, and step tests as indirect measures of $\dot{V}O_2$max. Detailed testing protocols are discussed in other chapters.

It is not clear, however, to what extent the experience of NCYFS I with low participation rates among older school-age females and the experiences in other major studies apply to a comprehensive assessment of health status. In NCYFS I, students were selected for participation as members of an intact class section, often a physical education class. They were tested in groups, wearing gym clothes, in a familiar school environment, usually the gymnasium and athletic field, with the assistance of familiar and not always beloved individuals, the school's physical education teachers.

NCYFS I was marketed to the students as a test of health. The competitive aspects of physical performance were downplayed. Nonetheless, students tended to associate the study strongly with either competitive sports or physical education class, which many of the older students shunned, rather than with an assessment of health status. A variety of contextual factors often contributed to a peer group dynamic that could drag down participation rates. The sheer competitiveness of the situation may have scared off many underconfident students. Other studies may succeed in achieving substantially higher participation rates among adolescent and adult subjects, especially females, if subjects are selected and tested individually in a noncompetitive medical environment, by medical personnel and technicians, with complete confidence in their anonymity. Pilot tests may permit estimation of the differential in participation rates contingent on variations in the testing protocol and other contextual factors.

Muscular strength and endurance

Muscular strength and endurance are both complex factors but are commonly measured through the same test items under field conditions. The AAHPERD Health Related Physical Fitness Test includes timed, bent-knee situps as its one measure of muscular strength-endurance (6). In our studies, muscle strength-endurance was measured by timed, bent-knee situps as well as by chinups (NCYFS I) and modified pullups (NCYFS II).

The NCYFS I panels had difficulty in deciding whether to merge strength and endurance measurements and, if not, how to differentiate them. Muscular strength was the one area in which the NCYFS I panel of school testing experts (Panel II) rejected the recommendation of the first panel regarding practical field measures. Panel I argued strongly to combine muscular strength and endurance tests under nonlaboratory conditions but still suggested separate tests. To measure muscular strength under nonlaboratory conditions, Panel I recommended a one-repetition maximum test (1–RM) bench press and biceps curl. To measure muscular endurance, it recommended two Level 2 procedures involving a bench press, biceps curl, leg press, and knee extension. The first procedure was to measure the number of repetitions that an individual could perform at a fixed weight, set individually on the basis of strength, in a fixed time. (The time element made this more a measure of power than endurance.) The second procedure measured the amount of weight that the individual could handle in 15 RM's. Ultimately, Panel I tended to endorse traditional field measures combining muscular strength and endurance but gave some support to a distinct measure of muscular strength (25). Designing a fitness test battery for a general population survey requires careful evaluation of the need to distinguish muscular strength from muscular endurance and to provide specificity is assessment of muscle function (35). The measures of muscle function from NCYFS I and II are primarily measures of dynamic muscular endurance, as are most field tests of muscle function; however, for a subject who must perform a maximal exertion to complete a single repetition, as is often the case for the traditional pullup or chinup, the test is more appropriately considered a test of strength. In contrast, traditional laboratory tests of muscle function have focused on strength rather than endurance. Recent developments in laboratory instrumentation, such as the Cybex II, permit controlled assessment of the various components of muscle function with a high degree of specificity. Even if a Cybex II or a similar system is not available, it is important to recognize that muscular strength has a high degree of specificity. It would be highly desirable to measure at least three areas of strength—upper body, lower body, and trunk—in a general population survey. It is not known whether muscular endurance has a comparable degree of specificity.

The idea of field measures of muscular strength and endurance should be entertained because of the educational value of readily usable norms. We suggest that strength be measured by 1-RM bench press, leg press, and biceps curl. For endurance, we recommend including tests of abdominal and upper body endurance. The timed, bentknee situp test can readily be standardized and, although not an optimal means of isolating the abdominals, would appear reasonable. Either the Baumgartner pullup or the modified (Vermont) pullup used in NCYFS II could be used to test upper body endurance. The Baumgartner pullup (36, 37) places the subject on a scooter board mounted on an incline. (See figure 4.) In a prone position, the subject pulls the chin up over a bar, sliding the scooter board on rails. There is evidence to support the reliability and a validity of the test; however, the bulkiness of the apparatus has limited its use under field conditions. In contrast, the modified (Vermont) pullup (31) places the subject in a supine position and requires the subject to pull his or her chin to a point 6 inches below a bar, which is set just beyond the fingertips. Heels remain in contact with the ground and the body remains straight throughout the test. (See figures 5 and 6.) Reliability and validity studies on the modified (Vermont) pullup are currently under way. The advantage of this test is that virtually all subjects achieve a nonzero score. Alternatively, endurance could be measured by the number of repetitions that an individual can perform at a fixed weight, determined as a percentage of his 1-RM, until exhaustion. This procedure represents a modification of a proposal by Panel I in NCYFS I. Each of these procedures should be pilot tested, primarily on

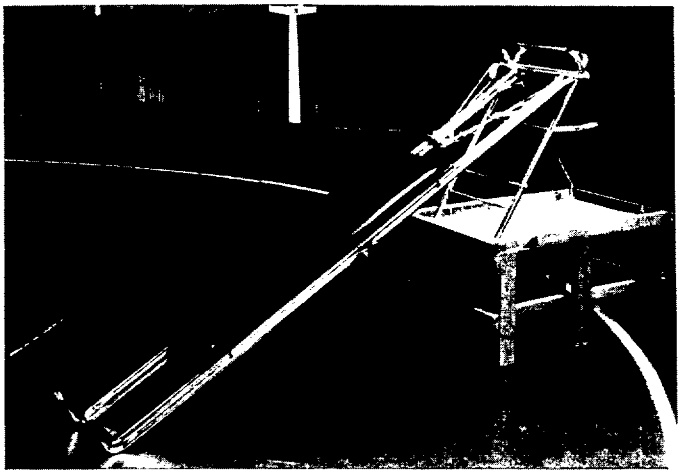


Figure 4. Baumgartner pullup.

Figure 5. Down position for modified (Vermont) pullup.

Figure 6. Up position for modified (Vermont) pullup.

younger children, women, and the elderly, to assess their impact on case and item nonresponse, to educate as to safety implications, and to determine if the protocols will require adaptation.

Flexibility

Flexibility has come to be regarded as a measure of health-related fitness primarily because of the association between lower back pain and inflexibility of the lower back and hamstring muscles. The most widely used field test of flexibility—the Wells and Dillon sit-and-reach test—is an element of the AAHPERD Health Related Physical Fitness Test (6) and was used in NCYFS I and II. This test offers a crude index of the flexibility of these muscle groups. It is not clear, however, how flexibility should be measured in a general population survey across the life cycle. There are questions regarding the validity of the sit-and-reach test because leg and arm length are confounded with flexibility. Moreover, flexibility, like muscular strength, has a high degree of specificity. Many students participating in NCYFS I questioned how their sit-and-reach score could be used as a general indicator of flexibility, especially because the score is not adjusted for arm and leg length. Neither the sit-and-reach test nor any other single test of flexibility yields an overall flexibility index, nor is it known what aspects of flexibility are health related and warrant inclusion in a test battery.

Flexibility should definitely be measured in a general population survey of health-related fitness status, at least with a variation of the sit-and-reach test. If an agreement can be reached regarding the critical joints or movements required for day-to-day living (e.g., neck, arm-shoulder girdle, lower back, ankles, wrists), then the flexibility of these joints should also be measured (e.g., with the Leighton Flexometer). For example, lateral flexibility of the neck might be considered a necessary function because of its safety implications for driving a car. If agreement cannot be reached on a set of three to six critical movements, the sit-and-reach test should be administered using a modified scoring system or apparatus. Procedures have been developed to adjust raw scores for differences in leg and arm length (38). In addition, a sit-and-reach apparatus that adjusts for arm and leg length has recently been marketed commercially (39). To limit test administration time, it may be appropriate to administer the full set of flexibility measures only to adults, perhaps only those over age 50.

Body composition

Body composition measurement is an essential component of a health-related physical fitness battery. Although many techniques for assessment of body composition have been used, most of them are not practical for a general population survey. The criterion method of underwater weighing requires bulky equipment, is time consuming, and imposes special requirements on subjects (e.g., arriving in a postabsorptive state). The prospect of being weighed underwater and these special requirements would jeopardize participation rates, especially among women, young children, the elderly, and those with a fear of water. It is difficult to conceptualize how the underwater weighing test, which typically requires approximately 10 trials, could be incorporated into a relatively brief test battery that will include a test of VO_2max, several anthropometric measurements, and possibily other tests of strength and flexibility. Respondent burden has to be a paramount concern. Moreover, although underwater weighing is still considered the best method for determining body composition, the validity of this technique has recently been criticized, especially for certain populations.

A more prudent course would be to expand the set of anthropometric measurements taken through a general population survey to include additional skinfolds and girths. We recommend that height and weight be taken to permit computation of body mass index. In addition, waist and hip girths should be measured. Although the validity of the skinfold technique has been criticized for certain populations, special equations have been developed to translate skinfolds into estimates of lean body mass. AAHPERD has released formulas for application to triceps and subscapula scores for young children (26). We recommend that at least three skinfold sites be measured: Triceps, subscapula, and medial calf. If feasible, we also suggest that the following sites be measured: Chest, suprailiac, abdomen, and anterior thigh.

The medial calf skinfold, although not an element in most formulas, is important because the educational community has adopted it as an alternative to the subscapula. The rationale for this shift to the medial calf is the hope that school administrators, physical education teachers, and parents will be more receptive to the test if a less sensitive and more accessible site than

Table 3. Proposed fitness tests

Dimension	Test
Cardiorespiratory endurance	Submaximal treadmill test
Muscular strength	1–RM [1] bench press, leg press, and biceps curl
Muscular endurance	60-second, bent-knee situps Leg press, maximum repetitions on weight set as percentage of 1–RM [1] Baumgartner or modified (Vermont) pullup
Flexibility	Sit-and-reach test with adjustments for arm and leg length Tests of flexibility of additional joints for adults over age 50
Body composition	Height and weight Waist and hip girths Triceps, subscapular, medial calf, chest, suprailiac, abdomen, and anterior thigh skinfolds

[1] 1-repetition maximum.

the subscapula is used. Lohman has documented the validity of replacing the subscapula with the medial calf (29).

Recommendations

Table 3 summarizes our proposed health-related physical fitness measures in each of five dimensions for a general population survey. Once again, we have assumed that the survey will be administered to subjects representing different points of the life cycle, starting at approximately age 5.

Developing a physical activities survey

The assessment of physical activity habits in conjunction with a general population survey of fitness status serves a range of research objectives. The most fundamental objective is to define in detail the relationship between historical and concurrent physical activity habits and health status. Much epidemiological research has focused on the relationship between regular, vigorous physical activity and reduction of CHD risk. Relatively little attention has been directed to the short- and long-term effects of aerobic activity on other health behaviors (e.g., smoking cessation or control of mild anxiety) or to the health implications of other properties of physical activity (e.g., the effects of weight-bearing activity on osteoporosis). Inadequate attention has also been directed to the determinants of physical activity or to the adverse health effects of physical activity, especially on the musculoskeletal system. In this section, we are primarily concerned with development of a physical activities survey for children and adolescents 5–18 years.

Topics for coverage in a physical activities survey

Several realms of physical activity compete for coverage in a comprehensive survey. An analysis of the health implications of physical activity habits requires a detailed description of work, school, household, and leisure-time physical activity. Information is needed on the frequency, duration, intensity, and seasonality of each specific leisure-time activity. Comparable information is needed for physical activity in school (in physical education class, during recess, on teams), in the workplace, and around the house. Sufficient information must be gathered not only to calculate a subject's energy expenditure but also to define other properties of activity. Different data collection strategies may be needed to assess physical activity in school, at work, at home, and during leisure hours. The subject's knowledge of the health implications of physical activity may also be an appropriate subject for a survey. The two most important topics to query for the survey are the characteristics of physical activity required for a training effect and basic energy balance concepts as related to weight control. Other potential topics include the definition of "lifetime" physical activity and the relationship between specific aspects of physical activity and disease prevention (e.g., the effects of regular walking on prevention of osteoporosis). For purposes of targeting organizational change, the data collected in the survey should be differentiated as to the contributions of school, the workplace, and community organizations in providing physical activity.

Problems in constructing a physical activities survey for children

Because most research on physical activity has concentrated on adult white males, relatively little is known about the construction of physical activities surveys for children and youth. The design task is quite complex for the reasons discussed next.

Developmental changes in physical activity patterns

One source of complexity is the rapid behavioral development through which most children pass in the brief 18 or so years from birth to adulthood. During the preschool years, the child has a low attention span, much play is solitary (even with other children), and activity gears shift rapidly. Enjoying play at its perfection, the child runs in spurts, stopping to smell a flower, seek solace, or encounter a coinhabitant of his or her space. In the early school years, physical activity begins to acquire more of a structure. The attention span is longer and the child has learned a lot about playing with other children. Spontaneous physical activity begins to give way to low-organized games. Such games offer limits within which much spontaneous activity continues to occur as the child also develops coordination and social skills. Gradually, low-organized games are forced to share the stage with competitive team sports requiring a long attention span, greater physical prowess, coordinated interactions with other children, and a sense of dedication and commitment. In time, the child outgrows teams, finds it more difficult to round up a street team, chooses the intimacy of playing with one other person, opts for a less injury-prone activity, or reclaims the pleasures of solitary physical activity. Team competitive sports give way to lifetime physical activity. The typical physical activity survey has been limited to competitive team sports and some forms of lifetime physical activity.

Individual differences in physical activity development

The task of survey design is made more complex by individual differences in the rate at which children move through these stages of physical activity. In any group of early schoolers, there will be many children who have already joined organized teams to play soccer, baseball, or other team sports. And in any group of high school students, children may be found who have never played a team sport and whose physical activity resembles the low-organized games of early childhood. Children progress through stages at varying rates. Occasionally, they will seem to skip a stage or even to regress.

Variations in intensity of activities

Other factors also make a physical activities survey for children hard to design. One problem is that two people can call an activity by the same name but play it very differently. In a typical high school physical education class, some students will be found playing an intense game of full-court basketball while others languish under a hoop, not even breaking a sweat. The typical 7-year-old backyard football player does not play the same game as the typical 17-year-old member of a varsity team. Measures of intensity and of the continuity of activity flow are needed.

Differential capacity to perceive intensity

A related problem, however, is that many children are scarcely aware of the behavioral signs of exercise stress. In part, this is because young children have a higher threshold for sweating and sweat at a lower rate than adults (40). Therefore, sweating should be deemphasized as an indicator of exercise stress in favor of rapid heart beat and breathing hard, as in the 1985 NHIS Supplement.

Identifying reliable informants

Another problem in designing a physical activities survey for children is deciding on the most appropriate informant. It seems safe to assume that children and adolescents 10–18 years (5th through 12th graders) can report their own data. In NCYFS I, 10- to 18-year-olds all seemed able to handle rather complex questions about the duration, frequency, seasonality, and sources of physical activity; very little was asked about intensity, unfortunately. Indirect validation of physical activity data against measured fitness confirmed the ability of 10- to 18-year-olds to report their own activity habits (19, 20). Similarly, for years, the National Assessment of Educational Progress has directed a wide range of complex questions, some of them requiring time estimates, to students in fourth grade through high school and it has been concluded that, if there was any distortion, it was constant across the grade span. In time-use studies by the University of Michigan, detailed complex reports have been gathered through time diaries from children and adolescents 9–18 years without parental involvement (41). The Michigan researchers support the reliability of the children's reported time use primarily on the grounds that their time estimates conform with intuitions about how children use their time. For example, on weekdays, adults report a greater variety of activities than children; on weekends, the reverse is true (41).

The major problem in identifying reliable informants for physical activity data applies to children under age 9 or 10. As noted earlier, in the Canada Fitness Survey, the fitness status of children as young as age 7 was tested, but the physical activities survey was not administered to those under age 10. It has occasionally been suggested that data on the physical activity habits of younger children will be improved if their responses are supplemented with those of parents and/or teachers; we proceeded on this assumption in NCYFS II. The evidence, however, is limited and mixed. The Michigan researchers involved parents in completing time-use diaries with children aged 3–8. Parents reported directly for children aged 3–5 years. However, children aged 6–8 years reported their own information to an interviewer, usually in the presence of a mother. In 8 percent of the interviews, the mother provided no information; in 30 percent, she gave "some"; and in 46 percent, she gave most or all of the information. The Michigan researchers concluded that this arrangement provided convergent validation of time use by children 6–8 years. For children 3–5 years, they concluded that the time-use diaries were not worthwhile because, with the large amounts of unaccountable time, little useful information was provided about the child's activities (41).

Baranowski (42) and others in the Galveston group tested various formats for completion of time diaries, limited in scope to bouts of APA lasting at least 20 minutes. Their studies have confirmed that APA is a rare event among third through sixth graders which may be reliably measured through daily self-report forms, especially when the form segments the day into time periods. However, parental assistance produced no effect (42). Most parents were not familiar with their children's exercise behavior and knew little about exercise behavior at school. Consistent with our experience in the NCYFS II pilot, parents did not assist the child in organizing memory and response to fit the form. However, we do not conclude from the Galveston studies that parents do not play a useful role in reporting certain aspects of the child's activity. It may be that bouts of APA among third through sixth graders are so rare, as Baranowski's research indicates (42, 43), that the child needs little help in organizing information. However, NCYFS II results suggest that fathers are much more frequently involved in their children's exercise activities (17) than Baranowski's research, which has focused on lower socioeconomic populations, would indicate (43). All factors considered, it may be prudent to limit the parent's role to assistance in describing behavior that occurs within and around the home, primarily for children under age 9. However, the younger the child is, the more diffuse and spontaneous are the activity patterns and the less information may be obtained from parental reports.

In general, it would appear that the less structured and continuous the child's activities, the more difficult they are to describe. One of the reasons 18-year-olds had an easier time than 10-year-olds in filling out the NCYFS I survey form was that the activity patterns of the older children were less complex. The average 18-year-old reported performing 15 different physical activities at least three times outside of physical education class over the past year; the average 10-year-old reported nearly twice as many, 27. As the child grows older and relinquishes the low-organized games of early school years, the variation in activity patterns decreases. For the older preschooler, spontaneous play is just beginning to acquire the structure of low-organized games. For the early schooler, activity still vacillates between low-organized games and spontaneous play. Therefore, it is extremely difficult for an observer to capture in detail the flow of physical activity of the typical young child. Once the child enters school, as far as the parent is concerned, the child spends 7 hours a day out of sight and out of mind. The observer of a young child's behavior may report casual impressions of the intensity of the activity, but this may not be congruent with the child's own perceptions of exercise stress. Structured daily self-reports for children over age 5 appear to offer the most reliable and valid measures of physical activity among the cost-feasible options. It may also be possible to train parents to assist in these reports. In a general population survey, such self-reports would best be collected on a randomly selected set of days over a period of a year. The cost, burden, and intrusiveness of such data collection methods could be a major issue in obtaining OMB clearance of a federally sponsored survey.

Variations in contexts or sources of activity

Estimating the physical activity of children and youth is further complicated by the many contexts in which physical activity is performed. Half the child's waking hours are spent at or in transit to and from school. Information is needed about how the child gets to and from school and about physical activity received through physical education class, at recess, on the playground before and after school, and on school teams (intramurals, interscholastic sports, and clubs). The school's physical education program will cover many activities over the school year, with activities often occurring out of synchrony with the seasons. Outside of school, the child finds opportunities for physical activity with his family, friends, in "pickup" games, or alone. Much of this activity happens outside of any organized context; however, a wide range of community organizations—including places of worship, scouting groups, parks and recreation programs, Y's, local teams, and private health clubs—help to provide a structure for physical activity. In addition, children help with housework or even assume regular jobs, on a full- or part-time basis, as they approach adulthood. Some children help with a family business until they leave home. During the summer, the school-year pattern of activity tends to break down as the child takes a summer job, goes away to camp, joins a special summer league, or tries to keep occupied at home.

Content of NCYFS I and II

In collecting physical activity data on children and adolescents 6–18 years in NCYFS I and II, development of baseline data related to the PFE objectives discussed earlier in this chapter was sought. Because of the wide age span of the respondents, a long list of 86 activities, including low-organized games, was used in NCYFS I. The same survey document was used for all grades, but two versions of the survey instructions were developed, one for elementary grades (5–8) and the other for secondary grades (9–12). The instructions clearly distinguished the sections that the survey administrator was to read silently from those to be read out loud to the students. The following information was collected:

- *Physical education (PE) class*—Enrollment in PE this term, last term; number of days per week; class length; whether time and place are provided to change clothes, take showers; activities participated in during the past calendar year; number of weeks per activity.
- *Community organizations*—Community organizations (checked off) through which the student received physical activity over the past calendar year.
- *Activities outside PE class*—Activities participated in at least 3 times over the past calendar year outside of PE class; for the 10 most frequent (the top 10), the seasons done the most and, during those seasons, the average weekly frequency, average duration, and the sources (school but outside of PE class, community organization, and home-neighborhood); also, the number of days per week, by season, that sweating and breathing hard was experienced for 20 minutes or more during exercise.

In NCYFS II, two survey instruments were used to gather data from parents and teachers. Data regarding the physical activity habits of the individual child were limited to activity level rating (given by parent and teacher), enrollment in physical education, community organizations through which the child obtained physical activity in the past year, "top five" activities through these community organizations, number of hours of television watching on weekdays and weekend days, and the number of days per week that mother versus father exercises with the child for 20 minutes or more. Additional information was collected about parental exercise habits and about the school program.

Issues not addressed in NCYFS I

NCYFS I generated a rich data base on the habitual physical activity of children and adolescents 10–18 years but failed to cover certain important areas adequately. On the plus side, the study produced a detailed profile of physical education programs and of activity patterns outside of the school physical education program. Sex differences and changes in activity patterns over the age span were described. Overall levels of participation in APA were estimated on the basis of activity reports and perceptions of exercise stress. Seasonality was also examined in relation to overall activity levels and receipt of APA. Additional analysis of the NCYFS I data will help to define seasonality in the nature of physical activity and assess relationships between activity types and sources. Further examination of the survey data in conjunction with the fitness data will help to refine relationships between activity patterns and measured fitness status and may allow the identification of criterion-referenced cutoff points for fitness norms. However, NCYFS I had several major weaknesses that future studies of physical activity among children should correct.

Direct reporting of intensity

Intensity data were intentionally not collected in NCYFS I by activity. Instead, a single summative question was asked about the frequency with which the child perceived certain behavioral signs of APA, as described earlier. This type of summative question should be retained, but "heart beating fast" should replace "sweating" for the reasons explained earlier. In addition, intensity data should be collected by activity, as in the 1985 NHIS supplement on physical activity. A global measure of the intensity of the physical education class activities should be collected.

The summative measures of intensity serve as a validation of the activity-specific estimates, which can be assumed to exaggerate actual behavior somewhat. In NCYFS I, three estimates of participation in year-round APA were produced. The first estimate was based on the detailed reports of "top 10" activities. Mean METS values from the Canada Fitness Survey (44) were adopted as "surrogate" intensity indicators. (A few values had to be guessed at.) An activity with a mean METS value of 6.5 or greater was regarded as a high-intensity activity that, if performed with sufficient frequency and duration, would yield an appropriate effect. The second estimate of APA was derived from the summative question on perceived exercise stress. Based on activity reports, 59 percent received year-round APA (i.e., participated in at least one high-intensity activity at least 3 days per week for at least 20 minutes in each season). The global self-perceptions of exercise stress indicate that only 41 percent participate in APA year round. Our third and most conservative estimate reflects the conjunction of these two sets, which yields 31 percent participation in APA on a year-round basis. In contrast, Baranowski reports 10.4 percent of third through sixth graders receiving daily moderate-to-vigorous physical activity using a strict definition and 47.9 percent using a less rigorous definition (43).

Of equal interest were the correlations between APA ratings based on activity reports and self-perceptions of exercise stress, which were all in the low-to-moderate range. These findings may be interpreted in several ways. That the higher estimate is derived from the activity reports may be seen as support for the belief that individuals exaggerate their physical activities. It could also mean that some students who participate in APA do so without being conscious of sweating and breathing hard. The less than optimal congruency between the two sets of estimates clearly indicates that two different behaviors have been measured. One obvious explanation of the incongruency is that the surrogate measures of intensity were inadequate. In other words, some students performed activities with a high mean intensity rating at lower intensity levels; other students performed activities with low and moderate ratings at higher intensity levels. Some students who participated in APA may not have perceived sweating because of the season, their high sweating thresholds, or the nature of the activity (e.g., swimming).

Energy expenditure was estimated in NYCFS I as a weekly estimate by season. These estimates were incorporated into the activity-fitness analyses. The lack of direct estimates of intensity was a clear deficiency in the analysis. A second deficiency was failure to adjust estimates of energy expenditure for body weight. It is essential that other general population surveys of physical activity include activity-specific measures of intensity as well as summative measures of self-perceived exercise stress. Third, if intensities are imputed by using mean METS values, it is important that certain extremely high-intensity activities, such as sprinting, not be included, even if reported at an appropriate frequency and duration.

Seasonal differentiation

NCYFS I had an apparent advantage over the 1985 NHIS Supplement in that NCYFS I included all four seasons, whereas only the most recent 2 weeks were examined in NHIS. This advantage was not entirely real for two reasons. First, many children presumably had to strain recall to report activities by season for an entire year. Second, NCYFS I respondents did not have the option of reporting different frequencies, durations, or sources for a given physical activity by season. For those seasons in which a respondent performed an activity the most, and only for those seasons, the respondent reported a uniform frequency, duration, and source. Thus, a student who played basketball for 2 hours a day 7 days a week in winter, for only 1 hour a day 5 days a week in the spring, once a week all day in the fall, and never in the summer had two reasonable options: average winter and spring or report winter only.

A better approach would be to collect data by season during the season in which the activity occurs. This can be accomplished by readministration of the physical activities survey or a portion of it at selected intervals during the year following its initial administration. Such an approach is the ideal for future general population surveys of either the same sample or a rotating sample.

Work behavior

In NCYFS I, no information was collected on household chores or about part-time, full-time, and summer jobs. This was an obvious deficiency in estimating energy expenditure levels and in estimating the percentage of the population participating in APA. One of the problems in examining work behavior among children and youth is that the behavior tends to be short lived. Even teenagers with jobs tend to change jobs fairly frequently. Collection of work data from youth requires closer attention.

Knowledge levels

In NCYFS I and II, no attempt was made to determine knowledge of important facts about habitual physical activity. The PFE objectives require estimation of knowledge among adults of the frequency, duration, and intensity of activity needed to produce a training effect. In the 1985 NHIS Supplement, this issue is addressed in a three-part question. Similar questions should be targeted to children. In addition, questions should be formulated regarding energy balance concepts in weight control.

Physical education class activity

NCYFS I provided a detailed profile of physical education class activities. Estimates were established regarding the proportion of physical education class time dedicated to lifetime physical activity, which, although not required by the 1990 PFE objectives, were consistent with the narrative supporting them. A few additional issues should also be examined:

- *Intensity of PE activities*—Students should be asked to rate the intensity of PE class activities. One method would be to ask students, "Out of 10 typical physical education classes, in how many would you experience no increase in breathing or heart beat? A small increase? A moderate increase? A large increase?"
- *Fitness testing*—The PFE objectives require estimates of the percentage of students who participate in school-based fitness testing. This issue was not included in NCYFS I because the student's participation in the study intruded into the organizational behavior of interest. It is recommended that general population studies of physical activity and fitness status in school settings obtain from the school information about the nature of the school's fitness testing program, the percentage of students recognized for participation in fitness testing, and whether fitness test results are routinely sent home to parents as a "report card."
- *Lifetime physical activity*—Several definitions of lifetime physical activity were considered in NCYFS I. If assessment of the amount of time spent on lifetime activities, either in school or in the community, is sought in future general population surveys, then all the potential definitions of "lifetime" should be reexamined before calculating the lifetime portion of the total physical education program. In NCYFS I, any activity that may typically be done readily by one or two people was regarded as "lifetime". Thus, basketball would not be lifetime, because it requires a team, but tennis would be. There were a few exceptions to the rule. All forms of dancing, including square dancing, were considered lifetime. Boxing was not, even though it requires only two people. A second potential definition of lifetime would categorize all physical activities commonly reported in physical activities surveys of adults as lifetime. A third potential definition would adapt the first by adding the proviso that activities that require only one or two people but typically require an organizational structure (e.g., diving and wrestling) should not be considered lifetime.
- *Community organizations*—One of the PFE objectives involved an interest in monitoring patterns of participation in physical activity, especially through public recreation programs in community facilities. In NCYFS I and II, "community" was defined to include public and private community organizations, and respondents were asked to report the ones through which physical activity was received in the past year; NCYFS II added the condition "at least three times." This line of inquiry generated preliminary estimates of the exposure of children to different types of community organizations—including churches, Y's, and public parks and recreation programs—as providers of physical activity. It is recommended that a similar question be asked in future surveys with two changes. First, children should be asked to report the community organizations through which they received physical activity during the past 2 weeks. Second, for each community organization, respondents should be asked to report the number of times; average duration, frequency, and intensity; and major activities (top two).

Recommendations

Table 4 lists the data elements recommended for inclusion in a survey of physical activity patterns among children ages 9–18. We assume that special surveys will need to be developed for children age 8 and under, based on the considerations we have described in this chapter.

Several of the data elements may need pilot testing. One task for a pilot test will be to ensure that the survey items, now sequenced thematically, are resequenced nonreactively. The proposed survey should also be tested for its applicability to younger children. If feasible, a separate version of the survey should be prepared for joint administration to children 5–8 years with their parents. We are not satisfied that this approach to gathering data from younger children has been adequately explored.

An activity list needs to encompass activities that are likely to be reported. The 1985 NHIS Supplement listed approximately one-third the number of activities included in NCYFS I. Based on the responses to NCYFS I,

Table 4. Data elements recommended for survey of habitual physical activity among children and youth

Exercise, sports, and physically active hobbies (last 2 weeks)

- Activities:
  - Number of times done/played
  - Average minutes per time
  - Size of increase in heart rate and breathing
- Comparison with parents:
  - A lot more active
  - A little more active
  - As active
  - A little less active
  - A lot less active
- Comparison with peers:
  - A lot more active
  - A little more active
  - As active
  - A little less active
  - A lot less active
- Time spent in watching television:
  - Weekdays
  - Weekends

Community organizations

- Organizations through which physical activity was obtained (last 2 weeks):
  - Number of times
  - Average length of time
  - Size of increase in heart rate or breathing
  - Major activities at each (up to 2)

Schools

- Distance and means of transportation to and from school
- Whether enrolled in physical education:
  - Now (most recent term)
  - Prior term
- Number of days of physical education per week (when last enrolled this year)
- Class length in minutes
- Location of class
- Availability of time and place to:
  - Shower
  - Change clothes
- Amount of time spent in physical activity
- Most frequent physical education class activities (up to 6) in past year

Schools—Con.

- Of 10 typical physical education classes, number in which:
  - No increase in heart rate or breathing
  - Small increase in heart rate or breathing
  - Moderate increase in heart rate or breathing
  - Large increase in heart rate or breathing
- Number of recesses
- Participation in sport teams (last 2 weeks):
  - Nature of sport
  - Average frequency
  - Average duration
  - Average intensity
- Participation in fitness testing in physical education over prior year
- Types of fitness tests
- Feedback to parents
- Recognition for participation in testing

Work

- Household chores during last 2 weeks (weekdays)
  - Nature of activity
  - Average frequency
  - Average duration
  - Average intensity
- Household chores during last 2 weeks (weekends)
  - Nature of activity
  - Average frequency
  - Average duration
  - Average intensity
- Paying jobs or volunteer work
  - Nature of activity
  - Average frequency
  - Average duration
  - Average intensity

Knowledge

- Activity needed to strengthen heart and lungs
  - Frequency of activity needed
  - Duration of activity needed
  - Intensity of activity needed
- Concept of energy balance

Summative

- Overall frequency with which physical activity made heart beat and breathing a lot faster for 30 minutes or more in last 2 weeks:
  - Number of days

Table 5. Recommended additions to the 1985 National Health Interview Survey Supplement for administration to children based on responses in the National Children and Youth Fitness Study I

- Badminton
- Cheerleading
- Dodgeball
- Fishing
- Gymnastics
  - Apparatus
  - Free exercise
  - Tumbling
- Horseback riding
- Hunting
- Jumping or skipping rope
- Karate, judo, martial arts
- Kickball
- Relays
- Tag games
- Wrestling

the activities in table 5 should be added to the NHIS list for administration to children and adolescents 9–18 years. It will be noted that remarkably few low-organized games were reported frequently. Tag and relays ranked among the activities taking the greatest portion of time outside physical education class among only fifth and sixth graders. However, if the survey is to be applicable to children as young as age 5, provision must be made for reporting additional low-organized games and perhaps even spontaneous play. Three approaches may be considered: (1) Develop a long list of low-organized games by name and ask parents and children to review the list; (2) construct a matrix to allow children and parents to describe the physical requirements of the child's activities; (3) gather free-form responses without precoding. The first approach is unlikely to work because few low-organized games are consistently known by the same name; most carry multiple names to reflect ethnic, cultural, and regional variations. The second is a variant of an observational technique. The third was followed in NCYFS II.

The proposed survey approach focuses on the two most recently completed calendar weeks. The obvious shortcoming of this approach is its insensitivity to seasonal variations in the nature and dimensions of activity, which are known to be significant. To measure seasonal variations, it is recommended that all questions limited to a period of 2 calendar weeks be readministered in subsequent seasons. For those ages 9–18 years, telephone callbacks will probably be adequate.

Conclusion

Incorporation of protocols for assessment of health-related physical fitness and of physical activity habits is entirely feasible under varied conditions of rigor. Special considerations apply in testing the physical fitness of children under age 5 and in measuring the physical activity habits of those under age 9 or 10. Some of the experiences in conducting NCYFS I and II may prove instructive in attempts to target future studies to children and youth ages 6–18.

References

1. Ross, J.G., and Gilbert, G.G. The National Children and Youth Fitness Study: A Summary of Findings. *Journal of Physical Education, Recreation and Dance,* 56(1):45–50, 1985.
2. Ross, J.G., and Pate, R.R. National Children and Youth Fitness Study II: A Summary of Findings. *Journal of Physical Education, Recreation and Dance,* 58(10), 1987.
3. U.S. Department of Health and Human Services. *Promoting Health/Preventing Disease: Objectives for the Nation.* Washington, D.C.: Government Printing Office, 1980.
4. Hunsicker, P.A., and Reiff, G. Youth Fitness Report: 1958–1965–1975. *Journal of Physical Education and Recreation,* 48:31–33.
5. Pate, R.R. A New Definition of Youth Fitness. *Physician and Sports Medicine.* 11(4):77–83, 1983.
6. *AAHPERD Health Related Physical Fitness Test Manual.* Reston, VA: AAHPERD Publications, 1980.
7. Hunsicker, P., and Reiff, G. *Youth Fitness Test Manual.* Washington, D.C.: AAHPERD Publications, 1976.
8. Ross, J.G., Dotson, C.O., Gilbert, G.G., and Katz, S.J. New Standards for Fitness Measurement. *Journal of Physical Education, Recreation and Dance,* 56(1):62–66, 1985.
9. Pate, R.R., Ross, J.G., Delpy, L.A., Svilar, M., and Gold, R.S. National Children and Youth Fitness Study II: New Health-Related Physical Fitness Norms. *Journal of Physical Education, Recreation and Dance,* 58(10), 1987.
10. Errecart, M.T., Ross, J.G., Gilbert, G.G., and Ghosh, D.N. The National Children and Youth Fitness Study: Sampling Procedures. *Journal of Physical Education, Recreation and Dance,* 56(1):54–56, 1985.
11. Errecart, M.T., Saavedra, P., Svilar, M., and Ross, J.G. National Children and Youth Fitness Study II: Sampling Procedures. *Journal of Physical Education, Recreation and Dance,* 58(10), in press.
12. Ross, J.G., Katz, S.J., and Gilbert, G.G. The National Children and Youth Fitness Study: Quality Control. *Journal of Physical Education, Recreation and Dance,* 56(1):57–61, 1985.
13. Ross, J.G., Delpy, L.A., and Damberg, C. National Children and Youth Fitness Study II: Study Procedures and Quality Control. *Journal of Physical Education, Recreation and Dance,* 58(10), 1987.
14. Ross, J.G., Dotson, C.O., Gilbert, G.G., and Katz, S.J. What Are Kids Doing in School Physical Education? *Journal of Physical Education, Recreation and Dance,* 56(1):73–76, 1985.
15. Ross, J.G., Dotson, C.O., Gilbert, G.G., and Katz, S.J. After Physical Education. . .Physical Activity Outside of School Physical Education Programs. *Journal of Physical Education, Recreation and Dance,* 56(1):82–85, 1985.
16. Ross, J.G., Dotson, C.O., Gilbert, G.G. Are Kids Getting Appropriate Activity? *Journal of Physical Education, Recreation and Dance,* 56(1):82–85, 1985.
17. Ross, J.G., Pate, R.R., Casperson, C., and Svilar, M. Physical Activity Patterns at Home and in the Community. *Journal of Physical Education, Recreation and Dance,* 58(10), 1987.
18. Ross, J.G., Pate, R.R., Corbin, R., Delpy, L.A., and Svilar, M. What are Kids Doing in the School Physical Activity Program? *Journal of Physical Education, Recreation and Dance,* 58(10), 1987.
19. Dotson, C.O., and Ross, J.G. Relationships Between Activity Patterns and Fitness, *Journal of Physical Education, Recreation and Dance,* 56(1):86–89, 1985.
20. Pate, R.R., and Ross, J.G. Factors Affecting the Health-Related Fitness of Children. *Journal of Physical Education, Recreation and Dance,* 58(10), 1987.
21. Geiff, G. *The President's Council on Physical Fitness and Sports 1985 National School Population Fitness Survey.* Washington, D.C.: President's Council on Physical Fitness and Sports, 1986.
22. Department of Health and Human Services. *The 1990 Health Objectives for the Nation: A Midcourse Review.* Washington, D.C.: Public Health Service, 1986.
23. Powell, K.E., Spain, K.G., Christensen, G.M., and Mollenkamp, M.P. The Status of the 1990 Objectives for Physical Fitness and Exercise. *Public Health Reports,* 101(1):15–22, 1986.

24. Gilbert, G.G., Montes, J., and Ross, J.G. The National Children and Youth Fitness Study: History. *Journal of Physical Education, Recreation and Dance,* 56(1):51–53, 1985.
25. Granville Corporation. Set of Measures for Physical Fitness Measures Development. Washington, D.C.: Office of Disease Prevention and Health Promotion, 1982.
26. *AAHPERD Health Related Physical Fitness Technical Manual.* Reston, VA: AAHPERD Publications, 1984.
27. Gold, R.S., Damberg, C., Christensen, G.M., and Ross, J.G. History of the National Children and Youth Fitness Study II. *Journal of Physical Education, Recreation and Dance,* 58(10), 1987.
28. Bennett, W.J. *First Lessons: A Report on Elementary Education in America.* Washington, D.C.: Government Printing Office, 1986.
29. Lohman, T. Calf Skinfolds as an Alternative to Subscapular Skinfold in the Health Related Physical Fitness Test. Presented to the Annual AAHPERD Conference in Atlanta, Georgia, 1985.
30. Sparks, R.E. Modification of the AAHPERD Youth Physical Fitness Test. Unpublished doctoral dissertation, Springfield College, Springfield, MA, 1965. (Published in the University of Oregon Microcards, UO–66 #72, listed as: 613:7 Physical Fitness Testing GV 436, Sparks, Raymond E. "Modification of the AAHPERD Youth Physical Fitness Test" (Springfield, MA), 1965. Thesis (DPE) Springfield College.)
31. Pate, R.R., Ross, J.G., Baumgartner, T.A., and Sparks, R. A New Physical Fitness Test: The Modified Pull-Up. *Journal of Physical Education, Recreation and Dance,* 58(10), 1987.
32. Bar-Or, O., and Inbar, O. Relationships Among Anaerobic Capacity, Sprint and Middle Distance Running of School Children. In: Shepard, R.J., and Lavallee, H. (eds.) *Physical Fitness Assessment.* Springfield: C.C. Thomas, 1978, 142–147.
33. Krahenbuhl, G.S., Pangrazi, R.P., Petersen, G.W., Burkett, L.N., and Schneider, H.J. Field Testing of Cardiorespiratory Fitness in Primary School Children. *Med. Sci. Sports,* 10:208–213, 1978.
34. Wenger, N.K., Hellerstein, H.K., Blackburn, H., and Castranova, S.J. Physician Practice in the Management of Patients With Uncomplicated Myocardial Infarction—Changes in the Past Decade. *Circulation,* 65:421, 1982.
35. Clarke, H.H. Toward a Better Understanding of Muscular Strength. *Physical Fitness Research Digest,* 3:1–20, 1973.
36. Baumgartner, T.A. Modified Pull-Up Test. *Research Quarterly,* 49:80–84.
37. Baumgartner, T.A., East, W.B., Frye, P.A., Hensley, L.D., Knox, D.F., and Norton, C.J. Equipment Improvements and Additional Norms for the Modified Pull-Up Test. *Research Quarterly for Exercise and Sport,* 55:64–68.
38. Frishberg, B., Way, J., Brown, J., and Hunter, M. An Investigation into the Effectiveness of the Sit-and-Reach Box in Assessing Leg and Back Flexibility. Presented to the Annual AAHPERD Conference in Atlanta, Georgia, 1985.
39. Health and Education Services, Division of Novel Products, advertising flyer dated May 1, 1987.
40. Bar-Or, O. *Pediatric Sports Medicine for the Practitioner.* New York: Springer-Verlag, 1983.
41. Timmer, S.G., Eccles, J., and O'Brien, K. How Children Use Time. In: Juster, F.T., and Stafford, F.P. (Eds.) *Time, Goods, and Well-Being.* Ann Arbor: Institute for Social Research, The University of Michigan, 1983.
42. Baranowski, T., Dworkin, R., Cieslik, C.J., Hooks, P., Rains, D., Ray, L., Dunn, J.K., and Nader, P.R. Reliability and Validity of Self-Report of Aerobic Activity: Family Health Project. *Research Quarterly,* 55(4):309–317, 1984.
43. Baranowski, T., Tsong, Y., Hooks, P., Cieslik, C., and Nader, P.R. Aerobic Physical Activity Among Third to Sixth Grade Children. *Journal of Developmental and Behavioral Pediatrics,* in press.
44. Goden, G. Multiples of the Resting Metabolic Rate (METS) of Physical Activities. Presented to the Canada Fitness Survey, 1982.

Appendix: Instruments from NCYFS II

PARENT SURVEY

To assist in understanding the physical fitness of your child, we would appreciate your taking a few minutes to answer several questions about your child. You are not required to complete the survey, but doing so will provide more complete information. Answers will be kept strictly confidential.

1. Compared to other children of the same age/sex, is your child: CHECK ONE.

46 ☐1 A lot more physically active than most
☐2 A little more physically active than most
☐3 A little less physically active than most
☐4 A lot less physically active than most
☐5 Average--same as most

2. How much television does your child usually watch on the typical <u>school day</u>? CHECK ONE.

47 ☐1 1 hour or less
☐2 2 hours
☐3 3 hours
☐4 4 hours
☐5 5 hours
☐6 6 hours or more
☐7 None

3. How much television does your child usually watch on the typical <u>weekend day</u>? CHECK ONE.

48 ☐1 1 hour or less
☐2 2 hours
☐3 3 hours
☐4 4 hours
☐5 5 hours
☐6 6 hours or more
☐7 None

4. In the <u>past 12 months</u>, did your child get exercise or physical activity at least three times through any of the following organizations? CHECK ALL THAT APPLY.

49 ☐1 Public park or recreation center
50 ☐1 Church or other place of worship
61 ☐1 Sports teams or leagues
52 ☐1 YMCA, YWCA, or similar organization
53 ☐1 Health club, private spa, or private lessons
54 ☐1 Cub scouts, brownies, or other scouts
55 ☐1 4-H or other farm club

5. What types of exercise or physical activity did your child receive through the places you checked above in Question 4? LIST UP TO FIVE.

1. ____________________ (56-57)
2. ____________________ (58-59)
3. ____________________ (60-61)
4. ____________________ (62-63)
5. ____________________ (64-65)

6. In the typical <u>week</u>, on <u>how many days</u> do the child's parents or guardians (whomever the child lives with) get exercise that causes rapid breathing and a fast heart beat for 30 continuous minutes or more? LIST NUMBER OF DAYS PER WEEK FOR EACH PARENT. PUT NA IF CHILD DOES NOT LIVE WITH THIS PARENT.

Mother or Female Adult: 66 ☐ Father or Male Adult: 67 ☐

7. Compared to other <u>adults</u> of the same age and sex, how physically active are the child's <u>parents</u> (whomever the child lives with)? CHECK ONE FOR EACH PARENT.

<u>Mother or Female Adult</u>	<u>Father or Male Adult</u>
68 ☐1 A lot more physically active than most	69 ☐1 A lot more physically active than most
☐2 A little more physically active than most	☐2 A little more physically active than most
☐3 A little less physically active than most	☐3 A little less physically active than most
☐4 A lot less physically active than most	☐4 A lot less physically active than most
☐5 Average--same as most	☐5 Average--same as most
☐6 NA	☐6 NA

8. In the typical <u>week</u>, on <u>how many days</u> do the child's parents or guardians (whomever the child lives with) exercise with the child for 20 minutes or more? LIST NUMBER OF DAYS PER WEEK FOR EACH PARENT.

Mother or Female Adult: 70 ☐ Father or Male Adult: 71 ☐

TEACHER SURVEY

THIS SURVEY IS TO BE ADMINISTERED FACE-TO-FACE WITH WHOMEVER TEACHES PHYSICAL EDUCATION. IF TEACHING RESPONSIBILITY IS SHARED BY A SPECIALIST AND A CLASSROOM TEACHER, YOU SHOULD INTERVIEW THE SPECIALIST.

School: ______________________________ School ID: ☐☐☐ 1-3 Grade: ☐ 4

I'd like to ask you a number of questions mostly about the physical education program for children in this class. INDICATE THAT, IF THE TEACHER TEACHES PHYSICAL EDUCATION TO MORE THAN ONE CLASS, SEPARATE QUESTIONS WILL BE ASKED FOR EACH SELECTED CLASS THAT HE/SHE TEACHES.

1. Which children on this list (SHOW DATA COLLECTION CHECKLIST) are <u>not</u> enrolled in physical education class <u>this term</u>? ENTER ON SCORE CARDS.

2. Who teaches physical education to the children in this class? READ OPTIONS. CHECK ONE.

5 ☐ 1 A physical education specialist?
☐ 2 A general classroom teacher who also teaches physical education?
☐ 3 One of each: a specialist and general classroom teacher who share the responsibility?
☐ 4 Other (________________________)
Specify

3. In what field or fields are you (you and the other teacher) certified to teach? READ OPTIONS. CHECK ONE.

6 ☐ 1 Certified in physical education
☐ 2 Certified as a general classroom teacher
☐ 3 One person is certified as a specialist and the other as a general classroom teacher
☐ 4 You are certified in both PE and as a classroom teacher
☐ 5 Other (________________________)
Specify

4. How many days per week is the class usually scheduled to take physical education with the specialist? CHECK ONE.

7 ☐ 1 Not applicable--there is no specialist
☐ 2 1 day a week
☐ 3 2 days a week
☐ 4 2 days one week, 3 days the next week
☐ 5 3 days a week
☐ 6 4 days a week
☐ 7 5 days a week
☐ 8 Other (____________________)
Specify

5. How many days per week is the class usually scheduled to take physical education with the classroom teachers. CHECK ONE.

8 ☐1 Not applicable--classroom teacher does not teach PE
☐2 1 day a week
☐3 2 days a week
☐4 2 days one week, 3 days the next week
☐5 3 days a week
☐6 4 days a week
☐7 5 days a week
☐8 Other (________________)
Specify

6. How much time is allotted for the usual physical education class (from bell to bell)? I know this varies, but give me an average. IF RESPONSIBILITY IS SPLIT, LIMIT RESPONSE TO THE SPECIALIST. RECORD ANSWER IN MINUTES.

9-10 ☐☐ minutes

7. Are the children usually provided with a time and a place to:
Change clothes before and after class?

11 ☐1 Yes
☐2 No

Take showers after class?

12 ☐1 Yes
☐2 No

8. In the usual physical education class you have with these children, which you said lasts ____ minutes (READ TIME FROM QUESTION 7), how much time is spent on each of 4 categories:

13-14 ☐☐ minutes-- Category 1: Getting to and from physical education class, changing clothes, showering, taking roll, making general announcements, and disciplining children
15-16 ☐☐ minutes-- Category 2: Explaining an activity or demonstrating skills
17-18 ☐☐ minutes-- Category 3: Most children actually doing vigorous physical activities that speed up breathing and make the heart beat fast, such as running, jumping rope, bike riding, or very heavy chores.
19-20 ☐☐ minutes-- Category 4: Most children playing less vigorous games and other physical activities that involve walking and occasional spurts of movement, such as running or ball throwing.

9. Over the course of the year, what five physical activities, games, and sports take up the most periods of physical education for this class? LIST UP TO FIVE ACTIVITIES BELOW. IF YOU ARE TOLD LESS THAN FIVE, ASK IF THERE ARE ANY MORE. IF YOU ARE TOLD MORE THAN FIVE, READ THEM BACK AND ASK WHICH WOULD NOT BE ONE OF THE TOP FIVE IN TERMS OF THE NUMBER OF CLASS PERIODS THEY TAKE UP.

1. ________________ 21-22
2. ________________ 23-24
3. ________________ 25-26
4. ________________ 27-28
5. ________________ 29-30

10. What is the one place where you usually hold physical education class with these children? READ LIST. CHECK ONE.

31 ☐1 Gymnasium?
☐2 Auditorium?
☐3 Cafeteria or multipurpose room?
☐4 Classroom?
☐5 School Grounds?
☐6 Public Park?
☐7 Other (________________)?
Specify

11. Does the school sponsor any sports teams or other extracurricular physical activity programs for children in this grade?

32 ☐1 Yes
☐2 No

12. (IF YES TO #12) What sports teams or other extracurricular physical activity programs are available to children in this grade? LIST UP TO FIVE.

1. ________________________________ 33-34
2. ________________________________ 35-36
3. ________________________________ 37-38
4. ________________________________ 39-40
5. ________________________________ 41-42

13. What fitness tests would the children in this class normally take in the course of the school year? CHECK ALL THAT APPLY.

43 ☐1 Broad jump or long jump
44 ☐1 High jump
45 ☐1 Vertical jump
46 ☐1 Squat thrusts
47 ☐1 Regular push-ups
48 ☐1 Chair or knee push-ups
49 ☐1 Flexed-arm hang
50 ☐1 Pull-ups
51 ☐1 Chin-ups
52 ☐1 Modified pull-ups
53 ☐1 Sit-ups
54 ☐1 Sit-and-reach test
55 ☐1 Softball throw
56 ☐1 Skinfolds
57 ☐1 Dash/sprint
58 ☐1 Side step
59 ☐1 Shuttle run
60 ☐1 600-yard run
61 ☐1 Half-mile walk/run
62 ☐1 Mile walk/run
63 ☐1 6-minute run
64 ☐1 9-minute run
65 ☐1 12-minute run
66 ☐1 Other (______________________________)
Specify

67 ☐1 No fitness testing is done

14. Based on precedents, what percentage of the students in this class will receive some form of recognition for participating in your school's fitness testing?

68-70 ☐☐☐ Percent

15. Do parents get any feedback on the actual results of the fitness tests?

71 ☐1 Yes
☐2 No

16. Between the starting bell and the dismissal bell, how many times a day do children have recess? This is something other than PE class. CHECK ONE.

72 ☐1 None--they don't get recess
☐2 One
☐3 Two
☐4 Three
☐5 Other

17. How many minutes long is the average recess? ANSWER IN MINUTES.

73-74 ☐☐ Minutes

18. What do you think are the main reasons that many elementary schools don't do much fitness testing? I'll read you a list. CHECK ALL THAT APPLY.

75 ☐1 There isn't enough class time.
76 ☐1 The classes are too big; too many children.
77 ☐1 Risk of injury to the children is too great.
78 ☐1 Grades aren't given in physical education, so it doesn't matter.
79 ☐1 The children don't like being tested.
80 ☐1 The school or school district doesn't allow it.
81 ☐1 The mandated curriculum doesn't include testing.
82 ☐1 The main interest is movement education, not fitness or sports skills.
83 ☐1 Many teachers don't like to encourage competition among the children.
84 ☐1 Many teachers don't know how to administer certain tests.
85 ☐1 Parents are opposed to testing.
86 ☐1 Many teachers don't like the available tests.

19. Of the children in the class, which ones would you say are more physically active than most (top 40 percent) and which ones are less physically active than most (bottom 40 percent). REVIEW INDIVIDUAL SCORE CARDS. What we want to do now is sort the Score Cards into three piles: top 40 percent, middle 20 percent, and bottom 40 percent. DO NOT BE CONCERNED WITH PROPORTIONAL DISTRIBUTION.

Of the children you said are more active than most, are any of them a lot more active (top 20 percent)? We want to sort the top 40 percent pile into 2 piles: top 20 percent and next 20 percent. CONDUCT SORT

Of the children you said are less active than most, are there any who are a lot less active (bottom 20 percent)? What we want to do is sort the bottom 40 percent into 2 piles: bottom 20 percent and next higher 20 percent. CONDUCT SORT.

YOU NOW SHOULD HAVE 5 SORTED PILES. RECORD EACH STUDENT'S CLASSIFICATION DIRECTLY ONTO HIS/HER INDIVIDUAL SCORE CARD AS SOON AS YOU COMPLETE THE FINAL QUESTION.

20. As of today, how many children are there enrolled in this entire school in the __________(INSERT GRADE) Grade? We may need to get this figure from the office. IF TEACHER CANNOT ANSWER ACCURATELY, OBTAIN INFORMATION FROM PRINCIPAL'S OFFICE.

87-89 ☐☐☐ Boys
90-92 ☐☐☐ Girls

Thanks a lot for your help.

NAME: ______________________________

CHILD SURVEY

THIS SURVEY IS TO BE ADMINISTERED TO THE CHILD ORALLY PRIOR TO THE INDIVIDUAL FITNESS TESTING SESSION IN A PRIVATE SETTING.

1. First, could you tell me what grade you are in?

   ☐ 1
   ☐ 2
   ☐ 3
   ☐ 4

   Now I'd like to ask you a few questions about exercise and about physical education. All right?

2. Some boys (or girls) get more exercise than others. Would you say that you play games that make you breathe hard and make your heart beat fast:

   READ FIRST FIVE OPTIONS. IF CHILD HESITATES OR DOES NOT APPEAR TO UNDERSTAND, RE-READ SHORTENED VERSION OF EACH OPTION: " A LOT MORE," "A LITTLE MORE," "A LITTLE LESS," "A LOT LESS," OR "THE SAME." CHECK ONE.

   ☐ A lot more than most boys (or girls) you know?
   ☐ A little more than most boys (or girls) you know?
   ☐ A little less than most boys (or girls) you know?
   ☐ A lot less than most boys (or girls) you know?
   ☐ The same as most boys (or girls) you know?
   ☐ Don't know. DO NOT READ THIS OPTION.

3. Good. Now I'm going to read you some things that some people might say about exercise or about PE class. You tell me which ones are true and which ones are not true for you, ______________ STATE CHILD'S NAME. What you think is true for you is all that matters. No one else will hear what you tell me. AFTER READING FIRST STATEMENT, ASK, "TRUE OR NOT TRUE?"

   CHECK ONE BOX FOR EACH STATEMENT.

True	Not True	Don't Know	
☐	☐	☐	1. The things we do in PE class are always fun.
☐	☐	☐	2. My PE teacher is nice.
☐	☐	☐	3. We get all sweaty in PE class and I think that's gross.
☐	☐	☐	4. My PE class is too crowded.
☐	☐	☐	5. My PE teacher yells a lot.
☐	☐	☐	6. We do lots of different things in PE class.
☐	☐	☐	7. One thing I like about PE is getting out of class.
☐	☐	☐	8. Some kids in PE class make fun of me.
☐	☐	☐	9. I like to sweat when I exercise.
☐	☐	☐	10. We always do the same old things in PE class.
☐	☐	☐	11. Sometimes the place we have PE is too hot.
☐	☐	☐	12. Sometimes the place we have PE is too cold.
☐	☐	☐	13. The PE teacher sometimes makes fun of me.
☐	☐	☐	14. The PE teacher only pays attention to the kids who are good at sports.
☐	☐	☐	15. A problem with PE is we have to go outdoors in bad weather.
☐	☐	☐	16. I like to show how fast I am in PE.
☐	☐	☐	17. Lots of times the PE teacher holds up class when kids make trouble.
☐	☐	☐	18. PE would be better if we had it on more days.
☐	☐	☐	19. The PE teacher spends too much time teaching me things I already know.
☐	☐	☐	20. My mother thinks that PE is important.
☐	☐	☐	21. I like to breathe hard when I exercise.

True	Not True	Don't Know		
☐	☐	☐	22.	I'm better than most kids when it comes to running, throwing, and catching.
☐	☐	☐	23.	My father thinks that other classes are more important than PE.
☐	☐	☐	24.	I think it's really important to run fast and to throw and catch a ball well.
☐	☐	☐	25.	Nearly all of my friends like PE.
☐	☐	☐	26.	The PE teacher likes to exercise.
☐	☐	☐	27.	I like it when my heart beats really hard during exercise.
☐	☐	☐	28.	I'm really good at being physically fit.
☐	☐	☐	29.	My best friend thinks that arithmetic and reading are more important than PE.
☐	☐	☐	30.	Being physically fit is really important to me.

4. In general, do you like your PE class?

DO NOT READ RESPONSE OPTIONS. CHECK ONE. PROBE IF STUDENT ANSWERS "MEDIUM" OR GIVES ANOTHER RESPONSE WITH NEGATIVE CONNOTATIONS.

☐ Yes
☐ No
☐ It depends. It varies.
☐ Don't know

Chapter 12

Assessing Fitness and Activity Patterns of Women in General Population Studies

Barbara L. Drinkwater, Ph.D.
Pacific Medical Center, Seattle, Washington

Introduction

The elements of health-related physical fitness—muscular strength and endurance, cardiorespiratory endurance, flexibility, and body composition—may be assumed to be independent of gender. Therefore, these qualities, which are deemed important for continued good health, overall vitality, and the prevention of some disease states, are equally important for women and men. Recognition of this fact has been rather recent. Only in the last 10–15 years have women been encouraged—even permitted—to participate in activities designed to develop the essential components of health-related fitness. As a result, there is very little information available to describe the current fitness status or activity habits of American women.

This report will examine each of the four basic components of fitness, describe how these components have been measured in the female population, provide estimates of sample means and variability within each decade, and discuss specific aspects of test protocols that may require modification to meet the needs of women. At the end of the report, a section will consider concerns specific to women, such as pregnancy, menstruation, aging, and osteoporosis.

Recent changes in societal attitudes toward women's participation in sports and exercise must be considered in the interpretation of the physical fitness data from the National Health and Nutrition Examination Survey. Women in their twenties and thirties have come through an educational system that offers equal opportunities for boys and girls to participate in sports. Most older women did not have this opportunity. In fact, prior to the early 1970's, young women were discouraged from participating in any type of vigorous activity following puberty. Thus, an apparent effect of aging on fitness scores may well be confounded by the changes in the sociocultural acceptance of the active athletic woman.

A basic knowledge of anatomical and physiological gender differences will also be necessary if direct comparisons are made between the data for males and females. Too often, the word "inferior" is used relative to women's performance scores such as strength and aerobic power (VO_2max) when compared with values for males. Such judgmental phrases may lead to the assumption that men's higher scores in these two areas confer an added health benefit. In fact, medical statistics relative to cardiovascular disease suggest otherwise. Women's fitness scores should be interpreted in terms of their physiological capacity and their health concerns.

Cardiorespiratory endurance

Less than 30 years ago, it was widely believed that young women reached their peak cardiorespiratory en-

Table 1. Effect of training relative to initial fitness level of women 18–29 years of age

Measurement	<40 ml O_2·kg^{-1}·min^{-1}				>40 ml O_2·kg^{-1}·min^{-1}			
	n	Pre	Post	Percent	*n*	Pre	Post	Percent
VO_2max (ml·kg^{-1}·min^{-1})	181	35.6	42.5	19.5	105	44.4	48.4	9.1
$V_E$max, BTPS (l·min^{-1})	146	84.8	92.0	8.5	79	90.0	95.8	6.4
Heart rate (bpm)	136	192.3	193.1	0.4	74	191.6	189.0	1.3
Weight (kg)	76	59.4	59.4	0	73	58.8	56.2	4.4
Body fat (%)	145	27.1	25.7	5.2	84	22.4	19.3	14.1
Stroke vol. (ml)	27	66.0	79.9	21.0				
CaO_2-CvO_2 (ml·l^{-1})	27	126.9	132.3	4.2				

Note: Data were compiled from references 25–37.

durance at puberty. It was assumed that women's ability to participate in strenuous activity declined from that point onward, although men continued to improve their fitness (1). A retrospective examination of these early studies demonstrates the fallacy of generalizing from one population to another. The data represented the fitness of sedentary women but were assumed to represent the potential of all women. As more recent studies have shown, this assumption was erroneous.

To avoid similar misinterpretation of the NHANES III data by the lay public, it is important to relate the current fitness levels of American women to their potential for fitness. For that reason, VO_2max values for female athletes are included in the report as a basis for comparison. Although the genetic basis of fitness cannot be discounted, there is ample evidence from training studies that sedentary women can improve their cardiorespiratory fitness through physical-conditioning programs. In fact, those women with the lowest levels of initial fitness often have the highest percent improvement following training (table 1).

Aerobic power

Aerobic power (VO_2max) is generally accepted as the best measure of cardiorespiratory endurance, and the techniques for determining VO_2max are basically the same for men and women (Wilmore, chapter 4). Because the level attained during an exercise stress test represents both biological potential and habitual level of physical activity, data from studies prior to 1970 probably underestimate both the mean and variability of women's aerobic power today. For that reason, only studies published since 1970 are included in this report.

Variability

It is apparent from data drawn from the literature that the variability of VO_2max is greatest among younger women and grows progressively smaller with each successive decade. The range of VO_2max means for women ages 20–29 years represented by the weighted mean in table 2 is 19 ml O_2 per kilogram of body weight per minute (ml·kg·min^{-1}); for women in their sixties, 4.5 ml·kg·min^{-1}. This difference does not represent an aging effect on interindividual variability per se but less diversity in activity patterns and fewer samples from the older age groups. The standard deviations and standard errors within studies are slightly greater for the age group 20–29 years but do not differ markedly for older groups. When the number of subjects is approximately equal, the standard deviations and standard errors are also similar:

Table 2. Women's response to maximal exercise stress test

Mean age (yrs)	Age group	Height (cm)	Weight (kg)	*n*	VO_2max (ml·kg·min^{-1})	HRmax (bpm)	$V_E$max (l·min)
20–29 years	22.1	164.0	57.9	135	38.7	191.6	78.3 (*n* = 92)
30–39 years	36.0	165.2	60.4	136	30.6	183.1	73.2 (*n* = 61)
40–49 years	44.3	166.0	63.6	117	27.4	179.9	71.3 (*n* = 20)
50–59 years	54.0	165.1	63.6	75	25.6	177.0	58.6 (*n* = 12)
60–69 years	64.2	160.6	61.1	51	25.6	158.0	55.6 (*n* = 45)
70–75 years	70.6	160.9	63.0	67	20.9	135.8	—

Note: Weighted means from several studies are presented within each decade. Data were compiled from references 2–4, 15, 25, 34, and 38–52.

Age	*n*	VO_2max	*SD*	*SE*
20.3	33	35.7	7.0	1.2
30–39	33	28.3	3.4	0.6
40–49	39	25.9	3.3	0.5
54.0	41	25.0	3.2	0.5
63.4	29	26.4	3.4	0.6
70.2	42	18.9	3.9	0.6

Although this homogeneity might suggest the feasibility of using relatively small samples within older groups, the data for the athletes in an age group indicate otherwise (table 3). With only one exception, the age group 60–69 years, the mean aerobic power of the athletes is greater than two standard deviations above the mean of nonathletic women of the same age. If it is desirable to assess the diversity of fitness within each decade as well as to estimate the mean, it will be necessary to sample more widely.

Aging trends

A number of investigators have attempted to quantify the rate of decrement in aerobic power relative to age using both cross-sectional studies (2–4) and longitudinal studies (5, 6). The rate of decline varies considerably, depending on the baseline values of the group. When former physical education students were followed across 40 years, the regression slope, -0.63 ml $O_2\cdot kg\cdot min^{-1}\cdot yr^{-1}$ (5), was almost twice that reported for less active women, -0.32 ml $O_2\cdot kg\cdot min^{-1}\cdot yr^{-1}$ (7) over a period of 5 years. When active women in longitudinal studies become less active as they age or older women are less active than younger women in cross-sectional studies, the effect of age on aerobic power is confounded by a detraining effect. This may be a factor in the NHANES III study. The more active lifestyle of today's young women may be reflected in a steeper aging

Table 3. Physiological response of female athletes to a maximal exercise stress test

Age	Height (cm)	Weight (kg)	*n*	VO_2max $(ml\cdot kg\cdot min^{-1})$	HRmax (bpm)	$V_E$max (l·min)	Reference
15–20 years							
Basketball	173.0	68.3	15	49.6	186	116.9	46
Running	163.3	50.7	145	52.6	198	92.8	9, 53, 54
Swimming	169.2	60.1	21	51.6	192	121.0	55, 56
Synchronized	165.5	54.8	35	43.7	196	92.2	57, 58
Skiing							
Nordic	165.1	57.2	30	55.5	193	124.8	59–61
Alpine	165.1	58.8	13	52.7	198	112.7	59
Tennis	168.7	58.0	10	48.0	202	101.0	62
21–30 years							
Cycling	165.0	55.0	7	50.2	—	—	63
Lacrosse	164.0	57.4	7	52.9	189	116.9	64
Orienteering	162.7	58.1	8	51.4	195	99.8	65, 66
Pentathlon	175.4	65.4	9	45.9	185	106.1	67
Running	164.4	52.6	59	55.8	188	101.6	9, 53–56
Skating	175.5	73.9	13	46.1	191	96.5	68
Swimming	—	—	10	43.9	179	77.0	69
Volleyball	177.6	68.9	48	50.0	186	113.9	70–72
31–40 years							
Golf	168.9	61.8	23	34.2	—	—	73
Running	165.8	55.2	124	55.9	181	103.4	44, 74–77
Swimming	—	—	10	42.1	178	79.0	69
Tennis	163.3	55.7	25	44.2	179	82.7	45
Climbing	167.5	64.7	6	51.9	184	96.4	77
41–50 years							
Running	161.5	53.8	10	43.4	177	82.1	78
Swimming	—	—	9	38.3	163	74.0	69
51–60 years							
Swimming	—	—	7	35.9	163	75.0	69
61–70 years							
Swimming	—	—	6	32.1	159	66.0	69
71–80 years							
Swimming	164.2	61.6	2	37.6	162	78.9	52

Note: Mean values are presented for selected sports within each age group.

slope, or even a curvilinear rather than linear relationship of VO_2max with age. It will be interesting to compare the aging curve derived from a large population study with the -0.307 linear slope calculated from data representing 29 studies from eight countries (8).

The age-related values for maximal heart rate, HRmax (table 2), illustrate the difficulty of predicting VO_2max when submaximal exercise is terminated at a specific percent of an age-predicted HRmax. For example, if 50-year-old women were stopped at 70% of 170 bpm (220 − age), the actual %HRmax might range from 58% to 78%. Similar estimates of variability are found in all age groups and argue against using predicted HRmax as a criterion for predicting VO_2max.

The abrupt decrease in HRmax noted at age 60 in all four studies has been reported previously (7). The decrease of 19 bpm from the sixth to the seventh decade is greater than the 13 bpm drop across the previous 30 years. Though a linear rate of decrease can be calculated (-0.435 beats/year), a third-degree curve describes the relationship more precisely. If this departure from linearity is corroborated in the NHANES III study, it may represent a real difference between the sexes because the decline in maximal heart rate of males with age has uniformly been expressed as a linear function.

Maximal ventilatory volume ($V_E$max) data are noticeably lacking for the older age groups (table 2). This is an area where additional information would be very valuable.

Test protocols

Because the techniques and protocols for assessing physical fitness have been described in detail in chapter 4, this report will discuss only those aspects that are specific to women. In general, the same types of test—treadmill, cycle ergometer, step test, and assorted "field" tests—have been used for both men and women. However, protocols are often modified for female subjects, usually by adjusting the initial power output and/or power increments so the time for the test will be approximately the same for men and women.

The protocol must be one that is appropriate for the wide range of ages, skill, and fitness found in a random sampling of the female population. Although time is a factor, limiting tests to those that can be completed in a single session within 6–15 minutes, the increment in power across time must not be so demanding that true VO_2max is not attained. One reason for the variability in VO_2max values for older women may well be the choice of protocol for testing. Women tested under the Bruce protocol, which has large increments in metabolic demand between stages, uniformly have lower scores than those tested using the more gradual Balke method. The VO_2max levels they might have reached in a more gradual protocol fall somewhere between the final stage attained and the following stage.

Although cycle ergometers have many advantages in terms of safety, cost, space, precision of power settings, and a stable body position, they have one disadvantage that may be more of a factor for women than for men. Unless one is an experienced cyclist, the test is more likely to be terminated because of local muscle fatigue than because of cardiorespiratory limitations. As a result, the VO_2max may be underestimated. Because the VO_2max achieved on a cycle ergometer is usually 8–10% less than that attained on a treadmill, the combination of these two factors may markedly underestimate the cardiorespiratory fitness of American women. If a cycle ergometer is selected as the testing instrument, it is suggested that a pilot study be carried out to determine appropriate initial loads and power increments for women with and without cycling experience across a wide age range.

Treadmill tests have been used successfully with women of all ages and levels of fitness (2, 4, 9). Although the protocol may involve either running or walking, a running test is appropriate only for active women who can maintain balance and pace while running on the treadmill. The selection of a walking protocol for NHANES III will ensure a higher proportion of women agreeing to participate in the testing, a speedier learning period, and less chance of injury to the participants.

Walking protocols may be subdivided into those that are continuous and those that are discontinuous, with incremental or fixed speeds and grades. The two most frequently cited protocols for women are the Bruce and some modification of the Balke. Theoretically, both tests could extend beyond 20 minutes, but it is unlikely that many women from a general population would continue that long. Time to complete the Bruce test ranged from 4 to 11 minutes for women ages 29–60 years in one study (4). The range for a Balke-type protocol was 4–25 minutes for women 20–70 years of age (7). The pressures of time must be weighed against the additional information gained from a more leisurely protocol in choosing an appropriate procedure.

The step test has been used successfully in the Canadian Fitness Survey as an indirect estimate of VO_2max and has the advantage of producing data that could be compared with the Canadian figures and perhaps those of other countries as well (10). However, the step test is seldom used in this country as a means of determining cardiorespiratory fitness because it is a submaximal test for most people and VO_2max must be predicted rather than measured.

Direct versus indirect assessment of VO_2max

The optimal technique for assessing cardiorespiratory endurance would be a direct measurement of oxygen consumption during a symptom-limited exercise stress test. Although the equipment and personnel for this procedure would add to the cost of the survey, nothing is more expensive than inadequate data. Unfortunately, most attempts to predict VO_2max have underestimated values for some groups and overestimated values for others. The popular Astrand-Ryhming nomogram (11), for example, which predicts VO_2max from heart rate at submaximal work rates, has been reported to underestimate VO_2max when actual values are low and overestimate it when values are high (12).

Women and fitness

When subjects span a wide age range, as they will in NHANES III, errors are introduced into prediction equations by the decrement in maximal heart rate with age. Adding a correction factor for age does not solve the problem (13). Because the correction factor is based on the predicted decline in HRmax, it simply introduces another source of error.

Direct measurement of VO_2max is not as difficult as one might imagine. There are a number of units available commercially that combine the measurement of the physiological factors with an online computer to produce valid and reliable data. Pilot studies might be used to test some of the more popular units to see which stand up best under the field conditions imposed by the NHANES requirements.

An obvious concern in deciding to exercise people to their maximum capacity is the safety of the participants. Generally, this concern relates to the possibility of a cardiac incident with potentially fatal results. Fortunately, this is a rare occurrence even in clinical populations. The risk is even smaller for the female participants because the incidence of sudden cardiac death in the general population is less among women than among men (14). Nevertheless, basic precautions should be taken to protect all subjects.

If the safety concerns outweigh the need for precision in assessing cardiorespiratory endurance, a modification of the Astrand-Ryhming test described by Siconolfi et al. (13) is worth considering. This modification is also a submaximal cycle ergometer test but starts at a lower initial loan (24.5 W) for women and older men and increases the workload by 24.5 W at 2-minute intervals until heart rate reaches 70% HRmax. After steady-state heart rate is attained at this final load, VO_2max is estimated from the Astrand-Rhyming nomogram. This value is then inserted into a regression equation that includes a correction for age. The equation for women is:

$$y = 0.302\, x_1 - 0.019 \text{ age (yr)} + 1.593$$

where x_1 is the VO_2max predicted from the Astrand-Ryhming nomogram uncorrected for age. The standard error of estimate for women is 0.199 l $0_2 \cdot \text{min}^{-1}$.

This test was originally validated on an inactive adult population 20–70 years of age and would require further validation for children and active adults. The authors recommend the test specifically for use in epidemiological investigations.

Sex differences in aerobic power

Mean values of VO_2max can be expected to be higher for males than for females, although there will be considerable overlap between active women and less active men. However, if female endurance athletes are represented in the sample, they may have scores well above all but the fittest men. The reason for this apparent contradiction in sex differences in VO_2max lies in the relationship between biological potential and training in determining maximal aerobic power. The biological advantages of the male are negated when the woman is more highly trained. To place the mean values for each decade in perspective, it is necessary to know the maximal attainable values for women in that age group. The best estimate of the highest aerobic power women can attain is that of the female athlete. Table 3 provides the available data by age and sport for women ages 15–71 years. It is obvious that there is a dearth of data for the older female athlete.

Much of the apparent difference in fitness between the sexes is related to body size. When VO_2max is expressed in absolute terms as $L \cdot \text{min}^{-1}$, women average 50–55% of the men's score (15, 16). A similar difference can be found between a male jockey (2.7 L $O_2 \cdot \text{min}^{-1}$) and a football lineman (5.3 L $O_2 \cdot \text{min}^{-1}$) and is simply a matter of differences in size (17). To eliminate the effect of size, it is customary to divide VO_2 by body weight. The jockey (53.8 ml $0_2 \cdot \text{kg} \cdot \text{min}^{-1}$) is now a match for the lineman (52.1 ml $0_2 \cdot \text{kg} \cdot \text{min}^{-1}$). When the aerobic power of men and women is compared after equating for body weight, the difference is much less. Depending on the groups compared, it may range from 11% to 34% for athletes and averages about 28% for a general population (15, 16).

Part of the remaining difference is related to the effect of estrogen and testosterone on body composition. Compared with men, women usually have a higher percentage of body fat, a tissue that adds to body weight but does not contribute directly to endurance performance. To determine the physiological basis for sex differences in aerobic power, the role of body fat must be negated by dividing absolute VO_2max by *lean* body mass. When this is done, the difference between male and female aerobic power averages about 5–10%. This residual difference presumably represents the true biological difference between the sexes and is most likely related to the lower hemoglobin concentration of the female, her smaller muscle-to-body-weight ratio, and perhaps an overall smaller functional capacity of the oxygen transport system.

Because NHANES is concerned more with health-related fitness than with performance, the values attained in the assessment of body composition should be utilized to calculate aerobic power relative to lean body mass (VO_2, $\text{ml} \cdot \text{kg LBM}^{-1} \cdot \text{min}^{-1}$). This value plus absolute VO_2max and VO_2 $\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ will provide the most accurate assessment of women's absolute work capacity, endurance performance when the workload is body weight, and the physiological status of the oxygen transport system.

Body composition

The precision of body composition data available in the literature is questionable because of difficulties inherent in the measurement techniques and disregard of the basic assumptions underlying the prediction of body fat from indirect measurements. These problems have been addressed at length by Wilmore in chapter 4 and should be considered in reviewing current data and in planning NHANES approach to this topic.

Although the densiometric method has been accepted as the most accurate method of determining percent body fat (%BF), the formulas that translate body density to %BF assume a constant value for the density of lean tissue for all individuals. Where this assumption is in error, the estimate of %BF will be in error. Because it is likely that the density of lean tissue is less in women than in men, the present estimates of %BF for women may overestimate the actual value. Even within the female population, there will be variation in the density

Table 4. Percent body fat (%BF) of sedentary, active, and athletic women, by age and sport

Age	n_h	%BF	5Sk	%BF	n_{sk}
15–20 years					
Sedentary	128	25.1			
Athletic:					
Basketball	95	19.8	102.4	19.2	49
Gymnastics	49	15.1	54.8	15.1	49
Distance running	135	15.0			
Swimming	7	18.9	59.1	18.9	7
Tennis	10	23.3			
21–30 years					
Sedentary	555	26.5	102.2	25.7	385
Active	36	20.0			
Athletic:					
Basketball	10	19.7			
Cycling	27	21.0			
Distance running	110	16.0	42.4	13.3	8
Field events	19	26.5	101.7	24.7	5
Field hockey	53	22.1			
Lacrosse	7	23.1			
Marathon	15	15.2	55.0		
Skiing, Alpine	20	18.7			
Skiing, Nordic	25	17.1	63.2	16.1	5
Swimming	20	18.0			
Volleyball	60	19.7			
31–40 years					
Sedentary	169	26.3	98.0	28.8	64
Active	17	21.2			
Athletic:					
Distance running	20	14.8			
Golf	23	24.0			
Lacrosse	25	20.3			
Marathon	51	15.5	64.0	15.5	42
Swimming	11	18.6			
Tennis	25	20.3			
41–50 years					
Sedentary	223	30.4	120.0	31.5	87
Active	23	22.7			
Athletic:					
Distance running	10	18.3			
Swimming	10	21.1			
51–60 years					
Sedentary	65	34.8	120.3	41.9	21
Active	23	25.0			
Athletic:					
Swimming	6	23.8			

Note: The term "athletic" is used to describe women who participate in competitive and/or recreational activities. The sum of five skinfolds (5Sk) with an associated %BF has been added for groups when available. Sample size is given for hydrostatic weighing (n_h) and skinfolds (n_{sk}) separately.

of lean tissue between active and sedentary women and between age groups. At the present time, measurement of body composition is in a state of flux. New techniques are being sought that will provide an accurate assessment of body composition in spite of confounding factors such as age and activity status. Unless these problems are resolved in time to incorporate the solution into NHANES III, the use of anthropometric measurements is recommended as an interim solution to the problem of estimating levels of body fat.

Although anthropometric measurements include circumferences, diameters, and skinfolds, it is suggested that NHANES confine its measurements to skinfolds. This will limit the number of techniques that must be mastered to one and also decrease the time required for this test.

At least 15 different sites have been used by various investigators in attempting to predict %BF from either the sum of skinfolds or weighted values of individual sites. Because men and women differ markedly in the areas where fat accumulates, sites selected for the NHANES project must reflect the sex-specific deposition of fat. For women, appropriate sites include the triceps, subscapular area, abdomen, suprailium, and thigh. The sum of all five has been used by a number of investigators in reporting the physical characteristics of their subjects, which ensures some comparative data (table 4). Three of the sites are used in the Jackson et al. (18) generalized equation predicting the body density of women. Although it is possible to use skinfold data to predict %BF, it should be remembered that the method has the same potential error as the hydrostatic technique in translating body density to %BF because both assume the same constant for density of lean tissue. An alternative to using skinfolds to predict either density or %BF is to use the sum of five or more sites to establish percentiles for the population and/or subgroups within the general population. In some ways this classification would be more useful than %BF itself because individuals would know precisely where they stood relative to other women of the same age and could observe how their relative position changed with age, diet, and activity habits. Percentiles would not be useful in deriving lean body mass.

At this time it is impossible to define the upper limits of %BF that are associated with good health, although it has been customary to define female obesity arbitrarily as greater than 30%BF. This has been unfair to older women, whose higher values probably reflect an age-related decrease in the density of lean tissue as well as an increase in adipose tissue. It is evident that the average value for each age group, the "norm," is not necessarily the optimum %BF and should not be interpreted as the goal women should try to attain. A more realistic value for the optimum %BF might be derived from studies of active women in each age group. These are listed in table 4 along with values of %BF for sedentary and athletic women of the same age.

Flexibility

Flexibility, or range of motion about the joint, is dependent on the functional status of the muscles, tendons, ligaments, cartilage, and synovial fluid surrounding the joint. Degenerative diseases, disuse, and perhaps some aspects of biological aging result in a decrease in flexibility in many older adults. The result is an impairment of movement and increased susceptibility to injury. Whether a decrease in flexibility is an inevitable part of aging is debatable. There is, for example, very little loss in the range of motion of those joints used most frequently in everyday activities (19). NHANES may make it possible to explore the relationship between habitual activity patterns, aging, and current levels of flexibility if specific items are added to the questionnaire for this purpose.

One of the difficulties in measuring flexibility per se is that range of motion is specific to the joint being tested. A single measure of flexibility, such as the sit-and-reach, will not adequately describe the overall flexibility status of an individual. Measurement sites should involve movements that are important in everyday living, where the range of motion has been shown to decrease significantly with age and where the pattern of change is similar for males and females.

Women are more flexible than men at all ages, although the difference between the sexes varies depending on the specific joint measured and the age of the subjects (20). Appropriate measurement sites for women where flexibility decreases linearly with age are neck flexion and extension, elbow flexion and extension, trunk flexion and extension, and knee flexion (20). A number of other movements have an erratic pattern of change with age or have less practical application to everyday tasks. One exception might be hip flexion and extension, which are important in locomotion and might be found to have a significant relationship with age if a curvilinear analysis were applied to the data.

An estimate of the variability to be expected in these measurements can be found in the paper by Bell and Hoshizaki (20). For men and women, ages 17 years and over, the range of standard deviations for the five sites was:

Neck flexion-extension:	12–24 degrees
Elbow flexion-extension:	19–33 degrees
Trunk flexion-extension:	13–20 degrees
Knee flexion:	15–23 degrees
Hip flexion-extension:	20–31 degrees

The two instruments available that can provide reliable and valid measurement of range of motion about a joint are the electrogoniometer and flexometer. The electrogoniometer could be interfaced with a microprocessor to provide a real-time digital score and a hard copy of the results. Once the software has been thoroughly validated, this procedure should save time and ensure accuracy in recording flexibility scores.

Muscular strength and endurance

The biological basis for the difference in strength of men and women is presumed to be a function of the gonadal hormones. The anabolic effect of testosterone, which is approximately 10 times higher in males, provides men with a larger muscle mass and a larger ratio of lean body mass to body weight. Because strength is primarily a function of muscle size, men generally tend to exceed women in absolute strength. As with the other facets of fitness, there will be some overlap between the sexes. Women can increase their strength with training, and the strength-trained woman may be stronger than an untrained man. However, the overlap tends to be less for strength than for aerobic power or body composition.

In general, the overall absolute strength of women averages two-thirds that of men. The difference varies considerably from one muscle group to another, ranging from 35% to 79% in the upper extremities and from 57% to 86% in the lower extremities (21). Therefore, the difference between the sexes in the NHANES study will depend in large part on the selected measurement sites. This should not be the primary factor in selecting sites but should be considered in discussing the resultant differences.

Strength can also be expressed relative to body weight or lean body mass. This procedure will decrease the differences between the sexes but not totally eliminate them in the younger age group (22, 23). The effectiveness of normalizing for body size in older age groups is unknown. In fact, there are very little data in the literature relative to the strength of older women.

As with other fitness variables, there may be a spurious relationship between age and strength for the female population. Until the last 10–15 years, there was a strong cultural bias in this country against large or well-defined muscles on women. As a result, women avoided activities that might conceivably increase muscle bulk, thereby avoiding activities that also increased muscle strength. In recent years this taboo has largely disappeared, and many women have turned to weight training as a means of improving sports performance and avoiding injury. Even women who do not train for strength are likely to be stronger because they are generally more active than women were prior to 1970. Apparent age-related decreases in strength may be magnified by these changes in lifestyle of the younger women.

Part of the difficulty in collating strength scores for women lies in the variety of techniques used to measure strength and the variety of units used to describe the results (table 5). NHANES can avoid this difficulty by following the International System of Units and expressing force as newtons (N), work as joules (J), and power as watts (W).

Use of an isokinetic dynamometer would be the optimal method of measuring the muscular strength of the female population for several reasons: (1) Strength and endurance can be measured on the same machine, (2) a microprocessor can be interfaced with the machine to provide immediate results, (3) the technique is easier to master and far safer for older women than attempting lifts with free weights, and (4) the technique is preferable to performance tests such as the vertical jump, which measure skill as well as strength. Because strength is specific to the muscle group tested, it is important to select sites representative of upper body, lower body,

Table 5. Selected strength measurements for women 15–69 years of age

	Ages				
Test	15–19 years	20–29 years	30–39 years	40–49 years	50–59 years
Grip (kg)	31.9	37.3	36.5	29.5	26.6
Leg press (kg)		100.5			
Bench press (kg)		34.1			
Arm curl (kg)		20.3			
Military press		30.4			
Dead lift (kg)		71.2			
Cable Tensiometer (kg):					
Shoulder extension	33.3				
Shoulder flexion	27.0				
Elbow flexion	27.3				
Elbow extension	20.8				
Hip flexion	40.6				
Hip extension	41.4				
Knee extension	84.6				
Isokinetic (J):					
Knee extension, 30'/sec		156.6	134.2	120.3	118.1
Knee flexion, 30'/sec		82.1	71.7	70.3	63.1
Shoulder extension, 180'/sec		48.1	39.6	42.2	36.5
Shoulder flexion, 180'/sec		32.2	28.2	30.7	24.9

Note: Data were compiled from references 23 and 79–82.

and trunk musculature. The choices for women might include shoulder flexion and extension, leg press, knee extension, and abdominal flexion. A grip strength test might be added because it has been a popular test in the past, and NHANES can provide normative data for those forced to rely on this inexpensive technique.

In the older groups, care must be taken to prevent injury to osteoporotic women. Spontaneous fractures can occur with minimal trauma in this group, so it is possible that a maximal effort in any type of strength test could result in a fracture. It may be that women with a history of osteoporotic fractures or showing overt signs of vertebral crush fractures should be excluded from the sample.

Special fitness concerns of women

Pregnancy

There is little to be gained from including pregnant women in the fitness testing. The questions that are of concern to women regarding exercise during pregnancy cannot be answered in a one-time test for physical fitness and should be approached in carefully designed longitudinal studies. There are, however, two aspects of the pregnancy question that are of concern to NHANES: (1) Is there a possibility of injury to the fetus or the mother if a pregnant woman is unknowingly included in the sample? and (2) Can information about the exercise habits of pregnant women be obtained from the questionnaire?

Although it is highly unlikely that any of the fitness tests recommended for NHANES could result in injury to the fetus or mother during the early stages of an unsuspected pregnancy, a pilot study should be carried out to check some factors that might adversely affect the course of a pregnancy. For example, one of the concerns regarding exercise in the first trimester is the possible teratogenic effect of high maternal body temperature. Monitoring core temperature of women as they progress through the test battery in a pilot study would provide evidence that core temperature can be maintained below 39° C. Both active and sedentary women should be included because core temperature is a function of the relative intensity and duration of exercise. A physician experienced in the research area of exercise and pregnancy should be retained as a consultant in planning this pilot study.

Activity patterns and pregnancy outcomes can be assessed via the questionnaire. Unfortunately, one of the most important factors, intensity of exercise, is difficult to quantify without knowing the fitness level of the respondent and actually monitoring her physiological response to the activity. However, it will be interesting to learn what proportion of pregnant women do maintain their exercise programs, what the activities are, and how they modify their activity as pregnancy progresses. Without access to medical records and information regarding other factors that can adversely affect the fetus, any conclusions relating exercise to pregnancy outcome will be tenuous at best.

Osteoporosis

Although cardiovascular disease remains a prime health concern for men in this country, osteoporosis is the major health problem for older women. It has been estimated that 15–20 million women in the United States have one or more symptoms of this disease, at an annual cost to the Nation's health care system estimated at $3.8 billion (24). NHANES provides a unique opportunity for building a bone mineral density (BMD) data base representative of the general population that can be used to determine the prevalence of this disease, quantify the changes in BMD with age, examine the relationship between BMD and other factors such as nutrition and activity patterns, and assess the effectiveness of current efforts to educate women in how to protect themselves against osteoporosis.

It is recommended that planners for NHANES consider adding a measurement of bone density to the test battery. If it is feasible to consider the purchase of the measurement devices, the tests could be done on site. Otherwise, tests could be planned for smaller groups in those geographic areas where the techniques are available. The least expensive technique is single-photon densiometry. Unfortunately, it is also the least sensitive. A better choice would be the dual-photon or CT scan, both more expensive but capable of assessing BMD of the lumbar vertebrae. The pros and cons of each technique should be examined in depth before a selection is made.

It is possible that osteoporotic women will be selected as subjects for the NHANES study. Because nontraumatic fractures are common in this group, it is possible that a fracture might be precipitated by the testing procedure. Unless there is some pressing need to assess their fitness on the general fitness battery, they should be excluded on medical grounds. As an alternative, a modified test could be designed to assess their capability to perform tasks considered important for maintenance of an independent lifestyle.

Because a number of older women are unaware of their osteoporotic condition, the screening examination should include questions relating to prior fractures, loss of height, and back pain, which might suggest the presence of the disease. Confirmation of a BMD more than two standard deviations below the age group mean for women over 60 years of age plus one or more positive physical indicators might be reasonable criteria for exclusion.

References

1. Morehouse, L.E., and P.J. Rasch. *Scientific Basis of Athletic Training*. Philadelphia: Saunders, 1958.
2. Drinkwater, B.L., S.M. Horvath, and C.L. Wells. Aerobic power of females, ages 10 to 68. *J. Gerontology* 30:385–394, 1975.

3. Hossack, K.F., and R.A. Bruce. Maximal cardiac function in sedentary normal men and women: Comparison of age-related changes. *J. Appl. Physiol.* 53:799–804, 1982.
4. Profant, G.R., R.G. Early, K.L. Nilson, F. Kusumi, V. Hofer, and R.A. Bruce. Response to maximal exercise in healthy middle-aged women. *J. Appl. Physiol.* 33:595–599, 1972.
5. Asmussen, E., K. Fruensgaard, and S. Norgard. A follow-up longitudinal study of selected physiologic functions in former physical education students—after forty years. *J. Am. Geriatr. Soc.* 23:442–450, 1975.
6. Astrand, I., P.O. Astrand, I. Hallback, and A. Kilbom. Reduction in maximal oxygen uptake with age. *J. Appl. Physiol.* 35:649–654, 1973.
7. Plowman, S., B.L. Drinkwater, and S.M. Horvath. Age and aerobic power in women: A longitudinal study. *J. Geront.* 34:512–520, 1979.
8. Drinkwater, B.L. Physiological responses of women to exercise. *Exerc. Sport Sci. Rev.* 1:125–153, 1973.
9. Burke, E., and F. Brush. Physiological and anthropometric assessment of successful teenage female distance runners. *Res. Q.* 50: 180–187, 1979.
10. Shephard, R.J., C. Allen, A.J.S. Benade, C.T.M. Davies, P.E. di-Prampero, R. Hedman, J.E. Merriman, L. Myhre, and R. Simmons. Standardization of sub-maximal exercise tests. *Bull. W.H.O.* 38: 765–776, 1968.
11. Astrand, P.O., and I. Ryhming. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *J. Appl. Physiol.* 7:218–221, 1954.
12. Astrand, P.O., and K. Rodahl. Textbook of work physiology. New York: McGraw-Hill Book Co., 1977, pp.349–353.
13. Siconolfi, S.F., E.M. Cullinane, R.A. Carleton, and P.D. Thompson. Assessing VO_2max in epidemiologic studies: Modification of the Astrand-Ryhming test. *Med. Sci. Sport Exerc.* 14:335—338, 1982.
14. Kannel, W.B., and H. E. Thomas, Jr. Sudden coronary death: The Framingham Study. In: *Sudden Coronary Death.* (Eds. H.M. Greenberg and E.M. Dwyer) *Ann. NY Acad. Sci.* 382:3–21, 1982.
15. Drinkwater, B.L. Women and exercise: Physiological aspects. *Exerc. Sport Sci. Rev.* 12:21–51, 1984.
16. Sparling, P.B. A meta-analysis of studies comparing maximal oxygen uptake in men and women. *Res. Q.* 51:542–552, 1980.
17. Wilmore, J.H. The application of science to sport: Physiological profiles of male and female athletes. *Can. J. Appl. Sport Sci.* 4:103–115, 1979.
18. Jackson, A.S., M.L. Pollack, and A. Ward. Generalized equations for predicting body density of women. *Med. Sci. Sports Exerc.* 12:175–182, 1980.
19. Adrian, M. Flexibility in the aging adult. In: *Exercise and Aging* (Eds. E.L. Smith and R.C. Serfass), Hillside, NJ: Enslow Publishers, 1981, pp.45–58.
20. Bell, R.D., and T.B. Hoshizaki. Relationship of age and sex with range of motion of seventeen joint actions in humans. *Can. J. Appl. Sport Sci.* 6:202–206, 1981.
21. Laubach, L.L. Comparative muscular strength of men and women: A review of the literature. *Aviat. Space Environ. Med.* 47:534–542, 1976.
22. Morrow, J.R., Jr., and W.W. Hosler. Strength comparison in untrained men and trained women athletes. *Med. Sci. Sports Exerc.* 13:194–198, 1981.
23. Wilmore, J.H. Alterations in strength, body composition, and anthropometric measurements consequent to a 10-week weight training program. *Med. Sci. Sports.* 6:133–138, 1974.
24. *Osteoporosis.* National Institutes of Health, Consensus Development. Volume 5, Number 3, 1984.
25. Bransford, D.R., and E.T. Howley. Effects of training on plasma FFA during exercise in women. *Eur. J. Appl. Physiol.* 41:151–158, 1979.
26. Burke, E.J. Physiological effects of similar training programs in males and females. *Res. Q.* 48:510–517, 1977.
27. Cunningham, D.A., and J.S. Hill. Effect of training on cardiovascular response to exercise in women. *J. Appl. Physiol.* 39:891–895, 1975.
28. Cunningham, D.A., D. McCrimmon, and L.F. Vlach. Cardiovascular response to interval and continuous training in women. *Eur. J. Appl. Physiol.* 41:187–197, 1979.
29. Daniels, W.L., D.M. Kowal, J.A. Vogel, and R.M. Stauffer. Physiological effects of a military training program on male and female cadets. *Aviat. Space Environ. Med.* 50:562–566, 1979.
30. Eisenman, P.A., and L.A. Golding. Comparison of effects of training on VO_2max in girls and young women. *Med. Sci. Sports* 7:136–138, 1975.
31. Fringer, M.N., and G.A. Stull, Changes in cardiorespiratory parameters during periods of training and detraining in young adult females. *Med. Sci. Sports* 6:20–25, 1974.
32. Kearney, J.T., G.A. Stull, J.L. Ewing, Jr., and J.W. Strein. Cardiorespiratory responses of sedentary college women as a function of training intensity. *J. Appl. Physiol.* 41:822–825, 1976.
33. Knowlton, R.G., D.S. Miles, M.N. Sawka, J.B. Critz, and C. Blackman. Cardiorespiratory adaptations of females to cross country training. *J. Sports Med.* 13:391–398, 1978.
34. Lesmes, G.R., E.L. Fox, C. Stevens, and R. Otto. Metabolic responses of females to high intensity training of different frequencies. *Med. Sci. Sports* 10:229–232, 1978.
35. Milburn, S., and N.K. Bates. A comparison of the training responses to aerobic dance and jogging in college females. *Med. Sci. Sports Exerc.* 15:510–513, 1983.
36. Pedersen, P.K., and K. Jorgensen. Maximal oxygen uptake in young women with training, inactivity, and retraining. *Med. Sci. Sports* 10:233–237, 1978.
37. Puhl, J.L., and W.S. Runyan. Hematological variations during aerobic training of college women. *Res. Q. Exerc. Sport* 51:533–541, 1980.
38. Stewart, K., C. Williams, and B. Gutin. Determinants of cardiorespiratory endurance in college women. *Res. Q.* 48:413–419, 1977.
39. Bransford, D.R., and E.T. Howley. Oxygen cost of running in trained and untrained men and women. *Med. Sci. Sports* 9:41–44, 1977.
40. Eddy, D.O., K.L. Sparks, and D.A. Adelizi. The effects of continuous and interval training in women and men. *Europ. J. Appl. Physiol.* 37:83–92, 1977.
41. Smith, D.P., and F.W. Stransky. The effect of training and detraining on the body composition and cardiovascular response of young women to exercise. *J. Sports Med.* 16:112–120, 1976.
42. Kollias, J., H.L. Bartlett, P. Oja, and C.L. Shearburn. Cardiac output of sedentary and physically conditioned women during submaximal exercise. *Aust. J. Sports Med.* 9:63–68, 1977.
43. Diaz, F.J., R.D. Hagan, J.E. Wright, and S.M. Horvath. Maximal and submaximal exercise in different positions. *Med. Sci. Sports* 10:214–217, 1978.
44. Upton, S.J., R.D. Hagan, B. Lease, J. Rosentsweig, L.R. Gettman, and J.J. Duncan. Comparative physiological profiles among young and middle-aged female distance runners. *Med. Sci. Sports Exerc.* 16:67–71, 1984.
45. Vodak, P.A., W.M. Savin, W.L. Haskell, and P.D. Wood. Physiological profile of middle-aged male and female tennis players. *Med. Sci. Sports Exerc.* 12:159–163, 1980.
46. Vacarro, P.D., D.H. Clarke, and J.P. Wrenn. Physiological profiles of elite women basketball players. *J. Sports Med.* 19:45–54, 1979.
47. Seals, D.R., J.M. Hagberg, B.F. Hurley, A.A. Ehsani, and J.O. Holloszy. Endurance training in older men and women: I. Cardiovascular responses to exercise. *J. Appl. Physiol.* 57:1024–1029, 1984.
48. Sidney, K.H., and R.J. Shephard. Maximum and submaximum exercise tests in men and women in the seventh, eighth, and ninth decade of life. *J. Appl. Physiol.* 43:280–287, 1977.
49. Niinimaa, V., and R.J. Shephard. Training and oxygen conductance in the elderly: I. The respiratory system. *J. Gerontology* 33:354-361, 1978.

50. Morse, C.E., and E.L. Smith. Physical activity programming for the aged. In: *Exercise and Aging.* (Eds. E.L. Smith and R.C. Serfass), Hillside, NJ: Enslow Publishers, 1981, 191 pp.
51. Goertzen, D., R. Serfass, G. Sopko, and A. Leon. The functional capacity and physical activity levels of women over 60 years of age. *J. Sports Med.* 24:30–36, 1984.
52. Vacarro, P., G.M. Dummer, and D.H. Clarke. Physiologic characteristics of female masters swimmers. *Physician Sportsmed.* 9(12):75–78, 1981.
53. Butts, N.K. Physiological profiles of high school female cross country runners. *Res. Q. Exerc. Sport* 53:8–14, 1982.
54. Ready, A.E. Physiological characteristics of male and female distance runners. *Can. J. Appl. Spt. Sci.* 9:70–77, 1984.
55. Bell, G.H., and P.M. Ribisl. Maximal oxygen uptake during swimming of young competitive swimmers 9–17 years of age. *Res. Q.* 50:574–582, 1979.
56. Holmer, I., A. Lundin, and B.O. Erickson. Maximum oxygen uptake during swimming and running by elite swimmers. *J. Appl. Physiol.* 36:711–714, 1974.
57. Poole, G.W., B.J. Crepin, and M. Sevigny. Physiological characteristics of elite synchronized swimmers. *Can. J. Appl. Spt. Sci.* 5:156–160, 1980.
58. Roby, F.B., M.J. Buono, S.H. Constable, B.J. Lowdon, and W.Y. Tsao. Physiological characteristics of champion synchronized swimmers. *Physician Sportsmed.* 11(4):136–147, 1983.
59. Haymes, E.M., and A.L. Dickinson. Characteristics of elite male and female ski racers. *Med. Sci. Sports Exerc.* 12:153–158, 1980.
60. Rusko, H., M. Havu, and E. Karvinen. Aerobic performance capacity in athletes. *Eur. J. Appl. Physiol.* 38:151–159, 1978.
61. Rusko, H., P. Rahkila, and E. Karvinen. Anaerobic threshold, skeletal muscle enzymes and fiber composition in young female cross-country skiers. *Acta Physiol. Scand.* 108:263–268, 1980.
62. Powers, S.K., and R. Walker. Physiological and anatomical characteristics of outstanding female junior tennis players. *Res. Q. Exerc. Sport* 53:172–175, 1982.
63. Burke, E.R., F. Cerny, D. Costill, and W. Fink. Characteristics of skeletal muscle in competitive cyclists. *Med. Sci. Sports Exerc.* 9:109–112, 1977.
64. Withers, R.T. Physiological responses of international female lacrosse players to pre-season conditioning. *Med. Sci. Sports* 10:238–242, 1978.
65. Knowlton, R.G., K.J. Ackerman, P.I. Fitzgerald, S.W. Wilde, and M.V. Tahamont. Physiological and performance characteristics of United States champion class orienteers. *Med. Sci. Sports Exerc.* 12:164–169, 1980.
66. Saltin, B., and P.O. Astrand. Maximal uptake in athletes. *J. Appl. Physiol.* 23:353–358, 1967.
67. Krahenbuhl, G.S., C.L. Wells, C.H. Brown, and P.E. Ward. Characteristics of national and world class female pentathletes. *Med. Sci. Sports* 11:20–23, 1979.
68. Maksud, M.G., R.L. Wiley, L.H. Hamilton, and B. Lockhart. Maximal VO_2, ventilation, and heart rate of Olympic speed skating candidates. *J. Appl. Physiol.* 29:186–190, 1970.
69. Vaccaro, P., S.M. Ostrove, L. Vandervelden, A.H. Goldfarb, D.H. Clarke, and G.M. Dummer. Body composition and physiological responses of master female swimmers 20 to 70 years of age. *Res. Q. Exerc. Sports* 55:278–284, 1984.
70. Kovaleski, J.E., R.B. Parr, J.E. Hornak, and J.L. Roitman. Athletic profile of women college volleyball players. *Physician Sportsmed.* 8(2):112–118, 1980.
71. Puhl, J.L., and W.S. Runyan. Hematological variations during aerobic training of college women. *Res. Q. Exerc. Sport* 53:257–262, 1982.
72. Spence, D.W., J.G. Disch, H.L. Fred, and A.E. Coleman. Descriptive profiles of highly skilled women volleyball players. *Med. Sci. Sports Exerc.* 12:299–302, 1980.
73. Crews, D., G. Thomas, J.H. Shirreffs, and H.M. Helfrich. A physiological profile of ladies professional golf association tour players. *Physician Sportsmed.* 12(5):69–74, 1984.
74. Christensen, C.L., and R.O. Ruhling. Physical characteristics of novice and experienced women marathon runners. *Br. J. Sports Med.* 17:166–171, 1983.
75. Hagan, R.D., T. Stratham, L. Stratham, and L.R. Gettman. Oxygen uptake and energy expenditure during horizontal treadmill running. *J. Appl. Physiol.* 49:571–575, 1980.
76. Wilmore, J.H., and C.H. Brown. Physiological profiles of women distance runners. *Med. Sci. Sports* 6:178–181, 1974.
77. Drinkwater, B.L., P.O. Kramar, J.F. Bedi, and L.J. Folinsbee. Women at altitude: Cardiovascular responses to hypoxia. *Aviat. Space Environ. Med.* 53:472–477, 1982.
78. Vacarro, P., A.F. Morris, and D.H. Clarke. Physiological characteristics of masters female distance runners. *Physician Sportsmed.* 9(7):105–108, 1981.
79. Conger, P.R., and R.B.J. Macnab. Strength, body composition, and work capacity of participants and nonparticipants in women's intercollegiate sports. *Res. Q.* 38:184–192, 1967.
80. Kindig, L.E., P.L. Soares, J.M. Wisenbaker, and S.R. Mrvos. Standard scores for women's weight training. *Physician Sportsmed.* 12(10):67–74.
81. O'Shea, J.P., and J. Wegner. Power weight training and the female athlete. *Physician Sportsmed.* 9:109–120, 1981.
82. Petrofsky, J.S., R.L. Burse, and A.R. Lind. Comparison of physiological responses of women and men to isometric exercise *J. Appl. Physiol.* 38:863–868, 1975.
83. Dummer, G.M., D.H. Clarke, P. Vaccaro, L. Vandervelden, A.H. Goldfarb, and J.M. Sockler. Age-related differences in muscular strength and muscular endurance among female masters swimmers. *Res. Q. Exerc. Sport.* 56:97–110, 1985.

Chapter **13**

Exercise Testing and Physical Activity Assessment of Persons with Selected Cardiac Conditions

Nanette Kass Wenger, M.D.
Emory University

Introduction

Exercise testing or other objective assessment of functional capability is of particular importance in a population whose disease is characterized by a hidden or inapparent disability, and patients with atherosclerotic coronary heart disease are the most prominent example of this problem. These patients typically appear to be physically nondisabled and thus are often expected to be totally self-sufficient; they, therefore, are generally thought to be able to be included in the group of active, assertive persons without disabilities. Objective assessments of their capabilities, when compared with their perceived abilities as well as with the observer's impression of their abilities, may provide valuable information.

Comparable concerns are appropriate for persons with other cardiovascular diseases, but most of these have been less comprehensively studied than has coronary disease in regard to activity assessment.

Advantages of exercise testing as an objective measure are that it is a relatively safe procedure and one within the capability of most patients; the cost is limited; it can be performed in many ambulatory settings, including hospitals, clinics, and private practice offices; and it is a test generally available in the community at large. Further, it is a test that lends itself to serial repetition, as well as to comparison with a considerable literature of exercise test data and correlations with cardiovascular disorders.

Exercise testing as a screening procedure for cardiovascular disease

No abnormality encountered at exercise testing can be considered "diagnostic" of any specific type or etiology of cardiovascular disease, in that the exercise test abnormalities may reflect a limitation of exercise capacity, an abnormality of electrical function (arrhythmia), the development of changes highly correlated with myocardial ischemia, or the occurrence of other features limiting the cardiac output. Any and all of these can be encountered with a variety of cardiovascular disorders and, indeed, are not diagnostic or specific even for cardiovascular diseases as a group.

For example, although in the patient suspected to have atherosclerotic coronary heart disease, there is major concern with the adverse prognostic implication of exercise-induced hypotension, virtually any cause of left ventricular dysfunction (obstructive or congestive cardiomyopathy, for example), in addition to coronary disease with exercise-induced ischemia, may result in exertional hypotension. Exertional hypotension may also be encountered as a result of a number of disturbances of cardiac rhythm that may be precipitated by exercise, may be seen with a variety of etiologies of valvular heart disease, and may occur related to the administration of a multitude of drugs, including antihypertensive agents, beta blockers, peripheral vasodilators, sympatholytic drugs, psychotropic drugs, and the like.

Similarly, the occurrence of arrhythmias at exercise testing, whether supraventricular or ventricular arrhythmias, is not specific for any defined etiology of cardiac disease. The prognostic significance of ventricular ectopy, when encountered at exercise testing, relates predominantly to the underlying cardiovascular disease and the adequacy of ventricular function rather than to the characteristics of the ventricular ectopy per se.

Further, a variety of cardiovascular drugs may affect the results of exercise testing because of the hemodynamic changes brought about by drug administration, the direct effects of the drugs on the myocardium, or both (1, 2). Most important is the recent delineation that antianginal drugs may limit the ability to identify patients with even severe atherosclerotic coronary heart disease through exercise testing (3, 4). Indeed, the exercise test response and exercise capacity of patients with atherosclerotic coronary heart disease may be significantly altered by a number of commonly used cardiovascular drugs, alone or in combination, related to complex interactions of hemodynamic and electrophysiologic changes. Thus, the effects of pharmacotherapy must be appreciated when evaluating the results of exercise testing. (See also the later discussion of Bayes' theorem.)

Of critical importance is the fact that exercise capacity at testing reflects both the habitual activity level and the disease-related impairment. The former may be limited by choice of lifestyle, by the degree of perceived disease-related impairment, by actual disease-related impairment, by appropriate or inappropriate activity restriction recommended by physicians; or a combination of these features. To what extent these components can be delineated from questionnaire data remains unknown but is unquestionably a challenge for NHANES III. A pilot study may help assess the feasibility of these determinations.

Approaches to physical activity assessment

Approaches to physical activity assessment are also discussed in chapter 5, 6, 9, 21, and 24. A sizable data base is available from the NHANES II activity questionnaire. Additional demographic and physical examination data relative to cardiovascular disease have also been collected in the past (see chapter 2). Comparable information should be collected in the next NHANES. As will be discussed subsequently, it will be important to correlate the objective documentation of exercise performance with the reported activity levels and perceived ability to exercise of persons with a variety of cardiovascular diseases; in persons with coronary disease, as well as in apparently normal individuals, it would be of interest to correlate exercise capacity with other coronary risk factors.

In the National Health Interview Survey (NHIS), a considerable data base is also available defining physical activities and their levels, perceptions about physical activity, and data on coronary risk factors such as smoking, stress, and hypertension, and indeed data about diagnosed heart disease. It would appear valuable to relate these features, as well as the results of exercise testing, to disability days, physician visits, work-loss days, hospital days, and activity limitations described as secondary to cardiovascular disease.

The Framingham Study Physical Activity Questionnaire (5) used a relatively simple scoring system in that the respondents were asked to specify the activities they performed for each hour of a typical day. Activities were categorized as resting, standing, slight, moderate, or heavy; a "weighting value," reflecting the energy requirements, was assigned such that sleep entailed a weighting of 1.0, standing 1.1, slight activity (walking) 1.5, moderate activity 2.4, and heavy activity 5.0. This allowed a classification of individuals into least, intermediate, and most active groups based on a 24-hour activity score of less than 29, 29–36, and greater than 37, respectively. A potential value of replicating this approach is that current comparisons can be made with the extensive prior Framingham data. After 10 years of followup, there was a significantly lessened incidence of coronary disease, using the end points of coronary death, angina pectoris, and myocardial infarction, in patients who had moderate or high levels of activity as compared with the least active group.

Another option would be to use the questionnaire from the activity study of Morris and associates, wherein male civil servants recorded their activities for a Friday and Saturday, so as to assess a working and a leisure day (6).

Although the issue of fitness is not addressed in this chapter, the physical activity-fitness interrelationships can be addressed, as was done in a variety of U.S. studies—Framingham offspring, Tecumseh (chapter 15), military populations (chapter 16)—as well as in studies conducted in other countries. The evaluation of these interrelationships is further discussed in chapters 24 and 25. A detailed summary of techniques and instruments used to assess activity is presented in chapters 9 and 24 and will not be duplicated here.

Statistical information

The prevalence of noninstitutionalized disabled adults in the United States is estimated at from 8 million to 17 million (7–9). In 1972, 3.2 million adults were estimated to have a cardiovascular disability, with more than half of these individuals classified as so severely disabled as to be either totally unable to work or unable to work regularly (10).

In 1975, atherosclerotic coronary heart disease in the United States was estimated to involve almost 4 million persons, about equally divided between those below age 65 and those ages 65 and over. More than half of these 4 million persons are believed to have some limitation of normal activity. Next to problems resulting from arthritis, coronary heart disease accounts for the greatest losses related to physical limitation and restriction owing to a chronic disease in the United States (11).

Despite the continuous substantial decline in mortality from atherosclerotic coronary heart disease in the United States in the last decade, almost 1½ million episodes of myocardial infarction still occur each year. Furthermore, about 200,000 persons initially develop angina pectoris each year. The fact that coronary heart disease remains the major cause of morbidity in the United States, as well as the leading cause of mortality, is evidenced by the 20% of all survivors of myocardial infarction who have varying degrees of physical, psychosocial, and vocational disabilities.

Atherosclerotic coronary heart disease is the leading cause of premature disability benefits (before age 65) awarded in the U.S. labor force under the social security system (12). Thirty-three percent of the almost 500,000 persons receiving Social Security Administration disability allowances in 1975 were permanently disabled due to coronary heart disease. Data from the 1978 National Health Interview Survey indicated that atherosclerotic coronary heart disease affected almost 4 million persons, of whom over 2 million (over 60%) suffered some disability; half of the individuals affected were under age 65. Atherosclerotic disease is the major cause of permanent disability for both men and women ages 40 and over and is the primary diagnostic category for days of hospital care. The number of persons ages 20–64 categorized as having disability from atherosclerotic disease numbers almost 9 million, accounting for one-quarter of the total number of persons reporting limitation of activity resulting from chronic conditions. A major concern with prevalence estimates, particularly in regard to morbidity and to limitation of activity, is that they are based on health *interview* information, i.e., self-diagnosis and reporting. No physical examination is done, and no confirmatory documentation is available, so the interview-derived data may underestimate or overestimate the true prevalence. Exercise testing or other objective evaluation of functional capacity may enable the estimation of the validity of interview data for the population at large, as well as for specific subsets of persons.

All cardiovascular diseases in the United States, the overwhelming majority resulting from arteriosclerosis, rank first as causes of limitation of activity, days in the hospital, and disabled worker benefits; second as causes of physician visits and days in bed; and fourth as causes of days lost from work.

Exercise testing and physical activity assessment: Atherosclerotic coronary heart disease

Benefits of rehabilitative care for patients with acute myocardial infarction

Early ambulation of appropriately selected patients during the hospitalization for acute myocardial infarction effects an improvement in the functional status of this group of individuals at the time of discharge from the hospital; this has been reported to be associated with an earlier and more complete subsequent return to activity and to work (13–15). Early ambulation also permits the appropriate performance of predischarge exercise testing.

Many persons who have sustained a coronary event or who are hospitalized for other manifestations of coronary illness initially complain of lack of energy or inordinate weakness when they resume activity after the period of immobilization; this reflects their increase in perceived exertion to perform a given task, because it entails a greater proportion of their physical work capacity. Deconditioning caused by protracted bed rest decreases the maximal oxygen uptake (physical work capacity, as shown in figure 1). Further, many symptoms of the often associated depression, particularly fatigue, may be misinterpreted as indicating organic illness.

Predischarge exercise testing, in addition to its pivotal role in risk stratification (16–19), allows a more precise definition of safely tolerated activity levels. This delineation may permit the relatively unimpaired patient

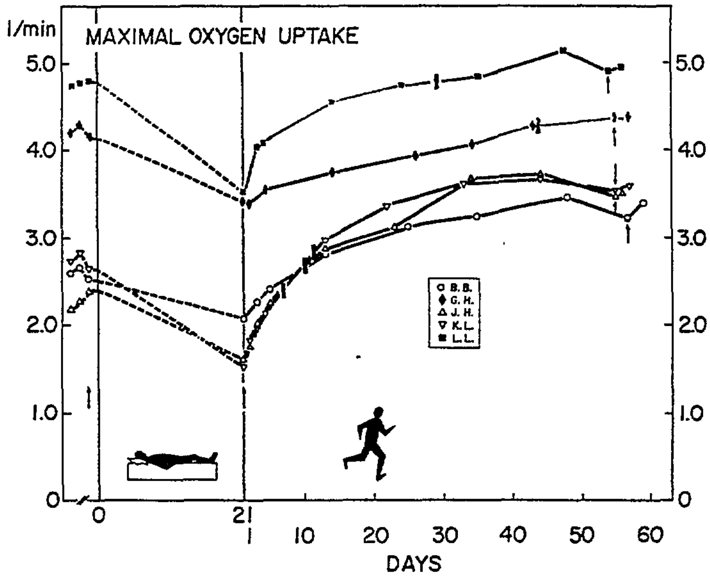


Figure 1. Changes in maximal oxygen uptake with bed rest and training. Individual data before and after bed rest and at various intervals during training. Arrows indicate circulatory studies. Heavy bars mark the time during the training period at which the maximal oxygen uptake had returned to the control value before bed rest. From Saltin B, Blomqvist G, Mitchell JH, Johnson RL, Wildenthal K, and Chapman CB: Response to exercise after bed rest and after training. *Circulation* 37–38(Suppl 7):1, 1968. By permission.

to more rapidly resume normal activity following a coronary event. It also demonstrates the anticipated level of activity that can be performed without adverse effects and thereby decreases the fear, common in persons after myocardial infarction, that physical activity may result in recurrent infarction or sudden cardiac death. It thus may also exert a positive psychologic impact and thereby improve functional status.

In the National Exercise and Heart Disease Project (20, 21) individuals recovered from myocardial infarction who had the least physical limitation were those who achieved high age-predicted heart rate levels or relatively high aerobic work levels; patients ages 40–49 years achieved a level of 9.5 METS; patients 50–59 years, 9.0 METS, and subjects 60–64 years, 8.5 METS. The mean levels at exercise testing for subjects recovered from myocardial infarction, in comparable age ranges, whose exercise tests were terminated because of adverse signs or symptoms approximated 7.5, 7.0, and 6.5 METS, respectively; thus the relative functional impairment in the latter three groups varied from 25% to 28%. Stated otherwise, sign- and symptom-limited patients recovered from myocardial infarction, as defined at exercise testing, possess 72–75% of the aerobic capacity of patients whose exercise test was terminated because they attained a predefined heart rate limit.

Predischarge exercise testing: Acute coronary episode

In recent years, formal low-intensity exercise testing prior to discharge from the hospital after myocardial infarction or other coronary events has had increasing application and acceptance (22–24). It is used both as a diagnostic and as a prognostic test, as well as to help guide therapy and to assess the residual functional capacity as a basis for recommending exercise rehabilitation. Persons with abnormalities at low levels of exercise or with a very limited exercise capacity appear to be at increased risk of proximate coronary events and are often considered candidates for early coronary arteriography; the predischarge exercise test can identify activity-induced left ventricular dysfunction, activity-induced ventricular arrhythmias, and activity-induced ischemia, all of which suggest multivessel coronary disease and myocardium in jeopardy. In terms of functional assessment, the exercise test performance can guide recommendations for initial activity levels at home and for subsequent return to work, and it can serve as a basis for prescriptive rehabilitative physical activity.

The general guidelines used to select patients who are inappropriate for predischarge exercise testing may also be considered exclusion criteria for exercise testing in NHANES designed to assess activity levels. In general, patients with angina pectoris at low levels of activity, angina at rest, decompensated congestive heart failure at rest or at low levels of activity, postural hypotension, untreated or uncontrolled hypertension, or serious ventricular dysrhythmias should not undergo this test. Additionally, a variety of musculoskeletal, neurologic, and pulmonary problems may limit the person's ability to perform activity and would thus render exercise testing unwise (25).

This test may have equally important applications in the context of assessing the activity levels of a variety of physically impaired individuals. The level of testing used for patients after myocardial infarction approximates the level of physical activity permitted during the last days of the hospitalization, i.e., exercise at a workload of 3–3.5 METS (1 MET = about 3.5 cc O_2/kg body weight/min) or that evoking a heart rate response below 120–130 beats/min. Thus the treadmill testing protocol shown in table 1 (26) may entail serial 3-minute stages of walking at 1.2 miles per hour, initially on the level and then at 3% and 6% elevations (24). A modified Bruce protocol maintains a constant speed of 1.4 miles per hour, but the grade increases from 0 to 5% to 10% in stages 0, one-half, and 1, respectively, with a MET expenditure of 1.7, 2.9, and 4.0 for the three stages (27). The modified Naughton protocol (28) has four stages of exercise with a constant speed of 2 miles per hour; the percent of the grade increases from 0 to 3.5 to 7 to 10.5, with a resultant expenditure of 2, 3, 4, and 5 METS, respectively.

In general, the occurrence of even mild anginal discomfort, dyspnea, or dizziness is an indication for stopping the test. The clinical signs warranting cessation of exercise testing include 0.1 to 0.2 mv of ischemic ST segment depression or elevation, a fall in systolic blood pressure, the occurrence of complex ventricular ectopy, development of an unsteady gait, or an alteration in the level of communication or responsiveness (the last signifying decreased cerebral perfusion). Finally, the attainment of the target heart rate (usually 60–70% of the age-predicted maximal value) or a target workload (usually 4–5 METS) is often another end point.

In addition to symptoms and ECG changes, the blood pressure response to effort (as noted earlier) is an important indicator in that exercise-induced hypotension, particularly early in the course of exercise testing (but after initial adaptation to walking), has an ominous implication, suggesting that modest activity may produce significant ischemic ventricular dysfunction and further suggesting that this may occur during usual daily activities. Exercise-induced hypotension at higher levels of activity is of less concern, and hypotension at peak exercise does not carry an adverse prognosis but rather suggests that the person has exercised beyond the aerobic threshold (29, 30). The pivotal characteristics of the predischarge tests that warrant emphasis are the low initial level of effort and the smaller serial increments in workload than are usual with standard exercise tests.

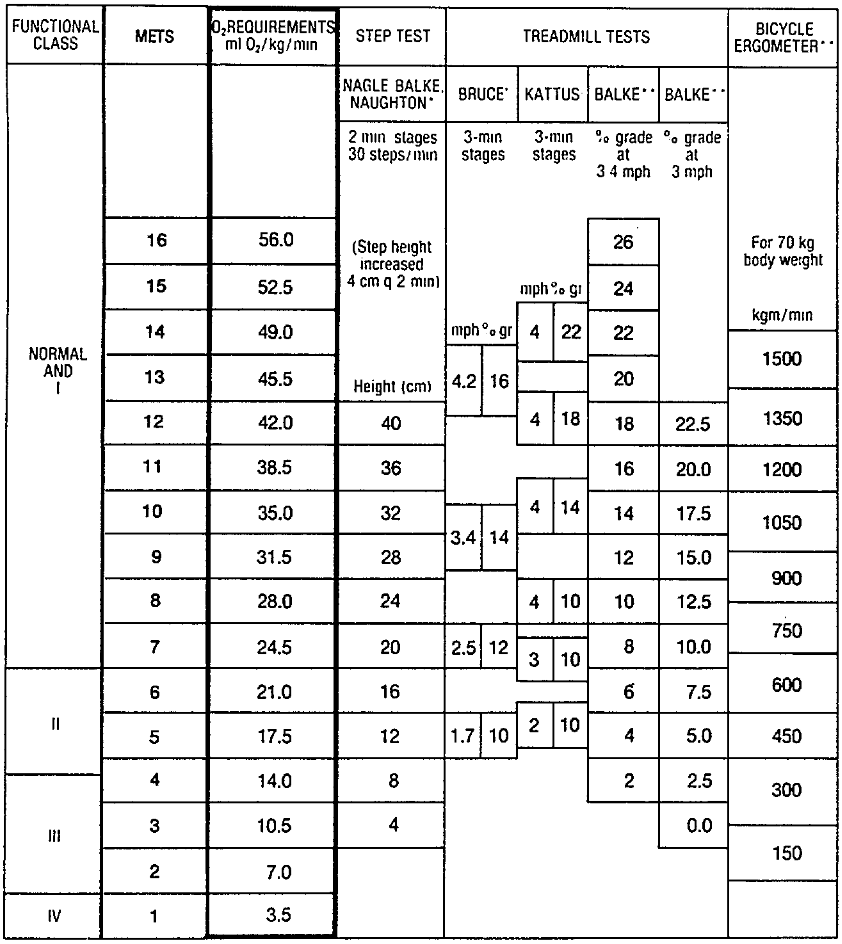
In experienced exercise testing laboratories, this low-level predischarge exercise testing has not been associated with appreciable complications in patients recovered from myocardial infarction. It appears to be a highly acceptable means for assessing functional capacity because of its gradual progression of activity and generally low level of imposed activity challenge; however, it may be inadequate to assess the functional capacity of less limited individuals. In some centers, patients are tested to a sign- or symptom-limited end point (26); often, there is relatively little difference in the intensity of the test enabled by a sign- and symptom-limited end point and by a predefined heart-rate-limited response. Thus either a heart-rate-limited or a sign- or symptom-limited end point test may be selected for NHANES, although I favor the latter owing to the inhomogeneity of the group to be tested.

There remains considerable controversy as to whether the addition of radionuclide procedures adds significantly to the information obtained at predischarge exercise testing. This uncertainty, coupled with the major cost and frequent lack of facilities or competent personnel to perform these procedures in many community hospitals, makes it unreasonable to consider radionuclide-based myocardial perfusion studies or radionuclide ventriculograms (either first-pass studies or multiple-gated acquisition studies) as part of the NHANES examination.

Although the application of low-level predischarge exercise testing has been predominantly for patients hospitalized for an acute coronary episode, this gradually progressive, low-level test appears to be among the most suitable formats for assessing patients with severe cardiovascular impairment of any etiology and, in particular, may be an appropriate test to document the functional capabilities of patients with congestive cardiac failure of a variety of etiologies.

Traditional exercise testing for patients with atherosclerotic coronary heart disease

Exercise testing is used to provide a quantitative assessment of the patient's functional cardiovascular status. Various methods and protocols for exercise testing (figure 2) may be employed in patients with stable angina pectoris or after recovery from myocardial infarc-



FUNCTIONAL CLASS	METS	O_2 REQUIREMENTS ml O_2/kg/min	STEP TEST	TREADMILL TESTS						BICYCLE ERGOMETER**
			NAGLE BALKE, NAUGHTON*	BRUCE†		KATTUS‡		BALKE**	BALKE**	
			2 min stages 30 steps/min	3-min stages		3-min stages		% grade at 3.4 mph	% grade at 3 mph	
NORMAL AND I	16	56.0	(Step height increased 4 cm q 2 min)					26		For 70 kg body weight
	15	52.5				mph	% gr	24		
	14	49.0		mph	% gr	4	22	22		kgm/min
	13	45.5	Height (cm)	4.2	16			20		1500
	12	42.0	40			4	18	18	22.5	1350
	11	38.5	36					16	20.0	1200
	10	35.0	32	3.4	14	4	14	14	17.5	1050
	9	31.5	28					12	15.0	900
	8	28.0	24			4	10	10	12.5	
	7	24.5	20	2.5	12	3	10	8	10.0	750
II	6	21.0	16					6	7.5	600
	5	17.5	12	1.7	10	2	10	4	5.0	450
III	4	14.0	8					2	2.5	300
	3	10.5	4						0.0	150
	2	7.0								
IV	1	3.5								

*Nagle FS, Balke B, Naughton JP: Gradational step tests for assessing work capacity. *J Appl Physiol* 20:745-748, 1965.

†Bruce RA: Multi-stage treadmill test of submaximal and maximal exercise. Appendix B, this publication.

‡Kattus AA, Jorgensen CR, Worden RE, Alvaro AB: S-T-segement depression with near-maximal exercise in detection of preclinical coronary heart disease. *Circulation* 41:585-595, 1971.

**Fox SM, Naughton JP, Haskell WL: Physical activity and the prevention of coronary heart disease. *Ann Clin Res* 3:404, 1971.

Figure 2. Oxygen requirements for step, treadmill, and bicycle ergometer. Oxygen requirements increase with workloads from bottom of chart to top in various exercise tests of the step, treadmill, and bicycle ergometer types. Adapted from American Heart Association: *The Exercise Standards Book,* AHA Publication 10–79–25M, p. 11. By permission.

Table 1. Summary data from low-level exercise tests in patients with recent myocardial infarction, angina pectoris, and coronary artery bypass surgery

Study	Patient no./classification	Test			Conclusion and/or prognostic indicator for future cardiac events or MVCD
		$\bar{x}$ time $\bar{p}$ event	Mode	End point	
MI and angina pectoris patients:					
Jelinek et al. (1977)	30/MI	8 days	B	S_s	(C) safe
Nixon et al. (1980)	61/Unstable AP	When stable	B	HR (120)	(C) safe (PI) S–T Dp
Theroux et al. (1979)	210/MI	11 days	T	5 METS/70% MHR	(PI) S–T Dp → higher 1 yr mortality
Corbett et al. (1981)	61/MI	19 ± 1 days	B	HR (130)	(PI) S–T Dp ≥ 0.1 mV
Markiewicz et al. (1977)	46/MI	3, 5, 7, 9, 11 wk.	T	S_s/HR (130) wk 3, HR (140) wk 5, HR (150) wk 7	(PI) S–T Dp
Mautner et al. (1981)	13/MI	13 days	T	S_s	(PI) S–T Dp → MVCD
Fuller et al. (1981)	40/MI	13 ± 2.1 days	T	HR (120)	(PI) S–T Dp → MVCD, S–T → ± MVCD
Smith et al. (1979)	62/MI	18 days	T	60% MHR	(C) LLET influenced clinical management in 21% cases (PI) S–T Dp
Vecchio et al. (1981)	74/MI (nontransmural)	4 wk	B	S_s	(PI) S–T Dp/AP
Cokkinos et al. (1981)	151/MI (65 anterior)	1–3 mo	T	S_s/6 min	(PI) AP (no S–T Dp) → MVCD in anterior MI patients
Sammel et al. (1980)	78/MI	1 mo	T	15 min	(PI) S–T Dp + AP → MVCD; no S–T Dp/AP → mild to moderate disease
Schwartz et al. (1981)	48/MI	18–22 days	T	75% MHR	(PI) S–T Dp → MVCD; no S–T DP → ± MVCD; ST → lower EF; AP → higher 2 yr mortality
Starling et al. (1981)	57/MI	14 ± 2 days	T	S_s	(PI) S–T Dp + AP/BP → MVCD + LVD
Sami et al. (1979)	200/MI (none with LVD)	3, 4, 7, 9, 11 wk; 6, 9, 12 mo	T	S_s	(PI) S–T Dp → early fatal events (>S–T Dp, >risk); VPBs on repeated testing → nonfatal events (VPBs on initial testing are less significant)
Ericcson et al. (1973)	100/MI	3 wk	T	HR (130)	(PI) VPBs
Granath et al. (1977)	205/MI	1st 3 wk, 2nd 9 wk	1st T, 2nd B	S_s/HR (140)	(PI) VPBs
Weld et al. (1981)	236/MI	16 ± 4 days	T	9 min/S_s	(PI) ex time <6 min; VPB's
Dillahunt et al. (1979)	28/MI	14 days	T	5 min/S_s	(PI) ex time <5 min; AP; VPBs → MVCD & LVD
Stein et al. (1980)	47/MI	17 ± 2 days	B	9 min	(C) test influenced clinical management in 47% of patients
Starling et al. (1981)	89/MI	1st $\bar{a}$ D/C, 2nd 6 wk	T	S_s	(C) S–T Dp reproducible test to test; AP, BP, VPB's vary test to test
Haskell and DeBusk (1979)	24/MI	1st 3 wk, 2nd 7 wk, 3rd 11 wk	T	S_s	(C) S–T Dp, AP, VBP's reproducible test to test
DeBusk and Haskell (1980)	200/MI	1st 3 wk, 2nd 3 wk	T	1st HR (130), 2nd S_s	(C) similar data from HR- and S_s-limited testing
CABG surgery patients:					
Wei-i et al. (1979)	37/CABG (ex hypotension $\bar{a}$ surgery)	1st $\bar{a}$ CABG, 2nd $\bar{x}$ 3.7 mo $\bar{p}$ CABG	T	S_s	(C) CABG reverses ex hypotension

Table 1. Summary data from low-level exercise tests in patients with recent myocardial infarction, angina pectoris, and coronary artery bypass surgery—Continued

Study	Patient no./classification	Test			Conclusion and/or prognostic indicator for future cardiac events or MVCD
		$\bar{x}$ time $\bar{p}$ event	Mode	End point	
Ormand et al. (1981)	20/CABG	1st $\bar{a}$ CABG, 2nd several mo $\bar{p}$	B	S_s	(C) work load & RPP increase $\bar{p}$ CABG; S–T Dp developed in 100% $\bar{a}$ CABG but only 13% $\bar{p}$ CABG
Rod et al. (1982)	86/CABG	14 ± 2.6 days	T	S_s	(C) LLET in CABG patients is feasible and safe; fatigue is end point in 88% patients
Johnston (1982)	66/CABG				
	Gp I = 32 patients	Gp I 8.7 days	1–B	75% MHR or S_s	(C) B & T LLET $\bar{p}$ CABG yield comparable data
	Gp II = 34 patients	Gp II 8 days	II–T		

Notes: $\bar{x}$ = mean, $\bar{p}$ = after, MVCD = multivessel coronary disease, MI = myocardial infarction, B = bicycle, S_s = sign/symptoms, C = conclusion, AP = angina pectoris, HR = heart rate (beats/min), PI = prognostic indicator, Dp = depression, T = treadmill, METS = metabolic equivalents (3.5 ml/kg/min), MHR = maximal heart rate, → = indicate, ↑ = elevation, LLET = low-level exercise test, EF = ejection fraction, BP = inadequate blood pressure response to exercise, LVD = left ventricular dysfunction, VPBs = ventricular premature beats, ex = exercise, $\bar{a}$ = before, D/C = hospital discharge, CABG = coronary artery bypass grafting, RPP = heart rate × systolic pressure, GP = group.
Source: From Johnston BL: Exercise testing for patients after myocardial infarction and coronary bypass surgery: Emphasis on predischarge phase. *Heart Lung,* 13:18, 1984. By permission.

tion, but the test is typically done using a bicycle ergometer or a motor-driven treadmill. Leg muscle fatigue may limit the performance if patients are unfamiliar with bicycle riding; walking on a treadmill, for these individuals, may be a better index of functional capacity. The patient is customarily tested to the limit of tolerance, that is, to a sign- or symptom-limited exercise end point, with the proviso that a cardiac patient should not exceed the maximum predicted heart rate for age. Even for patients with symptomatic atherosclerotic coronary heart disease, exercise testing is documented to be acceptably safe when appropriate guidelines are followed (31–33). McHenry (31) calculated the morbidity and mortality associated with exercise testing in recent years to be 1.8 myocardial infarctions per 10,000 tests and 0.25 cardiac death per 10,000 tests.

True "maximal" exercise tests are characterized by a failure to further increase oxygen consumption with an increase in workload; in clinical practice, the oxygen consumption is often not measured, but the limit of activity tolerance is used instead. However, in patients with symptomatic coronary disease, rather than a maximal test, a symptom- or sign-limited procedure is typically used. This means that the patient stops exercising because of chest discomfort, dyspnea, fatigue, claudication, palpitations, and so forth; or that the physician terminates the test because of an inappropriate response of the heart rate, blood pressure, or electrocardiogram—arrhythmia, conduction abnormality, repolarization abnormality, and the like (figure 3).

Each stage of exercise testing is characterized by a defined workload, described in terms of oxygen consumption or METS. This value can be translated into

1. Symptoms
    - Angina pectoris
    - Inappropriate dyspnea (? L.V. failure)
    - Fatigue, dizziness, lightheadedness, confusion
    - Leg pain
2. Signs
    - Cyanosis, pallor, mottling
    - Cold sweat, piloerection
    - Ataxia, glassy stare
    - Gallop heart sounds
    - Murmur of valve regurgitation
    - Abnormal cardiac impulse
3. Heart Rate
    - Excessive tachycardia
    - Bradycardia } Chronotropic incompetence
    - Failure to increase } Chronotropic incompetence
4. Blood Pressure
    - Failure to increase } Inotropic incompetence
    - Progressive decrease } Inotropic incompetence
5. ECG
    - ST segment elevation or depression < Time of onset / Configuration
    - Arrhythmia — Frequent or multiform PVC
        — Ventricular tachycardia
        — Atrial tachyarrhythmias
        — A-V block > first degree

Figure 3. Abnormal responses to exercise. Source: Wenger NK: Rehabilitation of the Patient with Symptomatic Coronary Atherosclerotic Heart Disease, Part I. In: *Cardiology Series,* edited by HD McIntosh, Continuing Education, Baylor College of Medicine, Houston, 1980, p. 18.

comparable occupational and recreational activities. However, most occupational work is intermittent, characterized by brief periods of more strenuous activity and long intervals of low-level activity. Because cardiac output, blood pressure response, and oxygen uptake do not approach the steady state until about 2 minutes after the onset of work, occupational myocardial work demand tends to be lower than that apparent for the same level of steady state exercise as determined at exercise testing. This explains why individuals with modest cardiac impairment, patients with limitation of their cardiac output, can tolerate significant workloads of short duration when adequate rest periods are interspersed. For example, the cardiac patient categorized as Class II in the New York Heart Association classification system (34), who can sustain up to 2.5 calories/min of continuous effort, can perform intermittent effort up to 4 calories/min; comparable values for the Class III cardiac patient are 2 and 2.7 calories/min, respectively. In general, when extrapolating the results of exercise testing to full-time work, recommendations for most cardiac patients are that they can work at a level of about 30% of their physical work capacity during an 8-hour day (35). Most coronary patients will develop symptoms when working at 65% of their maximal physical work capacity for prolonged periods of time. There are available tabular or graphic presentations that help define comparable intensity levels of daily living activities, occupational activities, and recreational activities (table 2, figures 4–7).

Rather than the more formal exercise testing cited earlier, physicians often use walking ability to gauge their patients' functional capacity. For example, walking at a speed of 3–3½ miles/hour entails a work intensity of 4–5 METS; most patients who can walk without difficulty at this pace should be able to perform most sedentary desk or bench jobs, as these entail an energy expenditure of 3 or 4 METS. Indeed, in a recent survey (36), it became apparent that most physicians used clinical judgment and their assessment of the patient's

Table 2. Approximate energy requirements of selected activities

Category	Self-care or home	Occupational	Recreational [1]	Physical conditioning
Very light >3 METS >10 ml/kg/min >4 kcal	Washing, shaving, dressing Desk work, writing Washing dishes Driving auto	Sitting (clerical, assembling) Standing (store clerk, bartender) Driving truck [1] Crane operator [1]	Shuffleboard Horseshoes Bait casting Billiards Archery [1] Golf (cart)	Walking (2 mph) Stationary bicycle (very low resistance) Very light calisthenics
Light 3–5 METS 11–18 ml/kg/min 4–6 kcal	Cleaning windows Raking leaves Weeding Power lawn mowing Waxing floors (slowly) Painting Carrying objects (15–30 lb)	Stocking shelves (light objects) [2] Light welding Light carpentry [2] Machine assembly Auto repair Paper hanging [2]	Dancing (social and square) Golf (walking) Sailing Horseback riding Volleyball (6 man) Tennis (doubles)	Walking (3–4 mph) Level bicycling (6–8 mph) Light calisthenics
Moderate 5–7 METS 18–24 ml/kg/min 6–8 kcal	Easy digging in garden Level hand lawn mowing Climbing stairs (slowly) Carrying objects (30–60 lb) [2]	Carpentry (exterior home building) [2] Shoveling dirt [2] Pneumatic tools [2]	Badminton (competitive) Tennis (singles) Snow skiing (downhill) Light backpacking Basketball Football Skating (ice and roller) Horseback riding (gallop)	Walking (4.5–5 mph) Bicycling (9–10 mph) Swimming (breast stroke)
Heavy 7–9 METS 25–32 ml/kg/min 8–10 kcal	Sawing wood [2] Heavy shoveling [2] Climbing stairs (moderate speed) Carrying objects (60–90 lb) [2]	Tending furnace [2] Digging ditches [2] Pick and shovel [2]	Canoeing [2] Mountain climbing [2] Fencing Paddleball Touch football	Jog (5 mph) Swim (crawl stroke) Rowing machine Heavy calisthenics Bicycling (12 mph)
Very heavy >9 METS >32 ml/kg/min >10 kcal	Carrying loads upstairs [2] Carrying objects (>90 lb) [2] Climbing stairs (quickly) Shoveling heavy snow [2] Shoveling 10/min (16 lb)	Lumberjack [2] Heavy laborer [2]	Handball Squash Ski touring over hills [2] Vigorous basketball	Running (≥6 mph) Bicycle (≥13 mph or up steep hill) Rope jumping

[1] May cause added psychologic stress that will increase workload on the heart.

[2] May produce disproportionate myocardial demands because of use of arms or isometric exercise.

Source: From Haskell WL: Design and implementation of cardiac conditioning programs, in Wenger NK, Hellerstein HK (eds): *Rehabilitation of the Coronary Patient.* New York, John Wiley, 1978, p. 203. By permission.

Figure 4. Energy costs of housework in Kcal/minute. Source: Wenger NK: Rehabilitation of the Patient with Symptomatic Coronary Atherosclerotic Heart Disease, Part I. In: *Cardiology Series,* edited by HD McIntosh, Continuing Education, Baylor College of Medicine, Houston, 1980, p. 9. (From Gordon. By permission.)

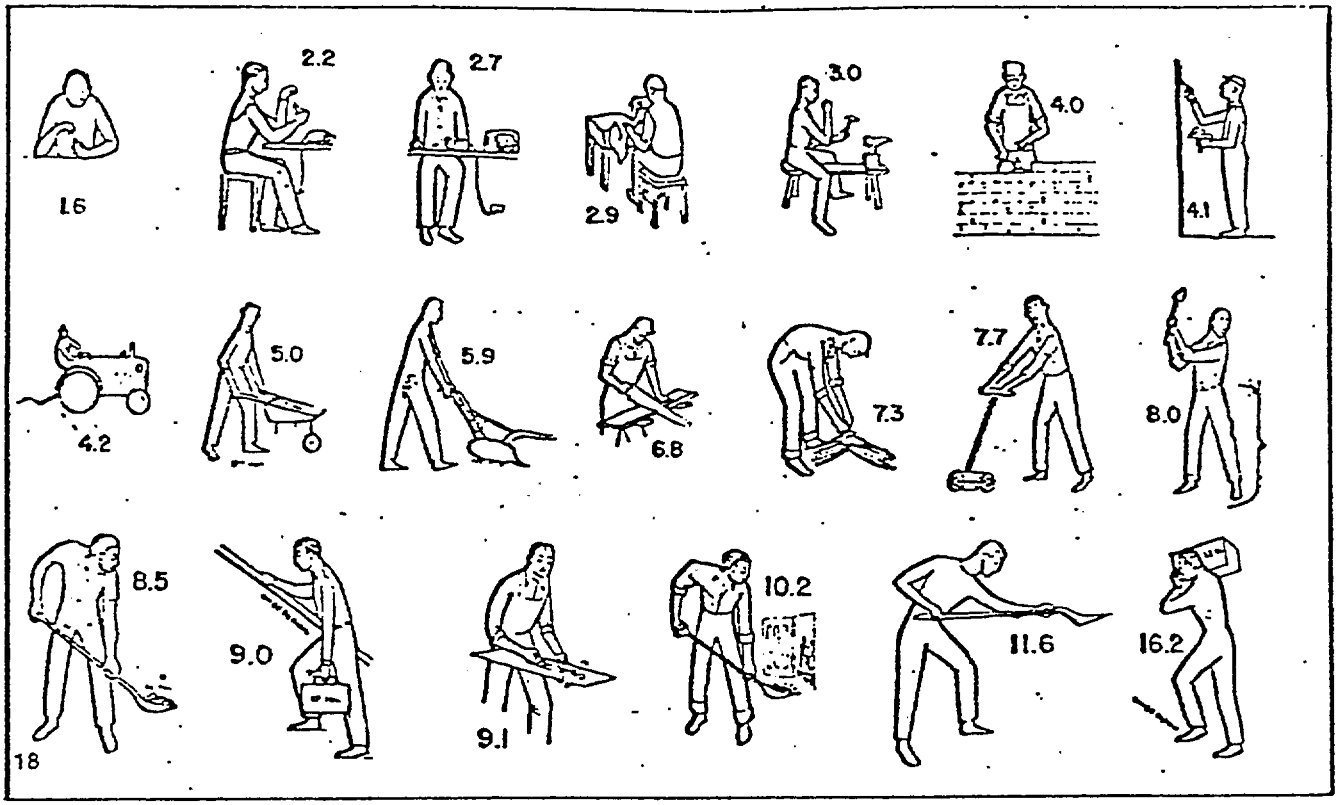


Figure 5. Graphic presentations such as those pictured are used by many physicians to define for their patients the comparable intensity levels of various activities. Occupational activities are shown here. Source: Wenger NK: Rehabilitation of the Patient with Symptomatic Coronary Atherosclerotic Heart Disease, Part I. In: *Cardiology Series,* edited by HD McIntosh, Continuing Education, Baylor College of Medicine, Houston, 1980, pp. 20–21. (From Gordon. By permission.)

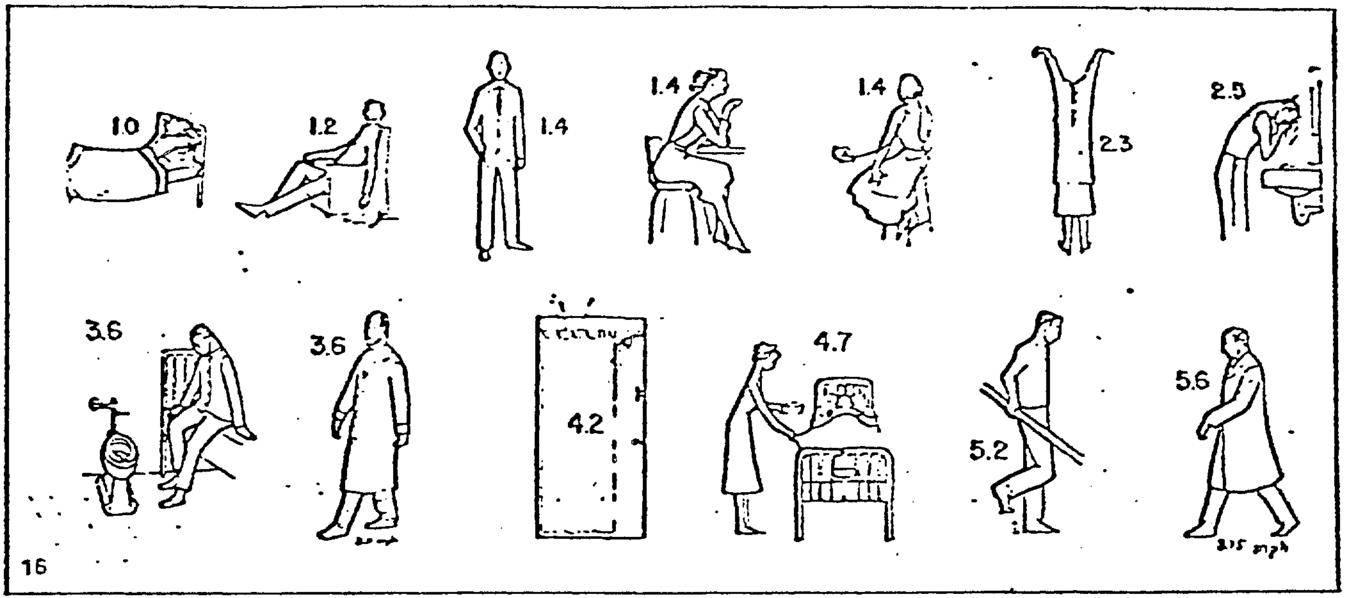


Figure 6. Graphic presentations showing self-care. Source: Wenger NK: Rehabilitation of the Patient with Symptomatic Coronary Atherosclerotic Heart Disease, Part I. In: *Cardiology Series,* edited by HD McIntosh, Continuing Education, Baylor College of Medicine, Houston, 1980, pp. 20–21.

tolerance to walking to recommend resumption of activities after myocardial infarction, including recommending return to work.

Standards for exercise testing of cardiac patients define the need for trained personnel and appropriate equipment for emergency cardiac care, including cardiopulmonary resuscitation (33). However, there has not been definition of the optimal community facilities for exercise testing and their appropriate staffing, with a determination as to whether they should differ qualitatively or quantitatively from those available in major medical centers. The community facilities for exercise testing could assess functional status and help quantify work capacity rather than having the exercise test serve as a diagnostic modality. The signing of an "informed consent" document is recommended before exercise

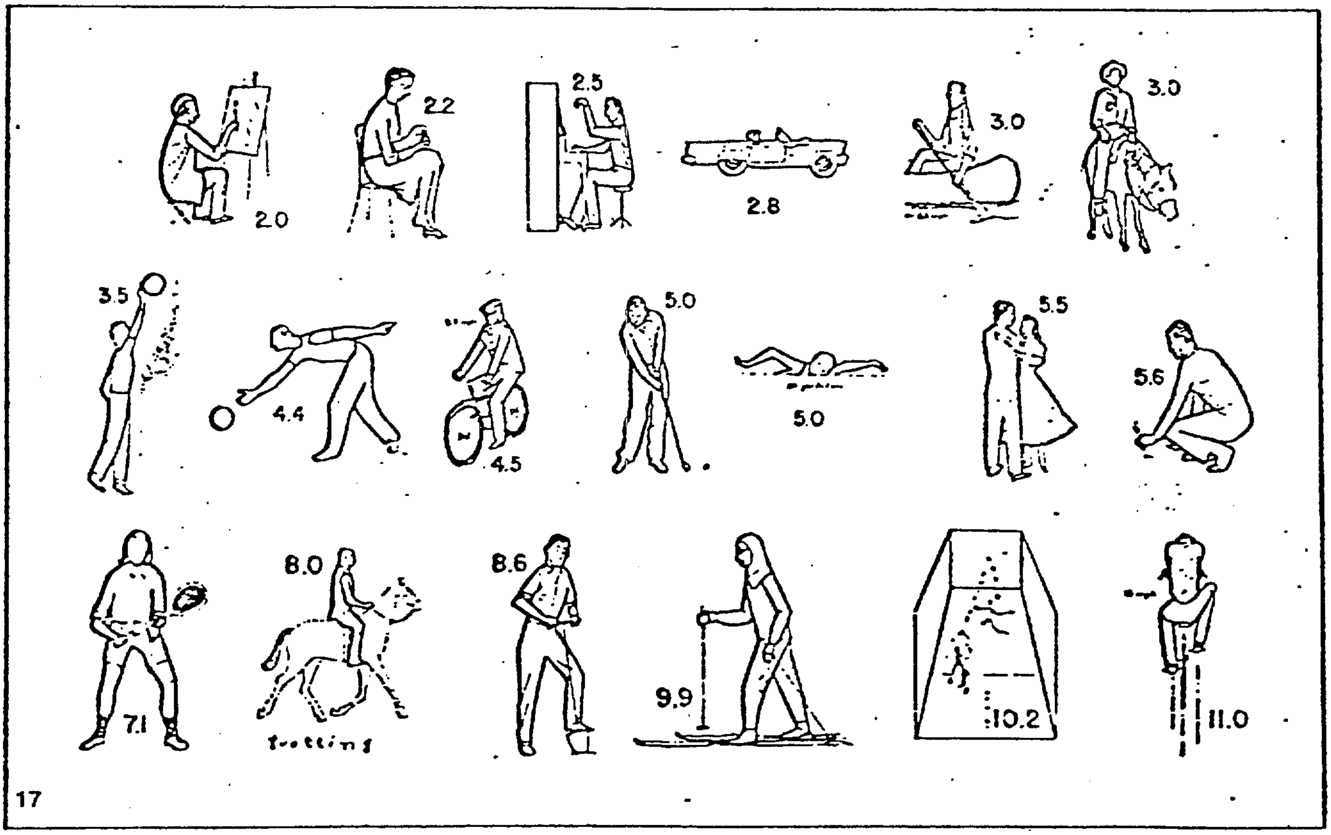
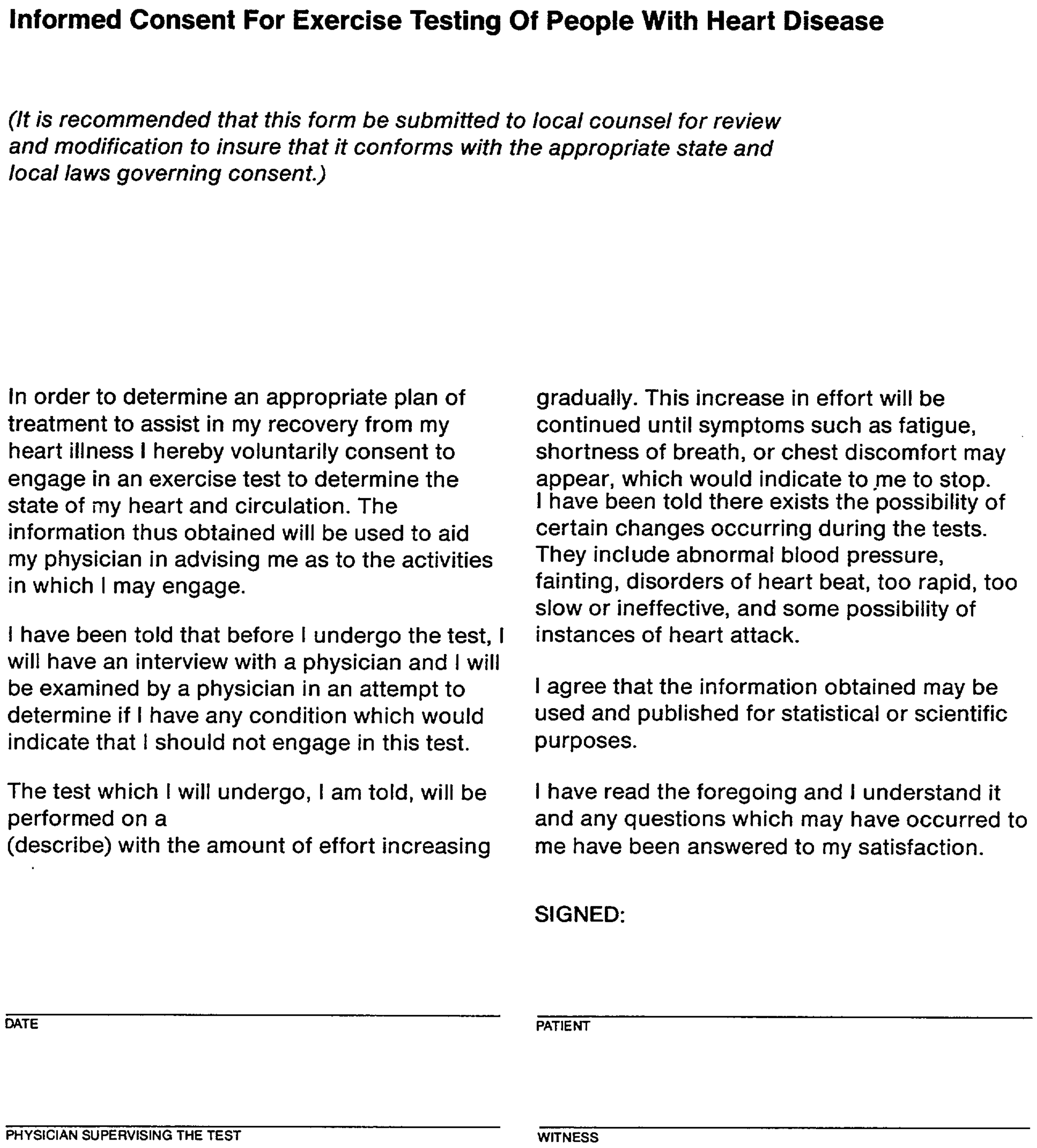


Figure 7. Graphic presentations showing recreational activities. Source: Wenger NK: Rehabilitation of the Patient with Symptomatic Coronary Atherosclerotic Heart Disease, Part I. In: *Cardiology Series,* edited by HD McIntosh, Continuing Education, Baylor College of Medicine, Houston, 1980, pp. 20–21.

testing is performed. The sample document recommended by the American Heart Association, shown in figure 8 (33), can be adapted for NHANES to indicate that the test is designed to assess the activity capability or capacity of the person tested and that the data will be used in the survey as well.

In assessing fitness and activity tolerance, both arm and leg exercise testing must be done, as the effect of training on their performance is only modestly interchangeable. This is important because most occupational as well as recreational activities entail predominant arm rather than leg work, mandating specific arm or leg exercise testing to assess performance ability. Because myocardial oxygen demand and cardiac output are greater with arm than with leg work of equivalent intensity (despite a comparable total body oxygen demand), arm exercise testing can more precisely identify the capability to perform arm work (17).

An important concern in exercise testing relates to the changes on the exercise electrocardiogram. It is currently accepted that the predictive accuracy of a test reflects to a great extent the prevalence of the disease in the population tested. Thus, exercise testing in an asymptomatic population with a low prevalence of disease



Informed Consent For Exercise Testing Of People With Heart Disease

(It is recommended that this form be submitted to local counsel for review and modification to insure that it conforms with the appropriate state and local laws governing consent.)

In order to determine an appropriate plan of treatment to assist in my recovery from my heart illness I hereby voluntarily consent to engage in an exercise test to determine the state of my heart and circulation. The information thus obtained will be used to aid my physician in advising me as to the activities in which I may engage.

I have been told that before I undergo the test, I will have an interview with a physician and I will be examined by a physician in an attempt to determine if I have any condition which would indicate that I should not engage in this test.

The test which I will undergo, I am told, will be performed on a
(describe) with the amount of effort increasing gradually. This increase in effort will be continued until symptoms such as fatigue, shortness of breath, or chest discomfort may appear, which would indicate to me to stop.
I have been told there exists the possibility of certain changes occurring during the tests. They include abnormal blood pressure, fainting, disorders of heart beat, too rapid, too slow or ineffective, and some possibility of instances of heart attack.

I agree that the information obtained may be used and published for statistical or scientific purposes.

I have read the foregoing and I understand it and any questions which may have occurred to me have been answered to my satisfaction.

SIGNED:

DATE

PATIENT

PHYSICIAN SUPERVISING THE TEST

WITNESS

Figure 8. Sample informed consent document. From The Exercise Standards Book, American Heart Association, 10–79–25M, p. 12. By permission.

is more likely to result in a false positive exercise test result than is the identical test performed in a population at high risk of having the specific medical problem (Bayes' theorem). Therefore, a potential problem with the widespread (and/or screening) use of exercise testing and exercise ECG monitoring is that, in populations likely to have a low prevalence of disease, a false positive test, and particularly a false positive exercise ECG, may subsequently considerably influence the planned diagnostic procedures and therapeutic interventions, as well as generate considerable costs. Further, the significant amount of anxiety, as well as the subsequent costs generated by the performance of other noninvasive tests or invasive procedures such as coronary arteriography to evaluate the results of exercise testing, may prove unacceptable. This aspect is of major concern in NHANES in the exercise testing of persons who believe and report that they have cardiovascular disease, when, in fact, this is not the case.

In the survey previously cited (36), when physicians were asked to identify which clinical or laboratory measurements they usually depended on to guide recommendations for activity and vocational planning for patients recovered from an uncomplicated acute myocardial infarction, 97% of all specialty groups cited clinical judgment, with assessment of the patient's tolerance to walking noted by 92%; these guidelines were not mutually exclusive. However, standard exercise testing is increasingly employed by all physician specialty groups, and multifactorial regression equations have been developed that may improve the accuracy of predicting functional capacity from treadmill testing (37). Standard exercise tests served as a guideline for 59% of family and general practitioners and internists and for 83% of cardiologists to guide recommendations for activity and make recommendations for return to work. Compared with the 1970 questionnaire (38), there was a substantial increase in the use of standard exercise tests by all physician specialties. The motor-driven treadmill was the most common apparatus, used by 80% of general and family practitioners, 83% of internists, and 93% of cardiologists. This suggests that treadmill testing is a reasonable approach to assessing functional capacity in NHANES, one that will be considered appropriate by the community physicians. Bicycle ergometer tests were used by fewer than 7% of physicians. The Master Two-Step Test was used by 12% of general practitioners, 8% of internists, and 2% of cardiologists. This test should not be considered for cardiac patients in NHANES. Because the intensity of the test is fixed, based on age and sex, rather than entailing serial steps or stages of progressive activity, the test may be submaximal for fit cardiac patients with minimal impairment but supramaximal, even at onset, for unfit or significantly impaired cardiac patients.

Exercise testing has been described as being advantageous in permitting a more precise assessment of function in coronary patients. It was advocated by the World Health Organization Expert Committee on Rehabilitation (39) to "estimate probable performance in specific life and occupational situations" for patients with cardiac disease. However, it must be emphasized that the demonstrated physical work capacity at exercise testing reflects not only the impact of the infarction on cardiovascular function but also the preinfarction fitness of the patient and the extent of deconditioning that occurred during the hospitalization, during convalescence, and during long-term surveillance. The results of exercise testing can identify the physical work capacity of the patient and help define the tolerated levels of heart rate and blood pressure under standardized conditions; this permits a far more precise assessment of work capability (40). Appropriate exercise testing correlates moderately well with vocational and recreational pursuits, but consideration must be given to differences in temperature, environment, relationship to meals, intellectual demands, emotional stress, clothing worn when working, etc., when comparing standardized exercise testing with real-life activities. It also must be appreciated that myocardial oxygen demand and cardiac output appear greater with arm than with leg work of equivalent intensity, although the total body oxygen demands are comparable; thus, arm testing is necessary to evaluate the activity tolerance for arm work and to establish correlations with the ability to do arm work (41–43).

When exercise testing is performed for exercise prescription, particularly in patients with symptomatic coronary disease, patients are tested on their optimal medical regimen, the one on which they are to be trained, to help ensure that their exercise prescription is appropriate (1). This suggests that a comparable approach should be used in the NHANES exercise testing of coronary patients (and probably other cardiac patients) to assess their functional status. However, it must be appreciated that both the exercise capacity and the exercise test response of patients with atherosclerotic coronary heart disease may be altered by a number of commonly used cardiovascular drugs, owing to interactions of hemodynamic and electrophysiologic changes. It will be necessary to assess the effects of drug therapy both in evaluating the results of exercise testing and in assessing the appropriate level of a physical activity regimen (3, 4).

Arm exercise testing

Many patients cannot perform the standard exercise test based on dynamic leg exercise, whether on a treadmill, a bicycle, or using steps, because of peripheral vascular, musculoskeletal, neurologic, or other complications. In these patients, arm exercise testing may be of particular value; it may be more readily correlated with the ability of these patients to perform arm work, but this is often a significant component of both their vocational and their recreational activities.

However, arm exercise produces a higher heart rate and systolic blood pressure than leg exercise and thus is characterized by increased myocardial work at any submaximal workload; mechanical efficiency is less in that there is more oxygen uptake at any given workload with arm than with leg exercise. Further, there is only limited correlation between the peak workload that can be performed with arm and with leg exercise. Thus, although arm testing can be an effective alternative test procedure to evaluate patients for the presence of myocardial ischemia, it probably cannot be used to estimate systemic functional capacity because of the large interindividual variations in the heart rate-blood pressure responses and in mechanical efficiency (44). It may, however, be of value in providing an objective level against which the perceived ability to perform arm work can be evaluated, as well as the actual described performance of arm work.

Psychosocial consequences of physical activity and their influence

Patients who exercise often have an improvement in self-confidence, self-esteem, and well-being and are more likely to view themselves as capable and effective adults. On standard psychometric tests, they show less depression, fear, and dependency (45); these features may, in turn, affect their subsequent exercise performance. Exercising patients are found to participate increasingly in leisuretime activities, and, in some studies, have been found to have a better work attendance record; in addition, an increased return to sexual activity is described (46). Good exercise capacity resulting from exercise training has also been described as associated with improvements in work capacity, in income, and in job responsibility (21). These data are derived predominantly or exclusively from patients with atherosclerotic coronary heart disease. Emphasis on psychosocial features is appropriate, as many patients after myocardial infarction and coronary bypass surgery are more disabled by the psychologic than by the physiologic consequences.

Many psychosocial outcomes after an acute coronary episode or following coronary bypass surgery appear related to the patient's perception of his or her health status, a feature that may be favorably altered by appropriate education and counseling. An important component of this counseling is a discussion of resumption of sexual activity (47–49). The time from a coronary event to resumption of sexual activity is typically related to the presence or absence of cardiac symptoms (47). However, even in patients without angina pectoris, a discussion that extrapolated exercise test data to the cardiac work demand of sexual intercourse increased the patients' perception of their ability to accomplish this activity (50). In patients with recent myocardial infarction, exercise training was associated with an improvement in resumption of sexual activity; a lesser effect was described with exercise training following a remote myocardial infarction (21).

The psychologic state has an important influence on the patient's perception of illness and of physical capabilities. Many patients remain anxious about exerting themselves physically for fear of precipitating pain, a recurrent coronary episode, or sudden cardiac death, and thus their level of activity may be less than that of which they are capable. This limits their return to the prior active lifestyle, whether it involves return to occupation, to leisure activities, to sexual activity, or to social pursuits; their reported activity may be far less than their capability in all these diverse areas. This fear of the incomplete nature of their recovery and of their excessive vulnerability to recurrent infarction or sudden death may limit activity to the extent that the physiologic consequences of immobilization may appear and confound the clinical presentation. At exercise testing, the limited exercise capability may reflect both their underlying disease and the prolonged curtailment of their activities. The same features are encountered in patients recovering from coronary bypass surgery as occur following myocardial infarction (51, 52).

Interventions designed to improve the patient's perception of his or her physical capabilities have had important benefits after uncomplicated myocardial infarction (50). In recent studies, the performance of leg exercise (treadmill exercise testing) did not alter the patient's self-estimated ability to perform lifting or pushing, whereas arm and leg strength exercise testing increased the self-perception of the ability to lift and push but had no effect on the perceived ability to walk, jog, or climb (53, 54). Further, the physician's counseling about the intensity comparability of test procedures and real-life activities has been shown to extend the patient's perceived ability to engage in apparently dissimilar occupational or recreational activities (50) and has been associated with an improved resumption of work (55).

The spouse's belief in the patient's ability to perform in various roles is also a feature that may influence activity performance. Although, in a recent study, husbands underestimated their own abilities to perform treadmill exercise following myocardial infarction, wives rated their husbands' abilities at even lower levels than the men did; wives who watched their husbands perform an exercise test and then walked on the treadmill themselves increased their perception of their husbands' ability to be physically active to an equal extent as performing the exercise test did for their husbands. Participating in treadmill testing may be an effective way to reassure spouses concerning their partners' ability to safely resume physical activity (56). Patients who are encouraged by their families, as well as by the medical staff, to function more independently may also improve their exercise status.

In view of these considerations, it appears important to correlate the patient's perceptions about the risks and benefits of exercise with the reported activity level and the actual performance at exercise testing.

Vocational aspects

Earlier assessments of patients with cardiovascular disease in regard to the severity of their disability showed that 36% were classified as having severe and 64% less severe disability (57). For the coronary patient, the presence or absence of complications of acute myocardial infarction cannot be directly equated with subsequent severe and less severe disability; nevertheless, it is likely that a substantially greater proportion of those whose acute illness was characterized by complications will be in the more severely disabled group. As evidence of activity capability, the return to productive or employed roles varied from 65–95% for less severely disabled cardiac patients to 25–33% for severely disabled persons. However, because more patients are limited by the psychologic than by the physiologic consequences of myocardial infarction (55, 58, 59), the analysis of disability data does not offer an appropriate reflection of activity capability. Only rarely, in the patient with marginal cardiac function, does the severity of the angina pectoris or congestive heart failure preclude or delay return to work, particularly to the relatively low-intensity workload of most jobs in today's mechanized society (60). Indeed, in a European study of coal miners and factory workers, patients recovered from myocardial infarction were compared with normal individuals holding the same jobs; the postinfarction patients used an increased percentage of their maximal oxygen uptake, 42–47%, compared with 39% for normal individuals, but they still had an adequate cardiac reserve (61).

In one series of patients studied 3 months after myocardial infarction, the average maximum working capacity was reported as 70% of that of a similarly aged reference group (62); although one-third of the patients were limited by angina pectoris and at least one-third experienced cardiac symptoms at a workload so low that they were not considered able to perform some of their daily life activities, 42% of patients limited their physical exertion because of fear rather than because of objective physical limitations. It also must be appreciated that these data and many of the earlier descriptions of symptoms limiting physical activity and limiting the performance of the patient at exercise testing relate to the care of coronary patients prior to the advent of current medical therapy. The availability and widespread use of long-acting nitrate drugs, beta adrenergic blocking agents, and the newer calcium antagonist drugs provide far greater symptomatic relief of angina pectoris and may enable increased activity performance. The widespread application of aortocoronary bypass surgery has had even greater implications regarding enhancement of functional status; however, the decrease in symptoms and improvement in function after coronary bypass surgery have not been associated with an increase in return to work (63–65). Only preliminary data are available regarding coronary angioplasty, functional status, and return to work, and the number of patients who have undergone coronary angioplasty is currently so limited as to make assessment in NHANES III unrealistic.

Patients in the New York Heart Association functional Class I (34), as shown in figure 9, can exercise at a workload of at least 7 METS and have little or no cardiac impairment; they remain asymptomatic when performing usual daily or occupational activities that are characterized by moderate exertion. The degree of cardiovascular impairment for these patients ranges from 0–15%. Class II patients, who become symptom limited with moderate and/or prolonged physical exertion, have a work tolerance of 5 or 6 METS as evidence of their modest cardiac impairment. Their degree of cardiac impairment ranges from 20% to 40%. Class III patients, typically those with severe angina or congestive heart failure, have a work tolerance in the 3 to 4 METS range and thus become symptomatic when performing usual daily tasks except at an extremely slow pace; for these individuals, only sedentary work is indicated. Their degree of cardiac impairment ranges from 50% to 70%. Class IV patients, symptomatic at rest, may have evidence of congestive heart failure, have a work tolerance of 2 METS or less, and are not considered employable (66); their degree of cardiac impairment ranges from 80% to 95%. However, a problem with the New York Heart Association classification is that patients with longstanding cardiovascular disease and its functional limitations decrease their perception of "ordinary" activity. Many of these patients progressively restrict their activity as they become more symptomatic, so that their definition of ordinary activity may falsely suggest a better New York Heart Association classification than is actually the case; their ability to perform clearly defined tasks with known MET equivalents may better reflect their functional capacity (67). This interrogation regard-

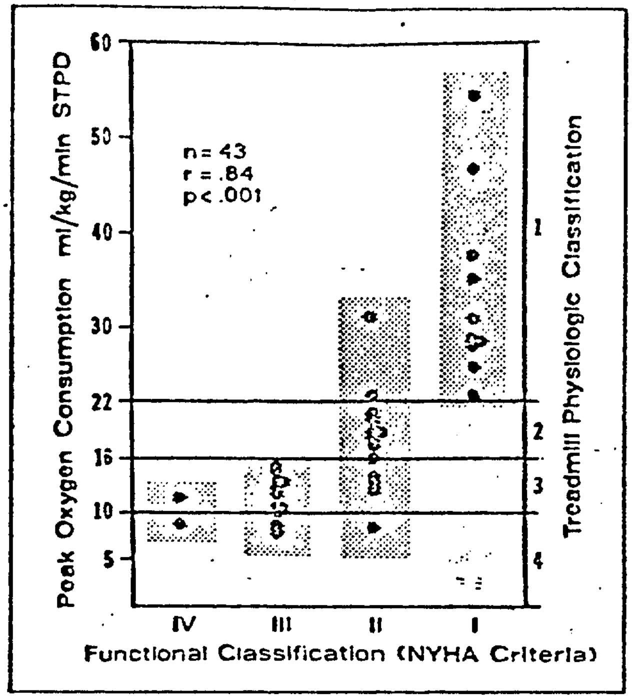


Figure 9. Relationship of treadmill physiologic classification to clinical functional classification. Source: Wenger NK: Rehabilitation of the Patient with Symptomatic Coronary Atherosclerotic Heart Disease, Part II, In: *Cardiology Series*, edited by HD McIntosh, Continuing Education, Baylor College of Medicine, Houston, 1980, p. 24. (from Naughton & Haider. By permission.)

ing specific tasks, rather than ordinary activity, should characterize NHANES III activity questionnaires for cardiac patients.

An important component of activity evaluation involves knowing the energy costs of a representative sample of occupations and work activities as currently performed. Data from prior years may not be applicable because of the changes in technology and improvement in the work environment, e.g., air conditioning, ventilation, and mechanization. A research component of the community rehabilitation program involving cardiac work evaluation units and physical reconditioning programs in North Carolina (68) studied the energy costs of representative occupations and work activities indigenous to the area; comparable data can be used to relate exercise test performance to actual work performance and to perceived ability for activity performance.

Leisure activities

For cardiac patients, more physical energy is often expended in recreational and leisure activities than in vocational activities. Because health professionals often have more difficulty in assessing the work requirements and skills needed for avocational and recreational activities, many patients may have these activities arbitrarily and inappropriately limited by medical advice; therefore it is important to determine whether patients do not engage in these activities because they are not able to do so or because it has been inappropriately recommended that they refrain from doing so.

Left ventricular dysfunction or congestive heart failure and exercise testing

Recent data suggest that the level of left ventricular function bears little or no relationship to the ability to perform physical activity. There is poor correlation between the ejection fraction as determined at the time of coronary arteriography or by radionuclide methods and the results of exercise testing used to infer functional classification. For example, there can be a decrease in the demonstrated ability to exercise in the presence of a well-maintained ejection fraction; on the other hand (as is often seen with the administration of some of the newer vasodilator and angiotensin-inhibitory drugs), there can be no improvement in exercise performance even though the ejection fraction increases, cardiac output improves, and there is a decrease in the left ventricular end diastolic pressure. Exercise performance seems significantly dependent on the changes in skeletal muscle blood flow, in that only when there is evidence of improved organ blood flow (in this instance, blood flow to skeletal muscle) is there an improvement in exercise performance and in maximal oxygen uptake. The decrease in symptoms appears to result from changes in organ blood flow in the periphery rather than from changes in cardiac output. The reduced exercise capacity of most patients with chronic congestive heart failure appears primarily related to impaired or limited blood flow to skeletal muscle, with resultant muscular fatigue with exercise (69).

Selected patients with severely impaired left ventricular function may have an adequate exercise capacity, and, indeed, some may even improve their exercise capacity by gradual appropriate, supervised physical activity training (70). Important peripheral adaptations appear to have beneficial effects on the functional capacity of these patients.

Exercise testing in patients with valvular heart disease

Mitral stenosis

As mitral valvular stenosis, characteristically of rheumatic etiology, becomes progressively more severe, there is a serial decrease in exercise performance; a good correlation is observed between the severity of mitral valvular obstruction and the decline in exercise tolerance (71). Perhaps the major exception is that, with the onset of atrial fibrillation, there is a precipitous decline in exercise tolerance related both to the tachycardia (which limits ventricular filling) and to the loss of the atrial kick that further decreases the stroke output. Additionally, patients with a controlled ventricular response to atrial fibrillation (typically those adequately treated with digitalis) reach their maximal heart rate at a low to moderate workload and then tend to have no further increase in heart rate despite an increase in workload (72).

In the patient with mitral stenosis, left atrial pressure rises progressively with increasing intensities of exercise, whereas the cardiac output tends to remain fixed or may even decline; this is because patients with significant mitral stenosis characteristically cannot maintain their stroke volume during higher levels of exercise because of decreasing diastolic flow time. The rise in left atrial pressure, transmitted to the pulmonary vascular bed, causes the dyspnea. Thus both the symptoms (fatigue, weakness, and exertional dyspnea) and the duration of exercise correlate well with the severity of mitral valvular obstruction (73).

Particularly in the patient with mitral stenosis, it is important to correlate the reported limitation of work capacity with objective (exercise test) data in that patients with mitral stenosis with progressive severity of their mitral valvular obstruction often decrease their work effort, eliminate selected higher level activities, or perform their work far more slowly and thus are unaware of progressive disability until it interferes with the performance of minimal daily activities. The inability to reach stage 3 of a Bruce protocol has been described to be associated with critical mitral stenosis in patients with isolated mitral valvular disease (74).

Mitral regurgitation

There is some correlation between the degree of mitral valvular regurgitation and the duration of treadmill exercise, but exercise limitation is often not documented in patients with early evidence of exercise-induced left ventricular dysfunction (75).

Mitral valve prolapse

Patients with mitral valve prolapse often have chest pain, develop ST segment abnormalities at exercise testing, and indeed may manifest supraventricular and/or ventricular ectopy with exercise; thus the exercise test results may mimic those of atherosclerotic coronary heart disease and, indeed, the two entities may coexist (76). In the exercise testing of patients with mitral valve prolapse in a survey examination, these features must be appreciated.

Aortic stenosis

A major concern in using exercise testing as a general screening procedure for cardiovascular disease is that this test is contraindicated in persons suspected to have hemodynamically severe aortic stenosis; exercise-induced syncope and/or sudden cardiac death may occur secondary to an inability to increase the stroke volume and to an actual or relative decrease in the cardiac output with exercise. In the younger patient, a basal systolic murmur with delayed peaking, a delayed and small volume upstroke to the carotid pulse, and an associated history of chest pain, heart failure, or syncope should suggest that exercise testing is unwise; similarly, the diagnosis of aortic stenosis, whether congenital or acquired, should require that verification of the degree of severity be performed prior to exercise testing.

Patients who have even modest severity of aortic valvular obstruction may have ST segment changes on the electrocardiogram that probably reflect unmet myocardial oxygen needs. However, it is impossible to differentiate these abnormalities from those resulting from the potentially associated atherosclerotic coronary heart disease.

Aortic regurgitation

An impairment of exercise tolerance is not an adequate way to assess the severity of aortic regurgitation, as both the history of exercise limitation and the objective documentation thereof occur late in the natural history of chronic aortic regurgitation; exercise tolerance is often well preserved even in the presence of an elevated left ventricular end diastolic pressure and/or a diminished left ventricular ejection fraction. This is because, with exercise, the decrease in peripheral vascular resistance favors forward flow, reducing the regurgitant volume; and the decrease in diastolic time during which the aortic regurgitation can occur, as tachycardia develops with exercise, further explains this disparity (77).

Exercise testing in patients with congenital heart disease

Exercise testing in congenital heart disease has been done most commonly in children or adolescents, but limited exercise test data are available for adults both with corrected and with uncorrected congenital heart disease. For example, a disproportionately large number of patients, both prior to and after the closure of a large atrial septal defect, have a low functional capacity, thought to be caused by a decrease in left ventricular stroke work (78). Supraventricular tachyarrhythmias are often present and may have their onset many years after corrective surgery. The same is the case with moderate- or large-size ventricular septal defects; in the latter, elevation of pulmonary artery pressures in response to exercise may be the responsible feature (78).

Despite adequate surgical correction of aortic coarctation, some patients have significant exercise-induced hypertension, particularly if the coarctation was corrected at an older age.

Tetralogy of Fallot is probably the most commonly encountered cyanotic congenital abnormality, both corrected and uncorrected, in the adult age group. A significant number of patients continue to have a low exercise capacity even after corrective surgery; one explanation is that time and possibly exercise training are required for the left ventricle to fully develop (79). Major arrhythmias are also encountered in the postoperative patient with tetralogy of Fallot, despite complete correction of the defects.

Exercise testing in patients with disturbances of cardiac rhythm or with conduction abnormalities

Patients with sinus node dysfunction may develop a wide variety of exercise-induced rhythm disturbances—both ventricular and supraventricular tachyarrhythmias, bradyarrhythmias, and progressive atrioventricular block; additionally, latent heart failure may become evident if there is underlying myocardial dysfunction (80).

The occurrence of supraventricular tachyarrhythmias related or unrelated to cardiovascular disease may limit the ability to exercise. However, limitation of exercise capacity may also be the result of drug therapy used to control the arrhythmias, particularly the beta blocking agents, which blunt the chronotropic effect of exercise. Patients with congenital complete atrioventricular block often increase both their atrial and ventricular responses to exercise, and many patients with varying degrees of atrioventricular block will have this problem reversed when exercise overcomes the vagal tone. These features are often responsible for the normal functional capacity of persons with variable atrioventricular block.

Ventricular ectopy may be increased or decreased with exercise testing. It is frequently present in normal individuals and therefore cannot be used to differentiate between normal persons and those with cardiovascular disease or to identify a specific cardiovascular disorder. Furthermore, and quite important, is that the reproducibility of exercise-induced ventricular arrhythmias is extremely limited (81).

Exercise testing in patients with miscellaneous cardiovascular disorders

Patients with hypertrophic cardiomyopathy, either obstructive or nonobstructive, may develop exercise-induced hypotension and arrhythmias, and caution is warranted in considering exercise testing for these individuals; medically supervised exercise testing is probably warranted if this procedure is undertaken. Exercise testing is contraindicated in the acute phase of myocarditis, but this is unlikely to pose a major problem in survey examinations.

Some patients with borderline hypertension, however, may have a significant increase in blood pressure with exercise testing, with a potential increase in the risk of a cerebrovascular or cardiovascular complication.

Exercise testing and current disability guidelines for patients with cardiac disease: Social Security Administration

The current guidelines for determining medical impairment for Social Security Administration disability determinations for patients with cardiovascular disorders are undergoing major revision. These revisions address clinical and exercise test data as well as other laboratory data.

Nevertheless, the contemporary decisions are based on the following formulations (82) published in 1979 for the guidance of the medical community, describing the characteristics of persons with cardiovascular diseases who are thought unable to engage in substantial remunerative employment. "Severe cardiac impairment is considered to result from one or more of three consequences of heart disease: (1) congestive heart failure; (2) ischemia (with or without necrosis) of heart muscle; and (3) conduction disturbances and/or arrhythmias resulting in cardiac syncope. The criteria for evaluating impairment resulting from heart disease or diseases of the blood vessels are based on symptoms, physical signs, and pertinent laboratory findings" (82). In the patient with ischemic heart disease, in addition to the clinical impression that the chest discomfort is of cardiac origin, confirmatory treadmill exercise data are required. One of the following, occurring at an exercise level of 5 METS or less, is considered evidence of impairment: "(1) Horizontal or down-sloping ischemic depression of the ST segment to 1.0 mm or greater, clearly discernible in at least two consecutive complexes which are on a level baseline in any lead; OR (2) premature ventricular systoles which are multiform or bidirectional or are sequentially inscribed (three or more); OR (3) ST segment elevation to 3 mm or greater; OR (4) development of second or third degree heart block." An alternative confirmatory exercise test is considered to be the "Double" Master Two-Step Test demonstrating one of the following: "Ischemic depression of ST segment to more than 0.5 mm lasting for at least 0.08 second beyond the J junction and clearly discernible in at least two consecutive complexes which are on a level baseline in any lead; OR development of a second or third degree heart block." In addition, a series of resting ECG abnormalities and angiographic data are considered confirmatory.

Exercise test data are not used to determine impairment in other forms of cardiovascular disease. Indeed, the instructions regarding the purchase of exercise testing are as follows: "Exercise tests should not be purchased in the absence of alleged chest pain of cardiac origin" (82).

Summary and recommendations

This paper addresses the exercise testing and the physical activity assessment of persons with selected cardiac conditions. Advantages of exercise testing as an objective measure of physical work capacity are that it is a relatively safe procedure and one within the capability of most patients; the cost is limited; it can be performed in many ambulatory settings, including hospitals, clinics, and private practice offices; and it is a test generally available in the community at large. Further, it is a test that lends itself to serial repetition, as well as to comparison with a considerable literature of exercise test data, often correlated with cardiovascular disease information.

Assessment of the physical activity of coronary patients has been performed using a variety of questionnaires, many of which are cited in the text. It appears important to correlate this described activity level with the objectively determined exercise capacity. However, it is important that the exercise capacity at testing reflects both the habitual activity level and the disease-related impairment. The former may be altered by choice of lifestyle, by the degree of perceived disease-related impairment, by the actual disease-related impairment, by appropriate or inappropriate activity restriction recommended by physicians, or by a combination of these features. To what extent these components can be delineated from questionnaire data remains unknown but is a challenge for the third National Health and Nutrition Examination Survey (NHANES III); a pilot study may help assess the feasibility of these determinations.

In considering exercise testing as a screening procedure for cardiovascular disease, it must be appreciated that no abnormality encountered at exercise testing is "diagnostic" of any specific type or etiology of cardiovascular disease. Exercise test abnormalities may reflect a limitation of exercise capacity, an abnormality of electrical function (arrhythmia), the development of changes correlated with myocardial ischemia, or the occurrence of other features limiting the cardiac output. Any and all of these can be encountered with a variety of cardiovascular disorders and, indeed, are not diagnostic or specific even for cardiovascular diseases as a group.

Formal low-intensity exercise testing prior to discharge from the hospital after myocardial infarction (predischarge exercise testing) has gained increasing acceptance in recent years. This test has been used for both diagnostic and prognostic determinations but may have equally important applications in assessing the activity capabilities of a variety of physically impaired individuals. Therefore, although the application of low-level predischarge exercise testing has been predominantly for patients hospitalized for an acute coronary episode, this gradually progressive, low-level test appears to be the more suitable format for assessing patients with severe cardiovascular impairment of any etiology and, in particular, may be an appropriate test to document the functional capabilities of patients with congestive cardiac failure of a variety of etiologies. The general guidelines used to select patients for whom predischarge exercise testing is inappropriate after myocardial infarction may also be considered exclusion criteria for exercise testing in NHANES III designed to assess activity capabilities. The pivotal characteristics of the predischarge test that warrant emphasis, then, are the low initial level of effort and the smaller serial increments in workload than are usual with standard exercise tests.

Because of the major concerns of the interrelationships of exercise capacity, cardiovascular disorders, ability to work, and disability determination, it appears important to compare the patient's perceived ability to exercise, the described habitual exercise levels, and the objectively determined exercise capacity as measured at exercise testing, and to correlate these with disability days, physician visits, work-loss days, days of hospitalization, days of limited activity, and indeed the determination of partial or total temporary or permanent disability.

References

1. Wenger NK: Cardiovascular drugs: Effects on exercise testing and exercise training of the coronary patient. In: Wenger NK (ed.): *Exercise and the Heart,* 2d ed, F.A. Davis Co., Philadelphia, 1985, p. 133.
2. Simoons ML, and Balakumaran K: The effects of drugs on the exercise electrocardiogram. *Cardiology* 68 (suppl 2):124, 1981.
3. Ho S W-C, McCormish MJ, and Taylor RR: Effect of beta-adrenergic blockade on the results of exercise testing related to the extent of coronary artery disease. *Am J Cardiol* 55:258, 1985.
4. Mukharji J, Kremers M, Lipscomb K, and Blomqvist CG: Early positive exercise test and extensive coronary disease: Effect of antianginal therapy. *Am J Cardiol* 55:267, 1985.
5. Kannel WB, Gordon T, Sorlie P, et al.: Physical activity and coronary vulnerability: The Framingham Study. *Cardiol Dig* 6:28, 1971.
6. Morris JN, Adam C, Chave SPW, et al.: Vigorous exercise and leisure time and the incidence of coronary heart disease. *Lancet* 1:333, 1973.
7. Haber LD: The epidemiology of disability. II. The measurement of functional capacity limitations. Social Security Survey of the Disabled: 1966. Report No. 10 (Table 4. Estimated 17.74 million), 1970.
8. National Center for Health Statistics, Chronic Conditions and Limitations of Activity and Mobility, U.S., July, 1965–June, 1967. Series 10, No. 61 (Estimated 8.2 million), 1971.
9. U.S. Bureau of the Census. 1970 Census of Population. General Social and Economic Characteristics. Final Report. PC(1)–C1 U.S. Summary. (Table 89. Estimated 11.7 million), 1970.
10. Treitel R: Rehabilitation of disabled adults, 1972. Social Security disability survey 1972: Disabled and non-disabled adults. Report No. 3, 1977.
11. Report of the Working Group on Arteriosclerosis of the National Heart, Lung, and Blood Institute: Summary, Conclusions, and Recommendations. *Arteriosclerosis 1981,* vol. 1, U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, NIH Publication No. 81–2034, 1981.
12. Social Security Administration: Social Security disability applicant statistics 1970, DHEW Publication No. (SSA) 75–11911, 1974.
13. Thocklock RM, Ho SC, Wright H, and Seldon WA: Is cardiac rehabilitation really necessary? *Med J Aust* 2:669, 1973.
14. Schiller E, and Baker J: Return to work after a myocardial infarction. *Med J Aust* 1:859, 1976.
15. Thornley PE, and Turner RWD: Rapid mobilisation after acute myocardial infarction: First step in rehabilitation and secondary prevention. *Br Heart J* 39:471, 1977.
16. Jelinek MV, Ziffer RW, McDonald JG, et al.: Early exercise testing and mobilization after myocardial infarction. *Med J Aust* 2:589, 1977.
17. DeBusk RF: Early exercise testing after myocardial infarction. In Wenger NK (ed): *Exercise and the Heart,* F.A. Davis Company, Philadelphia, 1978.
18. Smith JW, Dennis CA, Gassman A, et al.: Exercise testing three weeks after myocardial infarction. *Chest* 75:12, 1979.
19. Lindvall K, Erhardt LR, Lundman T, et al.: Early mobilization and discharge of patients with acute myocardial infarction. A prospective study using risk indicators and early exercise tests. *Acta Med Scand* 206:169, 1979.
20. Naughton J (for the Project staff): The National Exercise and Heart Disease Project. In: Cohen LS, Mock MB, and Ringqvist I (eds): *Physical Conditioning and Cardiovascular Rehabilitation,* John Wiley & Sons, Inc., New York, 1981, p. 247.
21. Shaw LW: Effects of a prescribed supervised exercise program on mortality and cardiovascular morbidity in patients after a myocardial infarction. The National Exercise and Heart Disease Project. *Am J Cardiol* 48:39, 1981.
22. Council on Scientific Affairs: Physician-supervised exercise programs in rehabilitation of patients with coronary heart disease. *JAMA* 245:1463, 1981.
23. Ericsson M, Granath A, Ohlsen P, et al.: Arrhythmias and symptoms during treadmill testing three weeks after myocardial infarction in 100 patients. *Br Heart J* 35:787, 1973.
24. Sivarajan ES, Lerman J, Mansfield LW, and Bruce RA: Progressive ambulation and treadmill testing of patients with acute myocardial infarction during hospitalization: A feasibility study. *Arch Phys Med Rehab* 58:241, 1977.
25. Weiner DA: Predischarge exercise testing after myocardial infarction: Prognostic and therapeutic features. In: Wenger NK (ed): *Exercise and the Heart,* 2d ed, F.A. Davis Co., Philadelphia, 1985, p. 95.
26. Johnston BL: Exercise testing for patients after myocardial infarction and coronary bypass surgery: Emphasis on predischarge phase. *Heart Lung* 13:18, 1984.
27. Bruce RA, and Hornsten TR: Exercise stress testing in evaluation of patients with ischemic heart disease. *Prog Cardiovasc Dis* 11:371, 1969.

28. Naughton J, Sevelius G, and Balke B: Physiological responses of normal and pathological subjects to a modified work capacity test. *J Sports Med* 3:201, 1963.
29. Hammermeister KE, DeRouen TA, Dodge HT, and Zia M: Prognostic and predictive value of exertional hypotension in suspected coronary heart disease. *Am J Cardiol* 51:1261, 1983.
30. Weiner D, McCabe C, Cutler D, et al.: Decrease in systolic blood pressure during exercise testing: Reproducibility, response to coronary bypass surgery, and prognostic significance. *Am J. Cardiol* 49:1627, 1982.
31. Rochmis P, and Blackburn H: Exercise tests: A survey of procedures, safety and litigation experience in approximately 170,000 tests. *JAMA* 217:1061, 1971.
32. McHenry PL, and Morris SN: Exercise electrocardiography—Current state of the art. In Schlant RC, and Hurst JW (eds): *Advances in Electrocardiography,* vol. 2, Grune & Stratton, New York, 1976, p. 265.
33. *The Exercise Standards Book*, American Heart Association, 70–041–A, Dallas, Texas, 1979.
34. Criteria Committee of the New York Heart Association, Inc.: *Diseases of the Heart and Blood Vessels (Nomenclature and Criteria for Diagnosis),* 6th ed, Little, Brown, Boston, 1964.
35. Astrand PO, and Rodahl K: *Textbook of Work Physiology,* McGraw-Hill, New York, 1970.
36. Wenger NK, Hellerstein HK, Blackburn H, and Castranova SJ: Physician practice in the management of patients with uncomplicated myocardial infarction—Changes in the past decade. *Circulation* 65:421, 1982.
37. Foster C, Jackson AS, Pollock ML, et al.: Generalized equations for predicting functional capacity from treadmill performance. *Am Heart J* 107:1229, 1984.
38. Wenger NK, Hellerstein HK, Blackburn HW, and Castranova SJ: Uncomplicated myocardial infarction: Current physician practice in patient management. *JAMA* 224:511, 1973.
39. Report of a World Health Organization Expert Committee: Rehabilitation of patients with cardiovascular disease. WHO Tech Rep Ser No. 270, Geneva, 1964.
40. Astrand PO, and Rodahl K: *Textbook of Work Physiology,* 2d ed, McGraw-Hill, New York, 1977.
41. DeBusk RF, Valdez R, Houston N, and Haskell W: Cardiovascular responses to dynamic and static effort soon after myocardial infarction: Application to occupational work assessment. *Circulation* 58:368, 1978.
42. DeBusk R, Pitts W, Haskell W, and Houston N: Comparison of cardiovascular responses to static-dynamic effort and dynamic effort alone in patients with chronic ischemic heart disease. *Circulation* 59:977, 1979.
43. Hellerstein HK: Prescription of vocational and leisure activities. Practical aspects. *Adv Cardiol* 24:105, 1978.
44. Blomqvist CG: Upper extremity exercise testing and training. In: Wenger NK (ed): *Exercise and the Heart*, 2d ed, F.A. Davis Co., Philadelphia, 1985, p. 175.
45. Carson P: Activity after myocardial infarction. *Br Med J*, 288:1, 1984.
46. Stein RA: The effect of exercise training on heart rate during coitus in the post myocardial infarction patient. *Circulation* 55:738, 1977.
47. Hellerstein HK, and Friedman EH: Sexual activity and the post-coronary patient. *Arch Intern Med* 125:987, 1970.
48. Stern MJ, Pascale L, and Ackerman A: Life adjustment post myocardial infarction: Determining predictive variables. *Arch Intern Med* 137:1680, 1977.
49. McLane M, Krop H, and Mehta J: Psychosexual adjustment and counseling after myocardial infarction. *Ann Intern Med* 92:514, 1980.
50. Ewart CK, and Taylor CB: The effects of early post-myocardial infarction exercise testing on subsequent quality of life. *Quality of Life and Cardiovascular Care* 1:162, 1985.
51. Heller SS, Frank KA, Kornfeld DS, et al.: Psychosocial outcome following open-heart surgery. *Arch Intern Med* 134:908, 1974.
52. Gundle MJ, Reeves BR, Tate S, et al.: Psychosocial outcome after coronary artery surgery. *Am J Psychiatry* 137:1591, 1980.
53. Ewart CK, Taylor CB, Reese LB, and DeBusk RF: Effects of early postmyocardial infarction exercise testing on self-perception and subsequent physical activity. *Am J Cardiol* 51:1076, 1983.
54. Ewart CK, Stewart KJ, Kelemen MH, et al.: Psychologic impact of circuit weight testing and training in cardiac patients. *Med Sci Sports Exerc* 16:139, 1984.
55. Krasemann EO, and Jungmann H: Return to work after MI. *Cardiology* 64:190, 1979.
56. Taylor CB, Bandura A, Ewart CK, et al.: Exercise testing to enhance wives' confidence of their husband's capability soon after clinically uncomplicated myocardial infarction, *Am J Cardiol* 55:635, 1985.
57. Graham S, and Reeder LG: Social Epidemiology of Chronic Diseases. In: Freeman HR, Levine S, and Reeder LG (eds): *Handbook of Medical Sociology,* 3d ed, Prentice-Hall, Englewood Cliffs, NJ, 1979, p. 71.
58. Wigle RD, Symington DC, Lewis M, et al.: Return to work after myocardial infarction. *Can Med Assoc J* 104:210, 1971.
59. Mulcahy R, and Hickey N: The rehabilitation of patients with coronary heart disease: A comparison of the return to work experience of national health insurance patients with coronary heart disease and of a group of coronary patients subjected to a specific rehabilitation programme. *J Ir Med Assoc* 64:541, 1971.
60. Kjoller E: Resumption of work after acute myocardial infarction. *Acta Med Scand* 199:379, 1976.
61. Denolin H: Personal communication, 1981.
62. Sanne H: Exercise tolerance and physical training of non-selected patients after myocardial infarction. *Acta Scand* (Suppl 551):1, 1973.
63. Wenger NK, and Hurst JW: Coronary bypass surgery as a rehabilitative procedure. In: Wenger NK, and Hellerstein HK (eds): *Rehabilitation of the Coronary Patient,* 2d ed, John Wiley & Sons, Inc., New York, 1984, p. 115.
64. National Institutes of Health Consensus Development Conference Statement. Coronary-artery bypass surgery: Scientific and clinical aspects. *N Engl J Med* 304:680, 1981.
65. Walter PJ (ed): *Return to Work After Coronary Bypass Surgery.* Psychosocial and Economic Aspects, Springer-Verlag, Berlin, 1985.
66. Naughton J, and Haider R: Methods of exercise testing. In: Naughton JP, Hellerstein HK, and Mohler IC (eds): *Exercise Testing and Exercise Training in Coronary Heart Disease,* Academic Press, New York, 1973, p. 79.
67. Goldman L, Cook EF, Mitchell N, et al.: Pitfalls in the serial assessment of cardiac functional status. How a reduction in "ordinary" activity may reduce the apparent degree of cardiac compromise and give a misleading impression of improvement. *J Chron Dis* 35:763, 1982.
68. North Carolina Heart Association: Program summary for the cardiac rehabilitation program: Cardiac work evaluation units, reconditioning, and outpatient counseling 1975–1978, Federal Grant, North Carolina State Grant No. 7503, 1978.
69. Wilson JR, Martin JL, Schwartz D, and Ferraro N: Exercise intolerance in patients with chronic heart failure: Role of impaired nutritive flow to skeletal muscle. *Circulation* 69:1079, 1984.
70. Conn EH, Williams RS, and Wallace AG: Exercise responses before and after physical conditioning in patients with severely depressed left ventricular function. *Am J Cardiol* 49:296, 1982.
71. Chapman CB, Mitchell JH, Sproule BJ, et al.: The maximal oxygen intake test in patients with predominant mitral stenosis. *Circulation* 22:4, 1960.
72. Blomqvist CG: Exercise testing in rheumatic heart disease. *Cardiovasc Clin* 5:267, 1973.
73. Hugenholtz PG, Ryan TG, Stein SW, et al.: Hemodynamic studies: The spectrum of pure mitral stenosis in relation to clinical disability. *Am J Cardiol* 10:773, 1962.
74. Almendral JM, Garcia-Andoain JM, Sanchez-Casco SA, et al.:

Treadmill stress testing in the evaluation of patients with valvular heart disease. *Cardiology* 69:42, 1982.

75. Borer JS, Gottdiener JS, Rosing DR, et al.: Left ventricular function in mitral regurgitation: Determination during exercise. *Circulation* 59,60 (suppl II):38, 1979.
76. Schlant RC, Felner JM, Miklozek C, et al.: Mitral valve prolapse. *Dis Month* 26(10):1, 1980.
77. Boucher CA, Wilson RA, Danarek DJ, et al.: Exercise testing in asymptomatic or minimally symptomatic aortic regurgitation: Relationship of left ventricular ejection fraction to left ventricular filling pressure during exercise. *Circulation* 67:1091, 1983.
78. Cumming GR: Maximal exercise capacity of children with heart defects. *Am J Cardiol* 42:613, 1978.
79. James FW, Kaplan S, Schwartz DC, et al.: Response to exercise in patients after total surgical correction of tetralogy of Fallot. *Circulation* 54:671, 1976.
80. Abbott JA, Hirschfeld DS, Kunkel FW, et al.: Graded exercise testing in patients with sinus node dysfunction. *Am J Med* 62:330, 1977.
81. Faris JC, McHenry PL, Jordan JW, et al.: Prevalence and reproducibility of exercise-induced ventricular arrhythmias during maximal exercise testing in normal men. *Am J Cardiol* 37:617, 1976.
82. Social Security Administration: Disability *Evaluation Under Social Security. A Handbook for Physicians.* U.S. Department of Health, Education, and Welfare. Social Security Administration, HEW Publication No. (SSA) 79-10098, Aug. 1979.

Chapter **14**

Health-Related Fitness of the Older Adult

Everett L. Smith, Ph.D.
Catherine Gilligan, B.A.
University of Wisconsin, Madison

Health-related fitness of older adults

In the general population, functional capacity and health-related fitness decline in many systems of the body after age 35. Many factors, both genetic and environmental, contribute to these declines. It is difficult to separate environmental from genetic contributions to aging declines. Below a critical level of function, regardless of the contributing factors, an older adult can no longer function in his or her environment. Work capacity, cardiovascular function, skeletal function, flexibility, strength, balance, and reaction time are all important aspects of health-related fitness in the older adult. When the skeletal system is compromised to the point at which spontaneous fractures occur, independence is significantly reduced and sometimes completely lost. Similarly, when the cardiovascular system decreases to such an extent that even activities of daily living require maximal effort, independent function is reduced, and often supportive environments are necessary. Strength, flexibility, balance, and reaction time are all important in permitting an individual to move freely and safely and to accomplish the activities of daily living. All of these factors can be improved through physical activity. In the following sections, we have reviewed aging and activity effects on work capacity, flexibility, and bone. In each of these sections, we have provided recommendations for testing the older adult on aspects of fitness and functional capacity. The reviews are not exhaustive but are intended to provide a background on the research in these areas.

Work capacity and cardiovascular function

Introduction

On the average, functional capacity declines with age. It must be kept in mind, however, that a number of factors other than genetically controlled aging contribute to the decline. Much of the difference in fitness between young and elderly adults may be caused by inactivity and disease. A primary goal of research on aging is to determine the relative contributions of genetic and environmental factors to the loss in functional capacity.

The aged population is very heterogeneous. In our laboratory, two apparently "healthy normal" aged men had work capacities of 7 and 14 METS. The man with lower work capacity had coronary disease and a lower level of physical activity than the man with the high work capacity.

Lakatta (1986) studied the effects of occult coronary disease on cardiovascular function and work capacity. In men free of coronary disease, most of the average changes in heart function commonly attributed to age were absent. In other studies, the work capacity of exercising older adults has been found to be similar to or higher than that of sedentary young adults. The average declines in flexibility, strength, and reaction times can also be countered by physical activity.

When doing cross-sectional surveys or longitudinal studies, it is important to clarify the goals of the research: (a) To describe the general functional capacity of the population with and without disease and disuse atrophy limitations; or (b) to define biological aging in the absence of disease and disuse atrophy. Both are valid approaches, but they are not equivalent.

Average changes in cardiovascular physiology with age

Various investigators have reported a decline in cardiovascular system function with age in the general population. To determine whether this decline is environmentally or genetically induced, further research is needed. On the average, maximum work capacity, heart rate, stroke volume, and cardiac output are decreased; resting and exercise blood pressures increase. Not all populations experience the same increase in blood pressure, and dietary factors may account for some of the rise in blood pressure with age in this country. Both anatomical and physiological changes in the heart and the vascular system contribute to these functional declines.

The heart and blood vessels undergo a number of changes in composition detrimental to contractility and elasticity (Harris, 1975; Skinner et al., 1982; Shephard, 1981). As elastin in the heart decreases, fat, connective tissue, and collagen build up (Harris, 1975; Skinner et al., 1982). The functional capacity of the heart is weakened through decreased muscle mass, capillary blood flow, and capillary-fiber ratio (Harris, 1975; Skinner et al., 1982; Shephard, 1981). Vessels lessen in compliance and internal diameter and are compromised by increased lamallae (Harris, 1975); atheromatous change; and calcification and ulceration, which produce rigidity and narrowing (Skinner et al., 1982).

In the average person, the left ventricle decreases and the left atrium increases in size with age (Harris, 1975), and left ventricular work at rest declines (Harris, 1975) as ventricle power output declines because of size and muscularity changes (Shephard, 1985). The ventricular wall is less compliant (Shephard, 1985), and the ventricular relaxation is slowed (Skinner et al., 1982). Preloading of the ventricle may be further reduced by poor venous tone and varicose veins (Shephard, 1985).

ECG (Shephard, 1981) and wall motion abnormalities (Port et al., 1980) are more frequent in the aged. The pre-ejection period (PEP) (Harris, 1975; Shephard, 1981; Shephard, 1985) lengthens from 80 in the young to 95–100 in the aged (Shephard, 1985), and QRS-axis changes are evident (Evans et al., 1982) because of slower tension development, loss of coordination in contractile process, and decreased myofibril ATPase (Shephard, 1978?; Shephard, 1981). Left ventricular ejection time increases 2 msec/decade (Harris, 1975), which may be associated with aortic resistance.

Congruent with decreased muscle mass, ATPase activity in the heart, tissue glycogen, and tension development decline (Harris, 1975; Skinner et al., 1982; Shephard, 1981; Shephard, 1985).

Narayanan (1981) studied ATP-supported calcium transport in the heart of young-old rats. Rates of accumulation were halved in the aged sarcoplasmic reticulum and doubled in the sarcolemma. The combined activity in the older animals was 38–47% of that in young rats. He hypothesized that this may contribute to the increased reaction time in aging hearts.

Some researchers have reported that the heart is less sensitive to catecholamines (Shephard, 1985), atropine, and beta-adrenergic stimulation (Harris, 1975; Rodeheffer et al., 1984) and more sensitive to carotid sinus stimulation (Harris, 1975). There is less endogenous norepinephrine in the myocardium (Harris, 1975) but a greater production of norepinephrine and catecholamines in response to isometric exercise (Skinner et al., 1982). Both parasympathetic and sympathetic drives are reduced (Skinner et al., 1982).

Valvular changes contribute to decreased cardiac functional efficiency with age. Valve thickness and rigidity are increased because of sclerosis and fibrosis (Harris, 1975), and the aortic valve may be calcified. The aorta widens and often shifts to the right (Harris). Stiffness and pulse-wave velocity are doubled from age 20 to age 60, with changes in the mechanical properties of the aorta and pulmonary artery (Gozna et al., 1974). The overall average heart volume is decreased or unchanged (Skinner et al., 1982; Shephard, 1981; Shephard, 1985).

Contractility of the heart and maximal heart rate decline with age (Evans et al., 1982). Resting heart rate is maintained or lowered slightly (Skinner et al., 1982). Heart rate recovery to resting levels after exercise is slowed in aged subjects (Harris, 1975).

Most investigators report that stroke volume decreases with age (Skinner et al., 1982; Bortz, 1982; Shephard, 1981), particularly over the last 20–30% of the individual's work capacity (Shephard, 1985). The loss in maximal stroke volume is estimated at 10–20% (Shephard, 1978?).

Peripheral vascular resistance (Shephard, 1981; Astrand and Rodahl, 1970) increases approximately 1% per year (Harris, 1975) at rest, submaximal exercise, and maximal exercise (Skinner et al., 1982). Plaques, muscle weakness, and decreased vasodilation contribute to the increased resistance (Shephard, 1981). Varicosities, and possibly decreased vasomotor tone, inhibit blood return to the heart (Shephard, 1981).

Maximum cardiac output declines .9% to 1% per year (Harris, 1975), or 20–30% by age 65 (Shephard, 1981). Some investigators report that resting (Skinner et al., 1982; Bortz, 1982; Shephard, 1985) and submaximal exercise cardiac outputs also decline (Skinner et al., 1982); others have found them unchanged (Rodeheffer et al., 1984).

Physical fitness, physical activity, and cardiovascular risk

Mortality of athletes

Kannel et al. (1985) reported that epidemiologic evidence linking physical inactivity to coronary heart disease is substantial but not conclusive. Training improved cardiovascular risk factors in some studies, but ordinary levels of physical activity may not be important determinants of the major cardiovascular risk factors. In prospective epidemiologic studies, level of physical activity was inversely related to overall mortality, cardiovascular mortality, and coronary mortality in men. The effect of physical activity was significant even when other risk factors were taken into account for men but not for women.

Brand et al. (1979) followed a group of 3,975 longshoremen for 22 years. A total of 410 men died from heart attacks. After adjustment for age, race, systolic blood pressure, smoking, body mass index, glucose intolerance, and ECG status, men with a high work activity had a rate of heart attacks approximately one-half that of subjects in the lowest work activity category.

Another longitudinal study of coronary risk was reported by Lie et al. (1985). Male cross-country skiers were divided into three age groups: 26–33 years (n = 35), 43–50 years (n = 48), and 58–64 years (n = 39). A second population of men 40–59 years of age (n = 2,014) were divided into quartiles based on physical fitness and by 5-year age group. At baseline, subjects performed a progressive bicycle ergometer test. After 7 years, subjects answered a comprehensive questionnaire on cardiovascular risk, including cardiac symptoms. Maximum work capacity declined cross-sectionally with age in both populations. Physical fitness was negatively related to all coronary risk factors and death from myocardial infarction.

In a study of Harvard male alumni (aged 35–74), Paffenbarger et al. (1978) found that risk of first heart attack was inversely related to reported energy expenditure. Men below 2,000 kilocalories per week in vigorous activity were at 64% higher risk for heart attack than men with an index over 2,000, and they were at 28% greater risk for all causes of death. Adult exercise was independent of other factors influencing heart attack risk, and peak exertion enhanced the effect of total energy expenditure. Ex-varsity athletes had a lower risk only if they maintained a high physical activity index.

Polednak (1972) studied mortality of athletes and nonathletes who had attended Harvard University from 1880 to 1912 and were born from 1860 to 1889 (n = 6,303). Major athletes consistently had the shortest lives, 1–3 years shorter than those of minor athletes and nonathletes.

College athletes (n = 629) and nonathletes (n = 583) did not differ significantly in life expectancy, cause of death, or type of death in a study by Montoye et al. (1956).

In a study of British men, Rook (1954) compared longevity of Cambridge sportsmen, random nonathletes, and intellectuals (academic distinction). Average age at death varied little, with a slight advantage to intellectuals. Cardiovascular conditions accounted for 31.5–42% of the deaths in sportsmen and 39.9–41.5% in the two groups of controls.

Physical fitness and cardiovascular risk factors

Submaximal exercise performance was measured in 12,866 men at high risk for coronary heart disease in the Multiple Risk Factor Intervention Trial (MRFIT). ST depression was significantly correlated with age, diastolic blood pressure, and serum cholesterol levels; it was negatively correlated with cigarettes per day, body mass index, and heart rate at rest. Physical fitness was evaluated by time on a graded treadmill exercise test to 85% of the predicted maximal heart rate for subjects 35–57 years old. Exercise duration was negatively correlated with age, cholesterol level, body mass index, and resting heart rate; it was positively correlated with cigarettes per day and leisure-time physical activity. The authors noted that the MRFIT selection process resulted in smokers who were significantly younger and had significantly lower levels of other coronary artery disease risk factors than nonsmokers (Crow et al., 1986).

Sedgwick et al. (1984) found only a weak relationship between physical fitness and coronary risk in 1,500 healthy men and women aged 20–65. Maximum oxygen uptake was estimated by a submaximal bicycle ergometer test. Fitter men had lower blood pressure, cholesterol, and triglycerides than less fit men, after control for age, physique, smoking, alcohol use, and stress; fitter women had lower blood pressure than less fit women. Fitness accounted for less than 5% of the variance in these factors.

Physical activity and cardiovascular risk factors

Berry et al. (1980) compared fitness profiles of previously selected and newly selected astronauts. Maximum $\dot{V}O_2$ (ml/kg/min) was 12% lower in the older astronauts (mean age 46.1) than the younger astronauts (mean age 33.5). Compared to a normative population, astronauts had lower triglycerides and higher high-density lipoprotein (HDL) levels, although cholesterol levels were similar. Cholesterol, triglycerides, and low-density lipoprotein (LDL) levels were significantly higher in the previously selected astronauts than the newly selected astronauts, but HDL levels and the cholesterol-to-HDL ratio did not differ between groups.

Men with clinical or electrocardiographic evidence of coronary heart disease participated in a 2-year fitness program and were retested 10 years later, with 46 of the original 260 men participating in the followup. Based on activity in the years since the exercise program, subjects were classified into high, moderate, and low activity groups. At the 10-year followup, the high activity group had lower diastolic blood pressure and higher work load at a heart rate of 150 (on a bicycle ergometer test) than the low and moderate activity groups (Franks et al., 1983).

Gyntelberg and Ohlsen (1973) evaluated the relationship of physical activity to serum cholesterol and fitness in 370 men aged 40–59, randomly selected from a large epidemiological study. Physical fitness and leisure-time activity were negatively correlated with serum cholesterol levels.

Hagan and Gettman (1983) compared maximal aerobic power and serum lipoproteins in 53 male distance runners (mean age 37.2) with 53 controls matched by age, height, body weight, and body fat. The runners had significantly greater maximal aerobic power (63.7 versus 47.4 ml/kg/min) and HDL (50.9 versus 46.5) and lower triglycerides (63.1 versus 96.8). Total cholesterol and LDL plus VLDL levels did not differ between groups.

Zauner (1985) studied 56 middle-aged male joggers (mean age 43.3) and compared the responses of those under and over 42 years of age. There was no significant difference between the two age groups in maximum $\dot{V}O_2$ (ml/kg/min), heart rate, cardiac output at a heart rate of 170, or cardiac index (cardiac output/body surface area). The cardiac index was strongly negatively correlated with running time for 10 km (r = .95). Age was correlated only with submaximal heart rate (r = .24) and 10-km running time (r = .57).

Middle-aged powerlifters (n = 5, mean age 52) studied by Hurley et al. (1987) had significantly higher risk factors for coronary artery disease than runners (n = 10, mean age 56) or sedentary men (n = 9, mean age 56). In the powerlifters, HDL was lower, the cholesterol:HDL ratio higher, and total area under the glucose and insulin tolerance curves higher than in either runners or controls.

Middle-aged white men (n = 7,106, ages 35–59) with primary type II hyperlipoproteinemia were rated by usual level of physical activity at work, outside of work, and the frequency of strenuous labor or exercise. Physical activity by any of these measures was related positively to HDL cholesterol and negatively to triglycerides in a comparison of physical activity groups. Physical activity was predictive of HDL and triglycerides even when age, Quetelet index, total cholesterol, cigarette smoking, alcohol intake, and clinic were included in the prediction. Low-density lipoproteins and total cholesterol were not significantly related to physical activity (Gordon et al., 1983).

Mann et al. (1969) trained 62 men ages 25–60 for 24 weeks, 5 times/week, 40 minutes/session. Maximum $\dot{V}O_2$ increased significantly, by 16%. Serum cholesterol decreased significantly but slightly.

Taylor et al. (1973) evaluated the effects of training on cardiovascular risk factors by randomizing men to treatment and control groups. The exercise group increased significantly in maximum work capacity, but no differences developed between control and treatment groups in total serum cholesterol or blood pressure. Exercise sessions were conducted three times per week, but adherence declined to approximately 50% within the first 6 months of the 15–21 month program.

Kramsch et al. (1981) studied the effects of moderate conditioning on the development of coronary artery

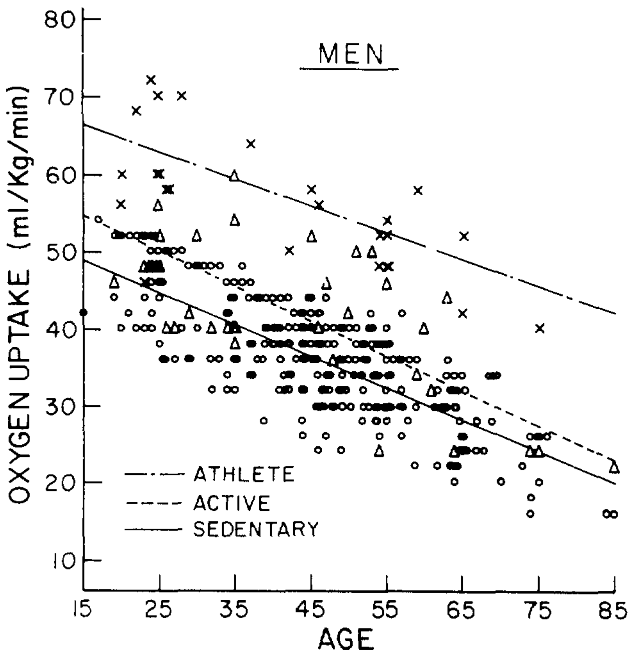


Figure 1. Maximum VO_2 (ml/kg/min) versus age in men. Sources are from table 1. The equations for the three groups are:

General: Maximum VO_2 = 55.6 − 0.423 × age
Active: Maximum VO_2 = 61.4 − 0.453 × age
Athlete: Maximum VO_2 = 71.9 − 0.366 × age

The slopes are not significantly different.

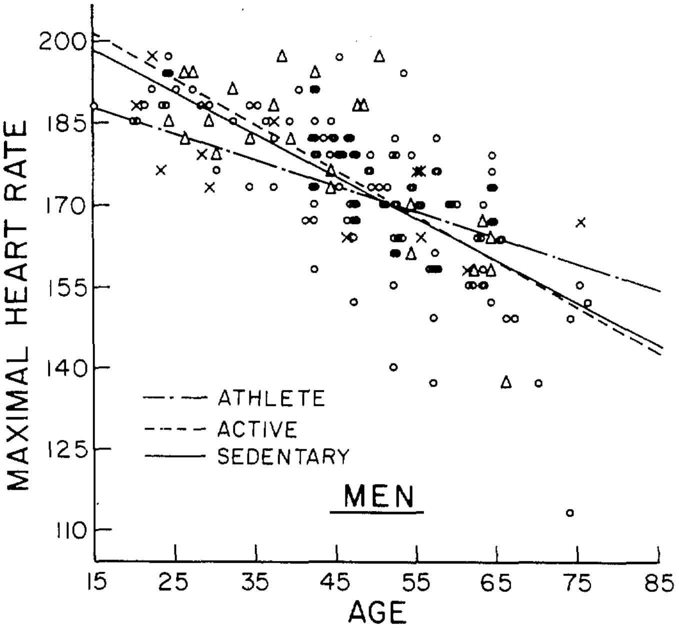


Figure 2. Maximum heart rate versus age in men. Sources are from table 1. The equations for the three groups are:

General: Maximum heart rate = 210.3 − 0.779 × age
Active: Maximum heart rate = 215.1 − 0.847 × age
Athlete: Maximum heart rate = 195.0 − 0.471 × age

The active group does not differ significantly from the general group. Both the slope and intercept are significantly different for the athletes. (p <0.01)

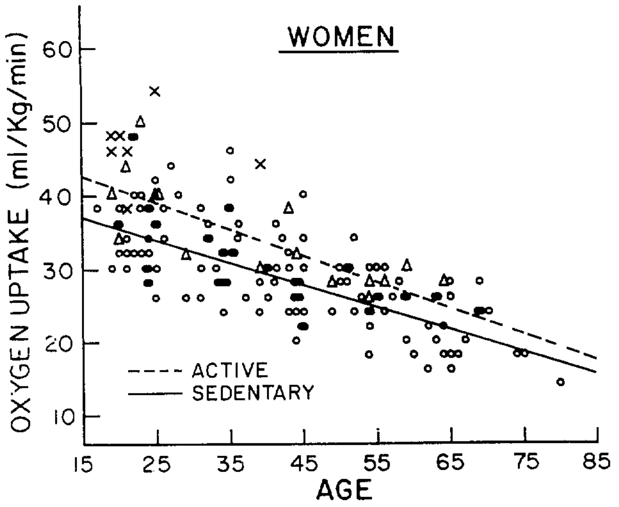


Figure 3. Maximum VO_2 (ml/kg/min) versus age in women. Sources are from table 1. The equations are:

General: Maximum $VO_2 = 42.0 - 0.305 \times \text{age}$
Active: Maximum $VO_2 = 48.5 - 0.370 \times \text{age}$

The slopes are not significantly different.

disease in monkeys on an atherogenic diet. Serum total cholesterol did not differ between control and exercise groups, but HDL was significantly higher and LDL and VLDL significantly lower in the exercise group than in the control group. Exercise was associated with substantially reduced overall atherosclerotic involvement, lesion size, and collagen accumulation; it also produced much larger hearts and coronary arteries.

Zir et al. (1972) evaluated cardiovascular risk factors in 15 male Masters Swimming Championship competitors aged 28–56. Mean cholesterol was 188 mg%, resting blood pressure 128/75, and heart rate 75. Four ECG's, however, were abnormal, despite the high level of conditioning.

Cross-sectional studies of fitness

Buskirk and Hodgson (1987) reviewed the literature, cross-sectional and longitudinal, for rate of change in aerobic power in men and women. Rates of change vary from 0.04 to 1.43 ml/kg/min per year of age in various studies. The authors proposed a curvilinear model for changes in maximum $\dot{V}O_2$ with age, influenced by changes in physical activity level and body weight. In cross-sectional studies, $\dot{V}O_2$ generally appears to decline 0.4–0.5 ml/kg/min per year of age in men and 0.20–0.35 in women. There is no clear distinction between the inactive and active populations in rate of change.

In table 1 and figures 1–4, we have detailed the findings on maximum work capacity and heart rate from a variety of studies. Regressions versus age on the mean values of these studies are similar to the values reported by Buskirk and Hodgson (1987). There are wide variations both within and among studies in maximum $\dot{V}O_2$ (ml/kg/min) and heart rate for any age range.

Cardiovascular function and aging

Seventeen men ages 57–72 with a history of ST segment depression were studied for their response to lifting work and cycling work. Lifting caused higher heart rates and blood pressures and greater ST depression than cycling relative to oxygen uptake (Astrand, 1972).

Benestad et al. (1968) evaluated the work capacity of 20 men and 10 women over age 70 recruited from older adult clubs in Norway. The authors noted that subjects were not randomly selected and were probably fitter than the average population. Women had substantially lower maximum oxygen uptake than men (30%), but maximal heart rates were similar in the two sexes. Pathological ECG's were detected in 40% of the subjects at rest and an additional 20% of subjects during exercise. Oxygen consumption on submaximal loads was higher in old men than in a group of young men evaluated by Hermansen and Andersen (1965). The authors estimated that, compared with sedentary Norwegians in their twenties, maximum oxygen uptake was reduced by approximately 50% in subjects over 70.

Cumming et al. (1973) evaluated the frequency of ECG abnormalities during exercise in women without known heart disease. Subjects exercised maximally or to 86% of maximal heart rate on a progressive bicycle ergometer or treadmill test. ST segment changes were found in 88% of subjects, with no major difference between the 10-year age groups (ages 20–83). For ECG's classed by the authors as abnormal, frequency increased from 25% for subjects 20–39 to 66% for subjects over 60.

Evans et al. (1982) studied the QRS axis in the male adult population of Pukapuka. There was no change with age in the QRS axis among the Pukapukans, although in a British population QRS shifted to the left with age. The authors concluded that the shift in the QRS axis in the Western populations was probably caused by environmental factors.

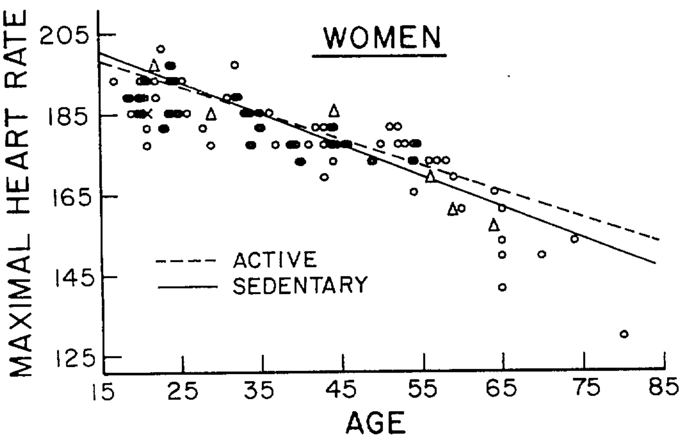


Figure 4. Maximum heart rate versus age in women. Sources are from table 1. The equations are:

General: Maximum heart rate $= 207.9 - 0.733 \times \text{age}$
Active: Maximum heart rate $= 204.7 - 0.629 \times \text{age}$

The two groups are not significantly different.

Table 1. Study findings on maximum work capacity and heart rate

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Adams and deVries, 1973	SB	V	F	6	52–79	66.7	8.7	19.6	4.7		
	SB	V	F	17	52–79	65.9	66.6	18.2	3.3		
Andersen and Hermansen, 1965	B	S	M	38	50–60	54.9		34.9		170.0	
	B	X	M	63	50–66	54.7		48.0		164.0	
Asano et al., 1976	M	X			40–50			49.7			
	M	X			50–60			45.1			
	M	X			60–70			42.2			
	M	X			70			38.9			
Asmussen et al., 1975	SB	A	F	6	22–26	23.5		51.0	11.1		
	SB	A	F	6	48–55	51.3		31.0	2.1		
	SB	A	F	6	60–66	63.2		26.0	2.8		
	SB	A	M	19	21–27	23.9		49.0	8.4		
	SB	A	M	19	41–52	48.7		33.0	4.5		
	SB	A	M	19	53–64	60.7		33.0	5.3		
Astrand, 1952		A	M	42	20–33			58.6	4.1	194.0	10.0
		A	F	44	20–25			48.4	3.2	198.0	10.0
Astrand, 1958	MB	A	M	46	50–54			33.9	5.2	161.0	10.0
	MB	A	M	27	55–59			30.9	4.3	158.0	10.0
	MB	A	M	8	60–64			31.0	4.4	158.0	10.0
Astrand, 1960	MB	A	F	8	20–29	25.0		39.9	4.7	187.0	10.0
	MB	A	F	12	30–39	34.7		37.3	5.2	185.0	7.0
	MB	A	F	8	40–49	43.9		32.5	2.7	178.0	8.0
	MB	A	F	16	50–65	55.6		28.4	2.7	170.0	9.0
	MB	A	M	4	20–29			52.2		186.0	
	MB	A	M	13	30–39			39.8	7.3	181.0	12.0
	MB	A	M	9	40–49			39.2	5.5	173.0	9.0
	MB	A	M	66	50–59			33.1	4.9	161.0	11.0
	MB	A	M	8	60–69			31.4	5.3	159.0	9.0
	MB	A	M	9	56–68					163.0	
Astrand et al., 1973	MB	A	F	35	20–25			47.6	4.0		
	MB	A	F	35	41–46			38.4	4.2		
	MB	A	M	31	20–33			58.7	4.0		
	MB	A	M	31	41–54			45.3	5.6		
Atomi and Miyashita, 1974	MB	S	F	56	20–29			32.4	3.7	186.3	7.5
	MB	S	F	19	30–39			27.9	4.8	176.0	11.8
	MB	S	F	18	40–49			26.0	4.2	173.8	37.8
	MB	S	F	9	50–62			24.4	3.6	172.2	15.9
	MB	A	F	17	20–29			37.5	3.7	186.1	8.2
	MB	A	F	8	30–39			31.9	2.0	178.2	13.3
	MB	A	F	12	40–49			28.8	3.1	179.3	2.6
	MB	A	F	9	50–59			27.3	1.8	173.2	4.8
Atomi and Miyashita, 1980	MB	V	F	7	23–40	31.0	7.0	29.4	3.2	187.9	8.1
Badenhop et al., 1983	MB	V	B	11	60–	67.0		21.1	3.6	156.0	12.0
	MB	V	B	10	60–	70.0		20.6	3.8	160.0	9.0
Bailey et al., 1974	SB	H	M	104	20–29	25.7	2.8	36.4	8.0		
	SB	H	M	163	30–39	33.9	2.7	32.2	6.6		
	SB	H	M	84	40–49	44.1	3.0	26.9	6.1		
	SB	H	M	63	50–59	54.1	2.9	25.7	4.9		
	SB	H	M	33	60–69	64.5	2.5	22.8	5.1		
	SB	H	F	138	20–29	23.8	2.9	30.6	7.2		
	SB	H	F	152	30–39	34.0	3.0	27.8	6.9		
	SB	H	F	93	40–49	44.2	2.9	24.3	5.9		

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Bailey et al., 1974—Continued	SB	H	F	88	50–59	54.2	2.7	21.9	5.6		
	SB	H	F	56	60–69	64.2	2.8	19.0	4.3		
Barry et al., 1969	MB	G	M	18	61–80	70.4	4.7	20.6	3.8	135.8	13.0
	MB	G	F	12	60–77	65.5	4.9	16.1	4.2	140.1	18.1
	MB	X	M	12	21–31	23.2	2.6	45.4	4.1	176.4	10.9
Barry et al., 1966	MB	V	B	8	55–78	69.5	11.0	16.1		139.5	18.4
	MB	V	B	5	58–83	72.6	9.5	16.1		139.2	29.3
Beall et al., 1985a	MB	A	M	8	60–77	66.0	5.0			136.5	12.5
	MB	S	M	25	60–77	66.0	5.0			147.9	22.9
Beall et al., 1985b		A	M	36	50–75					156.0	
		S	M	37	50–79					152.0	
Benestad, 1965	MB	V	M	13	70–81	76.0		27.0	2.5	153.0	13.0
Benestad et al., 1968	MB	V	F	10	70–79	76.0	3.4	19.0	2.5	153.0	13.0
	MB	V	M	20	70–79	76.0	3.3	27.0	3.1	150.0	15.0
Berry et al., 1980	M	A	M	54		46.1	4.1	41.7	8.3		
	M	A	M	29		33.5	3.3	47.5	7.0		
	M	F	M	6		29.3	3.6	36.5	5.9		
Brown and Shephard, 1967	SO	G	F	18	40–49			28.2	5.1		
	SO	G	F	15	50–59			29.6	5.6		
	SO	G	F	11	60–69			22.9	4.6		
Bruce, 1981	MT	A	M		25–34			42.5	5.1	186.0	10.0
(n =	MT	S	M		25–34			36.7	5.6	189.0	
183, 25–34;	MT	A	F		25–34			31.7	4.6	184.0	8.0
812, 35–44;	MT	S	F		25–34			26.1	6.4	176.0	13.0
1,020, 45–54;	MT	A	M		35–44			39.9	5.4	182.0	10.0
285, 55–64)	MT	S	M		35–44			36.6	4.3	184.0	11.0
	MT	A	F		35–44			29.9	5.3	178.0	14.0
	MT	S	F		35–44			24.1	3.2	177.0	12.0
	MT	A	M		45–54			37.0	5.3	176.0	12.0
	MT	S	M		45–54			32.7	4.7	176.0	12.0
	MT	A	F		45–54			27.6	6.2	173.0	10.0
	MT	S	F		45–54			23.1	4.0	175.0	14.0
	MT	A	M		55–64			33.3	4.4	169.0	12.0
	MT	S	M		55–64			29.8	4.8	169.0	12.0
	MT	A	F		55–64			29.7	4.7	162.0	11.0
	MT	S	F		55–64			20.2	4.3	169.0	14.0
Bruce, 1984											
(dyspnea)	MT	CD		337		53.0	7.7	18.1	8.4	145.0	23.0
(fatigue)	MT	CD		3482		48.5	9.0	27.8	9.5	165.0	23.0
(dyspnea)	MT	H		294		48.3	8.6	29.8	8.7	176.0	13.0
(fatigue)	MT	H		4268		46.4	8.7	31.7	8.2	174.0	17.0
Buccola and Stone, 1975	SB	V	M	20	60–79	67.6		23.7			
	SB	V	M	16	60–79	65.4		24.2			
Chiang et al., 1970	MT	G	M	304	10–19			42.5	7.4	189.0	45.0
	MT	G	M	123	20–29			39.5	5.5	187.0	13.0
	MT	G	M	213	30–39			34.3	6.1	174.0	20.0
	T	G	M	287	40–49			28.6	5.4		
	T	G	M	107	50–59			27.9	5.3		
	T	G	M	30	60–69			21.8	5.0		
Cumming and Borysyk, 1972	MB	G	M	22	40–45	42.5	1.7	31.9	4.5	178.0	10.0

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Cumming and Borysyk, 1972—Continued	MB	G	M	14	46–49	47.6	1.0	30.4	3.9	176.0	10.0
	MB	G	M	22	50–55	52.2	1.9	30.0	4.2	174.0	8.0
	MB	G	M	5	56–59	57.5	1.0	30.4	4.6	172.0	14.0
	MB	G	M	2	60–65	64.0	1.4	30.4	2.7	163.0	3.0
Cumming et al., 1973	MT	G	F		20–29					197.0	
(n = 136)	MT	G	F		30–39					192.0	
	MT	G	F		40–49					179.0	
	MT	G	F		50–59					167.0	
	MT	G	F		60–69					158.0	
	MT	G	F		70–79					145.0	
Cunningham and Hill, 1975	SB	V	F	17	21–48	31.0		26.0	6.4		
Cunningham et al., 1987	MT	G	M	100	54–68	62.9		30.3	5.9	155.9	15.7
	MT	G	M	100	55–68	62.5		29.6	5.8	155.5	17.7
Dehn and Bruce, 1972	MT	A	M	26	40–49	46.2	2.2	39.2	5.9		
	MT	A	M	21	50–59	54.6	3.0	37.7	4.5		
	MT	A	M	10	60–69	63.9	1.7	31.3	6.9		
	MT	S	M	15	40–49	45.7	2.1	33.2	4.5		
	MT	S	M	10	50–59	55.0	2.7	31.2	4.7		
	MT	S	M	3	60–69	64.7	1.5	28.8	3.1		
De Vries, 1970	SB	V	M	68		69.0		33.9			
	SB	V	M	8		69.0		33.7			
De Vries and Adams, 1972	SB	G	M	12	15–18	16.7		53.8			
	SB	G	M	12	60–75	69.2		33.6			
Drinkwater et al., 1975	MT	HF	F	16	19–29	21.9	2.4	40.4	2.3	189.8	10.0
	MT	HF	F	10	30–39	35.2	3.2	41.4	3.4	183.0	14.9
	MT	HF	F	7	40–49	44.8	3.7	39.2	2.0	179.0	18.5
	MT	HF	F	6	50–59	52.3	2.7	34.6	6.4	179.0	9.1
	MT	LF	F	10	19–29	22.3	2.8	32.7	2.3	193.6	7.9
	MT	LF	F	14	30–39	34.8	3.4	31.7	3.0	184.8	9.7
	MT	LF	F	13	40–49	43.4	2.9	29.5	2.7	177.2	8.3
	MT	LF	F	6	50–59	52.5	3.4	23.7	3.5	180.2	14.2
	MT	G	F	6	60–68	64.7	2.0	25.6	3.6	153.7	12.5
Drinkwater, 1977	MT		F	6		24.2		31.1	7.0	180.2	11.9
	MT		F	11		35.5		31.1	7.0	180.2	11.9
	MT		F	7		52.8		25.1	5.0	177.8	13.5
Epstein et al., 1981	SB	S	M	203	18–21			43.8	7.1		
	SB	S	M	72	22–30			35.6	7.0		
	SB	S	M	129	31–40			31.1	17.0		
	SB	A	M	164	18–21			45.6	16.6		
	SB	A	M	70	22–30			40.8	10.0		
	SB	A	M	78	31–40			38.7	9.7		
Erikssen et al., 1980	SB	G	M	244	40–44					178.0	7.0
(reached pred. max hr)	SB	G	M	385	45–49					171.0	8.0
	SB	G	M	386	50–54					165.0	9.0
	SB	G	M	294	55–59					160.0	9.0
(did not reach)	SB	G	M	111	40–44					159.0	
	SB	G	M	333	45–49					151.0	
	SB	G	M	64	50–54					141.0	
	SB	G	M	59	55–59					137.0	
Fardy et al., 1976	SB	G	M	106	25–34	32.1	1.7	31.3	8.4		
	SB	G	M	212	35–44	39.4	2.7	27.6	7.8		
	SB	G	M	130	45–54	49.1	2.7	24.6	7.0		
	SB	G	M	58	55–74	59.7	4.2	22.8	6.0		

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Fischer et al., 1965	MB	A	M	7	50–59			23.8	2.4		
	MB	A	M	36	60–69			23.4	1.6		
	MB	A	M	20	70–79			23.1	2.1		
	MB	S	M	48	60–69			20.7	1.4		
	MB	S	M	14	70–79			17.4	2.5		
	MB	S	M	5	80–89			16.1	1.8		
Flint et al., 1974	MT	V	F	7	21–49	31.8	11.1	36.3	7.9	189.0	12.2
Flint et al., 1977	MT	G	F	17		26.1	6.5	37.5	3.7		
	MT	G	F	13		41.0	3.6	35.1	6.1		
	MT	G	F	12		56.2	9.0	30.5	6.2		
Foster et al., 1986	MT	G	F	8	73–86	79.8	4.6	13.1	2.0	128.0	15.0
Froelicher et al., 1974	MT	H	M	44	20–24			40.3	6.0	192.0	10.0
	MT	H	M	117	25–29			36.8	5.9	190.0	10.0
	MT	H	M	99	30–34			36.0	5.2	186.0	10.0
	MT	H	M	113	35–39			33.8	5.1	182.0	11.0
	MT	H	M	59	40–44			34.0	5.0	182.0	10.0
	MT	H	M	68	45–49			33.5	6.0	181.0	13.0
	MT	H	M	19	50–53			34.0	5.5	172.0	10.0
	MT	A	M	72	25–29			39.7	5.7	194.0	9.0
	MT	A	M	97	30–34			40.3	5.8	190.0	10.0
	MT	A	M	16	35–39			38.8	4.4	187.0	11.0
	MT	A	M	6	40–44			37.8	6.4	194.0	6.0
Gordon et al., 1969	MB	X	F	3		19.0		47.9		189.0	
Grimby and Saltin, 1966	MB	X	M	14	42–49	46.0		57.0			
	MB	X	M	15	50–59	54.0		53.0			
	MB	X	M	4	61–68	65.0		43.0			
Grimby et al., 1966	B	A	M	9	45–55	51.0		50.9		171.0	
Grimby et al., 1970	MB	G	M	53		54.0		30.5		172.0	10.0
Hagan and Gettmann	MT	X	M	53		37.2	8.7	63.7	5.1	185.0	12.0
	ST	S	M	53		36.4	8.8	47.4	6.6	185.0	12.0
Hanson et al., 1968	B	V	M	7		48.9		35.8			
Hartung, 1974	MB	G	F	29	33–48	39.8		25.1	3.6	172.6	11.3
	MB	G	F	14	33–39	36.7		26.2	2.9	175.8	9.9
	MB	G	F	15	39–48	43.1		24.0	4.0	169.1	11.5
Havel et al., 1969	O		F	9	0–	21.3	0.3	33.5	5.4	193.0	12.0
	O		F	9	0–	21.4	0.0	32.1	6.3	181.0	26.1
Heath et al., 1981	MT	X	M	16	50–72	59.0	6.0	58.7	2.3	169.0	11.0
	MT	X	M	16	18–27	22.0	2.0	69.0	2.3	197.0	7.0
	MT	S	M	9		50.0	6.0	30.4	2.8	173.0	15.0
	MT	S	M	9		52.0	6.0	36.2	4.4	178.0	14.0
Hermansen and Andersen, 1965	MB	X	M	14		27.7	0.8	71.0	6.8	178.0	6.5
	MB	G	M	12		23.4	1.5	44.0	3.9	189.0	8.3
	MB	G	F	12		22.6		38.0	2.8	203.0	9.7
	MB	X	F	5		25.1	5.8	55.0	3.1	186.0	5.8
Hodgson, 1971			M		20–20			52.8			
			M		21–21			52.4			
			M		22–22			51.9			
			M		23–23			51.5			
			M		24–24			51.0			
			M		25–25			50.6			

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Hodgson, 1971—Continued			M		26–26			50.2			
			M		27–27			49.7			
			M		28–28			49.3			
			M		29–29			48.8			
			M		30–30			48.4			
			M		31–31			48.0			
			M		32–32			47.5			
			M		33–33			47.1			
			M		34–34			46.6			
			M		35–35			46.2			
			M		36–36			45.8			
			M		37–37			45.3			
			M		38–38			44.9			
			M		39–39			44.4			
			M		40–40			44.0			
			M		41–41			43.6			
			M		42–42			43.1			
			M		43–43			42.7			
			M		44–44			42.2			
			M		45–45			41.8			
			M		46–46			41.4			
			M		47–47			40.9			
			M		48–48			40.5			
			M		49–49			40.0			
			M		50–50			39.6			
			M		51–51			39.2			
			M		52–52			38.7			
			M		53–53			38.3			
			M		54–54			37.8			
			M		55–55			37.4			
			M		56–56			37.0			
			M		57–57			36.5			
			M		58–58			36.1			
			M		59–59			35.6			
Hollman and Knipping, 1961	B		M	80	20–40			39.6		176.0	8.0
	B		M	36	41–50			35.7		173.0	9.0
	B		M	42	51–60			29.6		169.0	10.0
	B		M	18	61–70			23.9		164.0	7.0
	B	A	M	127	20–40			51.3		178.0	6.0
Hossack and Bruce, 1982	MT	S	M	6	20–29			45.4	4.2	196.0	12.0
	MT	S	M	7	30–39			41.8	5.7	189.0	13.0
	MT	S	M	35	40–49			37.7	5.6	181.0	11.0
	MT	S	M	28	50–59			34.8	6.1	172.0	10.0
	MT	G	M	19	60–69			29.7	5.6	168.0	17.0
	MT	G	M	3	70–79			17.0	2.4	114.0	11.0
	MT	S	F	9	20–29			37.9	4.2	198.0	12.0
	MT	S	F	33	30–39			28.3	3.4	184.0	10.0
	MT	S	F	39	40–49			25.9	3.3	179.0	12.0
	MT	S	F	22	50–59			24.7	2.8	177.0	12.0
	MT	S	F	1	60–70			18.7		160.0	
Ikai and Kitagawa, 1972	MB	G	F	10	20–21			31.5	3.2	184.0	11.0
	MB	G	F	9	21–22			30.7	4.6	176.0	13.0
	MB	G	F	13	23–24			32.8	3.8	182.0	8.0
	MB	G	F	4	25–27			34.4	3.2	184.0	6.0
	MB	G	M	11	20–21			39.2	4.1	186.0	10.0
	MB	G	M	7	21–22			41.4	3.1	189.0	8.0
	MB	G	M	11	22–24			40.5	5.1	187.0	9.0
	MB	G	M	7	25–31			40.5	4.1	188.0	10.0
Ismail and Montgomery, 1979	MT	V	M	12		36.8	5.9	43.3	9.4	173.5	13.2
	MT	V	M	12		52.9	4.5	42.0	12.5	164.8	13.9

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Julius et al., 1967	MB	G	B	18	18–34					169.8	4.7
	MB	G	B	18	35–49					158.7	2.8
	MB	G	B	18	50–69					139.6	4.7
Kamon and Pandolf, 1972	MT	S	M	1		25.0		45.5		191.0	
	MT	A	M	5	19–34	26.2		50.8	2.5	183.0	11.0
	MT	X	M	5	17–26	20.2		60.8	2.1	185.0	8.0
	MT	S	F	2	17–18	17.5		38.2	0.1	194.0	2.0
	MT	A	F	6	18–23	20.8		44.1	1.5	195.0	7.0
	MT	X	F	4	19–21	20.0		47.7	2.8	184.0	5.0
Kasch et al., 1965	MO		M	18	40–49	45.1		37.3		198.0	
	MO		M	10	50–59	53.1		39.3		195.0	
Kasch et al., 1973	MB	V	M	7	39–60	49.0		32.6	3.3	180.0	
	MB	V	M	6	39–60	49.0		33.8	4.8	174.3	
Kasch, 1976		S	M	17		48.0		29.4		171.0	
		S	M	10		47.0		32.9		170.0	
Kasch et al., 1985	O	A	M	15	32–56	44.6	6.9	44.6	6.7	178.0	
	MT	A	M	15		55.2	6.9	45.2	10.0		
	MT	A	M	15		60.0	6.9	40.2	9.3		
	MT	A	M	13		63.1	6.9	43.1	8.4	166.0	
Kavanaugh and Shephard, 1977	S	X			0–40			49.6			
	S	X			40–50			49.9			
	S	X			50–60			46.0			
	S	X			60–70			41.6			
	S	X			70			29.0			
Kiens et al., 1980	MT	SV	M	24		40.0	3.4	37.8			
	MT	SV	M	13		39.0	5.0	40.2			
Kiessling et al., 1974	MB	V	M	10	46–62	54.0	6.3	34.8	4.4		
	MB	A	M	9	43–66	53.0	9.0	50.9	7.2		
Kilbom, 1971	MB	V	F	12	19–31			36.8	2.4	194.0	2.0
	MB	V	F	8	34–48			31.0	4.8	177.0	6.0
	MB	V	F	13	51–64			26.9	4.0	173.0	7.0
	MB	V	F	10	19–45			33.9	4.1	188.0	9.0
	MB	V	F	10	19–45			34.5	6.6	196.0	6.0
Klissouras et al., 1973	MT		F	14	32–46	40.3		29.9		174.0	
Lester et al., 1967	MB	S	M	10	40–44					190.6	
	MB	S	M	12	45–49					189.0	
	MB	S	M	11	50–54					178.5	
	MB	S	M		55–59					183.0	
	MB	S	M	9	60–					178.7	
	MB	MA	M	15	40–44					191.0	
	MB	MA	M	15	45–49					187.9	
	MB	MA	M	11	50–54					182.0	
	MB	MA	M	5	55–59					176.6	
	MB	MA	M	6	60–					176.6	
	MB	HA	M	2	40–44					181.5	
	MB	HA	M	1	45–49					180.0	
	MB	HA	M	3	50–54					171.0	
	MB	HA	M	2	60–					167.5	
Lie et al., 1985	B	X	M		26–33					174.0	11.0
	B	X	M		43–50					165.0	12.0
	B	X	M		58–64					157.0	11.0
	B	F1	M	101	40–44					168.0	15.0

Table 1. Study findings on maximum work capacity and heart rate—Continued

Source	T	G	S	N	Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
					Range	Mean	SD	Mean	SD	Mean	SD
Lie et al., 1985—Continued	B	F1	M	147	45–49					168.0	15.0
	B	F1	M	140	50–54					155.0	16.0
	B	F1	M	117	55–59					149.0	14.0
	B	F2	M	103	40–44					171.0	11.0
	B	F2	M	146	45–49					165.0	11.0
	B	F2	M	139	50–54					160.0	12.0
	B	F2	M	116	55–59					157.0	11.0
	B	F3	M	100	40–44					174.0	10.0
	B	F3	M	146	45–49					168.0	10.0
	B	F3	M	139	50–54					163.0	10.0
	B	F3	M	114	55–59					158.0	12.0
	B	F4	M	101	40–44					174.0	9.0
	B	F4	M	148	45–49					169.0	8.0
	B	F4	M	139	50–54					164.0	12.0
	B	F4	M	118	55–59					159.0	10.0
MacNab et al., 1969	MB	A	F	24		18.7	0.6.	39.1	5.1		
	MB	A	M	24		20.0	1.2	51.7	5.1		
Maksud et al., 1970	MT	X	F	13	15–30	21.5		46.1	5.7	191.4	10.5
	MT	X	M	10	15–32	20.1		56.1	1.0	187.6	6.6
Mann et al., 1969		V	M	62	25–60			34.0			
Massicotte et al., 1979	ST	V	M	23		35.5	10.4	36.5	4.8		
	ST	V	F	11		32.6	10.3	29.5	3.6		
McArdle et al., 1971	MT	A	F	6		20.3	0.7	33.6	3.7	189.0	8.6
	MT	X	F	6		20.1	0.7	35.7	6.1	191.0	8.6
McDonough et al., 1970	MT	G	M	10	40–44			40.5	4.7	184.0	9.6
	MT	G	M	24	45–49			38.4	5.3	178.0	12.2
	MT	G	M	20	50–54			37.5	5.3	171.0	9.3
	MT	G	M	19	55–59			36.2	5.7	175.0	12.6
	MT	G	M	9	60–64			32.6	4.7	165.0	14.9
	MT	G	M	3	65–69			27.7	4.2	150.0	6.8
	MT	S	M	15	40–49			36.8	5.4	185.0	9.3
	MT	S	M	7	50–59			33.1	5.8	180.0	14.3
	MT	S	M	5	60–69			29.4	5.0	174.0	
	MT	A	M	19	40–49			40.9	4.3	176.0	11.6
	MT	A	M	32	50–59			37.6	5.1	171.0	7.1
	MT	A	M	7	60–69			32.9	4.0	164.0	11.0
Metheny et al., 1942	MT	G	F	17	20–29			40.9		197.0	
	MT	G	M	30	19–23			51.3		194.0	
Michael and Horvath, 1965	MB	G	F	30	17–22	19.4	0.0	29.8	0.0	184.0	0.0
Myrhe et al., 1970		V	M	10		53.1		29.7			
Niinimaa and Shephard, 1978a	MT	V	M	9	60–76	64.4	3.3	33.4	3.1		
	MT	V	M	10	60–76	66.8	2.7	24.6	4.5		
Notelovitz et al., 1986	MT	S	F	13	35–44	39.3	3.9	29.0	4.9		
	MT	S	F	60	45–54	50.8	2.5	27.4	3.9		
	MT	S	F	70	56–65	58.7	3.1	25.9	3.8		
	MT	S	F	20	66–75	69.4	3.1	24.2	4.0		
	SB	S	F	20	35–44	40.5	4.9	29.9	6.5	30.0	45.0
	SB	S	F	47	45–54	49.8	4.9	27.7	8.0		
	SB	S	F	24	56–65	59.3	2.8	25.3	5.3		
	SB	S	F	10	66–75	68.7	2.3	23.3	3.1		
Olree and Stevens	MT	G	F	50	19–29			28.1		193.0	
	MT	G	F	30	30–60			22.4		176.0	

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Olree and Stevens—Continued	SB	G	F	50	19–29			30.5			
	SB	G	F	30	30–60			24.5			
Oscai et al., 1968	MT	V	M	14	26–64	37.0		38.5			
Patton et al., 1983	MT	G	M	64	40–41			39.7	6.6		
	MT	G	M	71	42–43			38.1	5.9		
	MT	G	M	53	44–45			37.6	6.0		
	MT	G	M	37	46–47			37.2	6.6		
	MT	G	M	21	48–49			37.2	5.1		
	MT	G	M	10	50–51			36.6	5.1		
	MT	G	M	4	52–53			36.8	9.8		
	MT	G	M	295	40–53			38.1	6.2		
	MT	G	M	295	40–53			38.1	6.2		
	MT	G	M	295	40–53			38.1	6.2	181.6	9.2
	MT	S	M	140		43.9	3.1	36.1	5.3	182.0	10.0
	MT	MA	M	53		43.5	2.6	39.8	5.4	183.0	9.0
	MT	HA	M	55		43.6	3.1	41.9	6.7	178.0	7.0
Plowman et al., 1979	MT	G	F	5		23.0	4.0	40.5	5.8	182.4	4.4
	MT	G	F	5		27.6	6.5	40.4	7.2	179.8	7.4
	MT	G	F	8		35.7	3.1	36.6	8.2	183.3	12.4
	MT	G	F	8		41.9	4.5	33.9	9.3	179.8	11.9
	MT	G	F	12		43.3	2.8	32.1	5.2	180.8	8.7
	MT	G	F	12		49.8	3.5	30.4	4.8	178.2	8.7
	MT	G	F	6		51.5	1.7	30.8	8.6	180.8	12.0
	MT	G	F	6		58.0	2.9	27.9	9.1	173.6	17.1
	MT	G	F	3		64.7	3.5	27.1	5.0	151.0	15.6
	MT	G	F	3		70.3	3.8	25.0	4.7	149.3	12.6
Pollock et al., 1971	MT	V	M	15	40–56	48.9		29.9			
Pollock et al., 1972	MT	V	M	12	30–45	38.7		36.0	3.0		
	MT	V	M	10	30–45	38.7		38.5	3.8		
Pollock, 1974	M	X	M	11	40–50			57.5		177.8	
	M	X	M	5	50–60			54.4		175.0	
	M	X	M	6	60–70			51.4		163.0	
	M	X	M	3	70			40.0		166.3	
Pollock et al., 1976	MT	V	M	22	49–65	55.0		31.0	4.0		
	MT	G	M	7	49–65			31.5	3.5		
Profant et al., 1972	MT	S	F	35	29–39			28.3	3.5	184.0	10.0
	MT	S	F	42	40–49			25.7	3.6	180.0	12.0
	MT	S	F	24	50–59			24.5	3.0	177.0	12.0
	MT	S	F	1	60			18.7		160.0	
	MT	A	F	7	29–39			31.8	2.4	184.0	9.0
	MT	A	F	16	40–49			29.0	3.7	185.0	8.0
	MT	A	F	17	50–59			25.8	3.5	175.0	12.0
	MT	A	F	2	60–69			27.1	5.8	159.0	1.0
Ribisl, 1969	MT	V	M	15		40.2	5.7	40.1	5.1	181.0	10.0
	MT	A	M	10		39.4	7.2	53.0	4.3	179.0	10.0
Robinson, 1938	T	G	M	13		24.5		48.7		193.0	
	T	G	M	10		35.1		43.1		188.0	
	T	G	M	10		44.6		39.5		180.0	
	T	G	M	10		51.5		38.4		170.0	
	T	G	M	7		63.0		34.5		164.0	
	T	G	M	3		75.0		25.5		156.0	
Robinson et al., 1975	MT	G	M	9		51.0	1.5	35.1			
	MT	G	M	12		41.3	1.2	37.6			

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Robinson et al., 1975—Continued	MT	G	M	12		49.2	1.2	40.1			
Robinson et al., 1976	MT	X	M	13		24.3		71.4			
	MT	FO	M	13	47–68	56.6		41.8			
Saltin and Grimby, 1968	MB	FO	M	10	44–49	47.4		44.0		182.0	
	MB	FO	M	14	50–59	53.6		38.0		175.0	
	MB	FO	M	5	60–67	63.0		37.0		170.0	
Saltin et al., 1969	MB	V	M	42	34–50	40.5		37.5	5.8	190.0	9.1
Schocken et al., 1983	MT	G	B	24	65	72.0	6.2			146.0	17.0
Seals et al., 1984	MT	V	B	11	61–67	63.0	2.0	25.4	4.6		
Sedgwick et al., 1984 (divided by fitness)	MB	LF	M	94	20–65	46.7	10.9	26.0	2.4		
	MB	LF	M	238	20–65	44.3	9.4	31.1	1.4		
	MB	MF	M	332	20–65	38.1	8.4	36.2	1.7		
	MB	HF	M	215	20–65	34.5	7.8	41.7	1.7		
	MB	HF	M	77	20–65	29.7	6.5	48.7	3.6		
	MB	LF	F	88	20–65	44.8	11.4	22.6	2.5		
	MB	LF	F	147	20–65	41.2	10.4	28.5	1.4		
	MB	MF	F	219	20–65	35.8	9.3	33.1	1.5		
	MB	HF	F	169	20–65	31.1	6.9	38.1	1.5		
	MB	HF	F	116	20–65	27.3	5.9	44.4	3.1		
Sheffield et al., 1978	MT	G	F		20–29					194.0	
	MT	G	F		30–39					186.0	
	MT	G	F		40–49					178.0	
	MT	G	F		50–59					166.0	
	MT	G	F		60–69					165.0	
Shephard, 1966 (Canada)		S	M	89	20–30			48.0			
		S	M	24	30–40			40.8			
		S	M	65	40–50			39.1			
		S	M	16	50–60			39.5			
		S	M	8	60–70			25.1			
(U.S.)		S	M	1005	20–30			37.6			
		S	M	333	30–40			36.2			
		S	M	112	40–50			35.7			
		S	M	11	50–60			35.7			
		S	M	16	60–70			32.1			
		S	M	3	70–80			25.5			
(Scandinavia)		S	M	511	20–30			59.1			
		S	M	57	30–40			43.7			
		S	M	377	40–50			44.6			
		S	M	176	50–60			34.5			
		S	M	6	60–70			33.2			
(Other)		S	M	767	20–30			44.9			
		S	M	370	30–40			40.1			
		S	M	136	40–50			35.1			
		S	M	203	50–60			30.4			
		S	M	166	60–70			24.6			
		S	M	25	70–80			20.3			
(Canada)		A	M	38	20–30			55.1			
(U.S.)		A	M	152	20–30			51.5			
		A	M	6	40–50			51.3			
(Scandinavia)		A	M	123	20–30			59.5			
		A	M	24	30–40			53.6			
		A	M	6	50–60			47.1			
(Other)		A	M	170	20–30			47.2			
		A	M	127	30–40			59.1			

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Shephard, 1966		A	M	36	60–70			25.0			
(Other)—Continued		A	M	20	70–80			24.7			
		A	M	3	80–90			21.9			
(All)		X	M	50	20–30			69.1			
		X	M	138	50–60			51.3			
(Canada)		S	F	30	20–30			36.0			
		S	F	9	30–40			38.3			
		S	F	9	40–50			30.0			
		S	F	9	50–60			25.5			
(U.S.)		S	F	38	20–30			30.3			
(Scandinavia)		S	F	45	20–30			42.6			
		S	F	24	30–40			46.3			
		S	F	8	40–50			34.8			
		S	F	16	50–60			30.4			
(Other)		S	F	25	20–30			27.0			
Siconolfi et al., 1985	SB	H	M	57	18–29	24.0	3.0	42.0	6.0		
	SB	H	M	43	30–39	33.0	2.0	36.0	6.0		
	SB	H	M	31	40–49	44.0	3.0	32.0	4.0		
	SB	H	M	31	50–59	54.0	3.0	28.0	4.0		
	SB	H	M	17	60–65	62.0	1.0	22.0	5.0		
	SB	H	F	86	18–29	24.0	3.0	29.0	5.0		
	SB	H	F	54	30–39	34.0	3.0	25.0	5.0		
	SB	H	F	24	40–49	44.0	3.0	21.0	4.0		
	SB	H	F	41	50–59	54.0	3.0	18.0	3.0		
	SB	H	F	17	60–65	62.0	1.0	16.0	3.0		
	MB	A	B	24	19–34	27.0	4.0	31.8	6.9		
	MB	A	B	14	35–50	41.0	5.0	29.8	7.5		
	MB	A	B	10	50– 0	56.0	7.0	21.7	4.1		
Sidney and Shephard, 1977a	MB	V	M	19	60–83	63.7		31.4	4.4	172.0	12.0
	MB	V	F	20	60–83	63.4		26.8	2.9		
Sidney and Shephard, 1977b	SB	V	M	11	60–70	63.9	2.6	21.2	3.4		
	SB	V	F	12	60–70	62.0	3.7	22.4	4.3		
Sidney and Shephard, 1978	SB	V	B	12	60–83	65.0		19.5	3.9		
	SB	V	B	8	60–83	65.0		21.2	7.9		
	SB	V	B	14	60–83	65.0		22.1	3.3		
	SB	V	B	8	60–83	65.0		23.4	3.3		
Siegel et al., 1970	MB	V	M	9	32–59	46.0		24.0	1.7	167.0	4.0
Sinning and Adrian, 1968	MB	X	F	7		20.7		38.7	4.5	184.6	6.9
	MB	A	F	8		20.8		35.2	3.3	187.5	5.4
Sprynarova and Pariskova, 1969	MT	X	F	10	15–25	19.5	3.2	45.9	3.8	191.0	
Sucec, 1969	MO	A	M	21	40–56			36.9	4.2	188.6	13.7
Suominen et al., 1977	SB	V	M	14		69.0		28.9	4.3		
	SB	V	F	12		69.0		27.9	4.3		
Taylor et al., 1973		H	M	11	40–44			34.4	4.2		
		H	M	13	45–49			31.6	5.6		
		H	M	16	50–54			33.1	4.6		
		H	M	12	55–60			30.9	3.8		
		H	M	17	45–49			33.6	4.0		
		H	M	24	50–54			31.7	3.9		
Thomas et al., 1985	MT	V	M	88		62.9	3.0	30.8	5.6	158.0	16.0
	MT	V	M	100		62.6	3.1	29.6	6.0	155.3	24.0

Table 1. Study findings on maximum work capacity and heart rate—Continued

					Age			Max $\dot{V}O_2$ (ml/kg/min)		Heart rate (beats/min)	
Source	T	G	S	N	Range	Mean	SD	Mean	SD	Mean	SD
Tietz and Valentin, 1957	O	S		30		63.8		26.5			
	O	S		30		73.3		23.1			
Tlusty, 1969	MB	H	M	25	50–59	55.5		28.8			
	MB	H	M	25	60–69	65.0		25.4			
	MB	H	M	13	70–79	73.0		22.2			
	MB	H	M	2	80	85.5		16.8			
	MB	H	F	13	50–59	55.5		25.4			
	MB	H	F	14	60–69	63.5		19.2			
	MB	H	F	7	70–79	75.0		17.2			
Tuxworth et al., 1986	SB	G	M	160	35–39			34.0			
	SB	G	M	287	40–44			33.0			
	SB	G	M	275	45–49			31.0			
	SB	G	M	288	50–54			28.0			
	SB	G	M	350	55–60			26.0			
Tzankoff et al., 1972	MT	V	M	7	44–53	48.4		31.0			
	MT	V	M	8	54–66	58.0		29.0			
Vodak et al., 1980	MT	X	M	25	31–55	42.0	6.0	50.2	5.7	178.0	9.5
	MT	X	F	25	31–55	39.0	3.3	44.2	5.4	179.0	6.1
Vogel et al., 1986	MT	G	M	210	17–25	19.7	2.2	51.1	5.1		
	MT	G	M	199		24.7	5.7	46.9	7.1		
	MT	G	M	168		43.4	2.7	36.7	5.5		
	MT	G	M	335		43.9	3.2	40.6	6.2		
	MT	A	M	315		23.5	7.9	51.9	6.0		
	MT	G	F	212	17–25	19.7	1.9	37.5	3.7		
Von Dobeln et al., 1967	MB	G	M	21	30–39			42.3		178.0	9.0
	MB	G	M	21	40–49			36.0		171.0	10.0
	MB	G	M	22	50–59			33.9		163.0	9.0
	MB	G	M	20	60–70			29.2		155.0	9.0
Wahren et al., 1973	MB	X	M	7	52–59	54.4	2.6	48.4	4.8		
	B	V	M	7		52.9		40.3			
Wilmore et al., 1970	MB	V	M	17	17–24	21.5	1.9	46.5	6.5	190.5	9.3
	MB	V	M	15	25–34	29.7	2.9	43.6	5.0	188.2	12.1
	MB	V	M	16	35–44	40.5	3.1	40.1	7.0	178.9	10.0
	MB	V	M	7	45–59	52.9	4.4	40.3	5.0	178.6	7.9
Zauner, 1982	MT	A	M	20	35–42			44.3	9.3	197.2	11.6
	MT	A	M	18	43–57			42.3	7.2	195.7	11.2

Gozna et al. (1974) studied age-related changes in the mechanics of the aorta and pulmonary artery in 43 men. Only subjects with pulmonary and arterial pressures in the normal range were included. Twenty-two subjects took part in the pulmonary artery study and 21 subjects in the aorta study. The pressure strain elastic modulus and pulse wave velocity increased linearly with age, doubling from ages 20 to 60, in both the aorta and pulmonary artery.

Hossack et al. (1980, 1981) compared the cardiac output of 99 normal men, 104 normal women, 77 men with coronary heart disease, and 16 women with heart disease. All subjects were classified as sedentary on the basis of participating in vigorous physical activity less than once per week. Subjects performed a maximal treadmill work capacity test. In the normal subjects, maximum $\dot{V}O_2$ (ml/kg/min) declined .445 per year of age in men and .333 per year of age in women; estimated cardiac output declined by .170 and .071; maximal heart rate declined 1.032 and 0.597; and estimated stroke volume declined by .31 and .20. The majority of subjects with coronary heart disease fell below the mean normal value for maximum $\dot{V}O_2$, heart rate, cardiac output, and stroke volume, and a substantial proportion of these patients fell below the 95% confidence limits for normal subjects.

Lester et al. (1968) evaluated the effects of age and athletic training on maximal heart rate in 190 healthy men aged 15–75. Maximal heart rate was measured during a maximal graded treadmill exercise test. Age and maximum heart rate were significantly negatively correlated. Athletes had a slightly but significantly slower maximal heart rate than untrained men at any given age. Decreases in maximal heart rate with age were not significantly different between groups ($205.02 - .411 \times$ age for sedentary subjects and $198.19 - .411 \times$ age for athletes).

To assess the effects of age on left ventricular ejection fraction (LVEF), Port et al. (1980) performed radionuclide angiocardiography at rest and during upright bicycle exercise in 46 male and 31 female volunteers ages 20–95. At rest, LVEF, end diastolic volume, and regional wall motion were not affected by age. During exercise, however, LVEF was less than 0.6 in 45% of subjects over age 60 and in only 2% of younger subjects. The difference between exercise and rest LVEF increased with age. Wall motion abnormalities during exercise occurred with increasing frequency in subjects over age 50.

In 62 patients over 75 years of age, LVEF averaged 0.60 in subjects without clinically identifiable heart disease and 0.49 in the 30 patients with clinically identifiable heart disease. In the heart disease patients, reduction of LVEF was correlated with cardiac enlargement or congestive heart failure (Luchi et al., 1982).

Shaw et al. (1973) investigated the effects of age, activity, and blood pressure on systolic time intervals in 315 men aged 20–89. Elevated diastolic pressure was significantly correlated with prolongation of PEP, PEPI, and QS2I. Elevated systolic pressure was significantly correlated with prolongation of the LVET, LVETI, QS2, and QS2I. Neither coronary heart disease, functional Class I (67 subjects), nor estimated daily physical activity were significantly related to systolic time interval. PEP, PEPI, and QS2I were slightly but significantly prolonged from the third through the sixth decade and declined in the seventh and eighth decades.

Rodeheffer et al. (1984) sought to delineate between age- and disease-related declines in cardiac output. He reported that exercise cardiac output is maintained in subjects free from coronary disease. Subjects were participants in the Baltimore Longitudinal Study of Aging and moderately physically active. Sixty-one subjects (47 men and 14 women) ages 25–79 met the selection criteria for coronary health based on thallium scans and ECG's at rest and during bicycle ergometer exercise and maximal treadmill exercise. Average daily caloric expenditure declined with age. There were no significant changes with age in resting cardiac output or heart rate, end diastolic volume, end systolic volume, and ejection fraction. Resting blood pressure, however, increased with age. During exercise at the same loads, cardiac output was not affected by age, but heart rate was significantly decreased, and end diastolic volume, end systolic volume, and stroke volume were increased. At higher workloads, the effect of age was more pronounced.

Another study attempting to differentiate between true aging and undetected heart disease and deconditioning was performed by Kostis et al. (1982). Of 1,500 patients referred for cardiac catheterization, 101 subjects with normal heart function were identified. All subjects had normal physical examinations of the cardiovascular system, ECG, right- and left-heart catheterization, and coronary arteriography and had no symptoms during a progressive maximal exercise test on a treadmill. Of the subjects selected, 50 were women and 51 were men (mean age 48.8, range 16–68 years). Age was significantly negatively correlated with maximum heart rate on the exercise test ($r = 0.27$) and with maximal work capacity ($r = 0.41$). Heart rate at progressive loads increased more rapidly with age ($r = 0.30$) and declined less rapidly during recovery ($r = .28$). Maximum work capacity was significantly correlated with maximal heart rate ($r = .42$).

In a 20-year cross-sectional study of a group of 21 male runners ages 35–55, cardiac output was not significantly related to age (Wright et al., 1982).

Sato et al. (1981) studied cardiac control function in 39 healthy men ages 15–75. Cardiac control was evaluated by systems analysis of heart rate response to supine and upright rest and to two levels of stepping exercise. The results suggested that activity of both sympathetic and parasympathetic nervous systems were depressed at rest; contributions of sympathetic and vagal nervous systems during exercise are more enhanced and more depressed, respectively, in older subjects than in younger subjects.

Sebban et al. (1981) investigated the relationships between arterial compliance (estimated by pulse wave velocity, or PWV) and systolic blood pressure, heart rate, oxygen consumption, and time to run 400 m in 29 elderly women aged 53–100. Faster PWV's were related to higher blood pressure both at rest and during exercise. Time taken over 400 m rose as arterial compliance fell.

Wahren et al. (1974) compared the circulation of young (ages 25–30, $n = 7$) and old (ages 52–59, $n = 7$) men, evaluating blood flow and oxygen uptake of the leg and release of lactate at rest and during exercise. The authors concluded that blood flow to the leg in the older men rose in a curvilinear manner in response to exercise of increasing intensity and that leg circulation during exercise becomes relatively hypokinetic with age. Leg blood flow rose less during exercise in the older group, compensated for by a larger arterial-femoral venous oxygen difference.

DeVries and Adams (1972a) compared the exercise responses of young ($n = 12$, mean age 16.7) and old ($n = 12$, mean age 69.2) men matched for age-normalized physical fitness. Subjects worked at each of five workloads on a bicycle ergometer. Resting heart rate was significantly lower for the older men, and the rate of increase with exercise was not significantly different between the two age groups. At the two lowest workloads, the heart rates of the older men were significantly lower than heart rates of younger men. Blood pressure was significantly higher for the older subjects at all workloads, but there was no difference between groups in the regression of systolic blood pressure versus oxygen uptake. The regression of cardiac effort (systolic blood pressure × heart rate) versus oxygen uptake had a nonsignificantly greater slope in the older men. DeVries and Adams (1972b) also evaluated ventilatory mechanics in this population. The old men breathed less efficiently for the same workload or unit of oxygen consumed, having significantly greater tidal volume and significantly lower respiratory frequency throughout the range of workloads.

Durnin and Mikulicic (1956) compared responses of young (ages 20–30, $n = 12$) and old (ages 55–65, $n = 12$) men to submaximal arm and treadmill exercise. Oxygen consumption, heart rate, and blood pressure did not differ between groups for arm ergometer work. Older subjects, however, had higher $\dot{V}O_2$ (l/min) and pulse rates at submaximal loads on the treadmill than younger men.

Fitness in the general population

Robinson (1938) performed a pioneering study on the effects of aging on 93 normal males aged 6–91. Studies included basal, submaximal, maximal, and recovery heart rates; work capacity; lung volumes; and blood.

Fardy et al. (1976) recruited 594 healthy men ages 27–74 to participate in a study of cardiovascular risk. Eighty-eight men were excluded for health factors, low work capacity (less than 10 ml/kg/min), and problems interpreting test results. Subjects took a submaximal bicycle ergometer test, and ECG's were evaluated at rest and after exercise. Estimated maximum $\dot{V}O_2$ (ml/kg/min) dropped approximately 25% between the groups aged 25–34 ($n = 106$) and 55–74 ($n = 58$). Resting blood pressures rose by 12 mmHg systolic and 5 mmHg diastolic.

Drinkwater et al. (1975) studied 109 females ages 10–68 recruited from the general population. Subjects completed a maximal work capacity test and a brief questionnaire on activity level. Women aged 50–59 had lower maximal $\dot{V}O_2$ than all other groups, followed by women aged 40–49. Maximum heart rate was higher in subjects under 30 than subjects over 30, but no significant changes in mean maximum heart rate occurred among age groups over 30. A regression versus age for all subjects predicted a 0.338 decrement in $\dot{V}O_2$ (ml/kg/min) for each year of age.

Grimby et al. (1970) analyzed a representative sample of men aged 54 in Gothenburg, Sweden. Maximal oxygen uptake was 30.5 ml/kg/min and maximum heart rate 172.

In a study of 54 sedentary men and women aged 18–68, Julius et al. (1967) reported that both maximum $\dot{V}O_2$ and maximum cardiac output declined with age. The relationship between cardiac output and $\dot{V}O_2$, however, was unchanged with age.

As part of the Tecumseh study, 1,064 healthy men aged 10–69 performed a maximal or near maximal exercise test. Of the original sample of 2,024, 677 were excluded from exercise testing for medical reasons, including coronary heart disease (36%), hypertension (28%), and electrocardiographic abnormalities at rest (25%). Subjects walked on a progressive treadmill test to maximum effort for subjects 20–39, to a heart rate of 160 for subjects 40–59, and to a heart rate of 150 for subjects 60 and over. The frequency of ST segment depression increased progressively between the age groups 30–39 (2.8% of subjects) and 60–69 (17%); J-junction depression increased from 2% to 13%; and premature ventricular contractions from 2% to 30%. Maximal oxygen uptake on the test declined approximately 45% between the age groups 20–29 and 60–69 years, although part of this difference resulted from the early termination of tests in subjects age 40 and over (Chiang et al., 1970). $\dot{V}O_2$ (ml/kg/min) at submaximal loads increased from ages 20 to 59. Efficiency of subjects 60 and over was not compared with that of other groups because the older subjects performed the treadmill test at a different speed (Montoye, 1982).

Olree and Stevens evaluated 330 girls (ages 12–18) and 80 women (ages 19–60) on a submaximal bicycle ergometer test and a maximal treadmill test. Maximal oxygen uptake (ml/kg/min) did not change significantly from ages 14 to 29 ($n = 337$) but was significantly lower in the 30–60 age group ($n = 50$) than in younger subjects. Although the subjects were instructed to exercise until exhaustion, the respiratory exchange ratios at maximum work level average lower than 1.0 for all age groups, which indicates that not all subjects reached maximal effort.

Sheffield et al. (1978) studied maximum work capacity and heart rate in 95 asymptomatic women fairly evenly distributed over the ages of 19 to 69. Maximum heart rate was inversely related to age, decreasing .88 beats/minute per year of age based on regression analysis ($r = -0.76$, CV 5.5%). The decrease in treadmill exercise time with age was not statistically significant, and exercise duration was only weakly related to maximal heart rate ($r = 0.22$). Resting and maximum heart rates were not significantly correlated.

Siconolfi et al. (1985) evaluated maximum oxygen uptake and blood pressure in 184 men and 227 women aged 18–65. Subjects who had hypertension (systolic pressure greater than 160 or diastolic pressure greater than 95), suspected angina pectoris, or systolic murmur or were on beta blockers were excluded from testing ($n = 134$). Maximum oxygen uptake, estimated from a cycle ergometer test ($n = 411$), was negatively correlated with both systolic and diastolic pressure. When the effect of age was partialed out, however, the proportion of variance in blood pressure accounted for by fitness was less than 3%. Despite the absence of overt coronary disease in the older subjects, maximum $\dot{V}O_2$ (ml/kg/min) declined almost 50% from the age groups 18–29 to 60–65 in men and women. Systolic pressure increased 17 mmHg in men and 28 mmHg in women, and diastolic pressure increased 9 and 13 mmHg, respectively.

Tlusty (1969) evaluated work capacity in men ($n = 65$) and women ($n = 35$) aged 50–90. Subjects were screened by a clinical examination, after which only subjects classified as healthy were accepted. Work capacity was measured by a maximal progressive bicycle ergometer test. Based on regressions, $\dot{V}O_2$ (l/min) declined significantly in both men and women, by 0.06 and 0.038 per year of age, respectively. Part of the apparent decline can be attributed to the lower weights of older subjects. Maximum heart rate also declined significantly, by 1.182 beats/min per year of age in men and 1.271 in women.

Vogel et al. (1986) studied aerobic capacity in several groups of U.S. Army members. Aerobic capacity was greatest immediately following basic training and declined an average of 10%, or 0.5 ml/kg/min per year of age.

Von Dobeln et al. (1967) tested 84 building workers aged 30–70 on submaximal and maximal bicycle ergometry tests. Maximum $\dot{V}O_2$ declined approximately 31% from ages 30–39 to ages 60–70.

Activity and fitness

Andersen and Hermansen (1965) compared the aerobic work capacity of well-trained men ($n = 72$, ages 50–66) to office and industry workers ($n = 38$, mean age 55). The well-trained men were cross-country skiers participating in the Birkebeiner race. Maximal $\dot{V}O_2$ (ml/kg/min), measured on a progressive bicycle ergometer test, was 37% higher in skiers than in the control group and 9% higher than in nonathletic younger students measured by Hermansen and Andersen (1965). Maximum oxygen pulse was 33% higher in skiers than controls. Significance levels were not reported.

Astrand (1958) studied the work capacity of truckdrivers from Stockholm breweries who were placed into three age groups: 50–54 ($n = 46$), 55–59 ($n = 27$), and 60–64 ($n = 8$). The subjects' employment required heavy but intermittent work. Work capacity tests were performed on a bicycle ergometer at 600 kpm/min, 900 kpm/min, and attempted maximum. Maximal heart rates differed little by age group (161, 158, and 158 beats/minute). Maximal oxygen uptake declined from 33.4 in the age group 50–54 years to 31.0 in subjects aged 55–59 and 31.1 in subjects 60–64. No levels of significance were reported.

Later, Astrand (1960) studied the effects of age and sex on work capacity, responses to submaximal exercise, and electrocardiograms. In one experiment, 44 women ages 20–65 were recruited from an organized calisthenics program (age groups: 20–29, $n = 8$; 30–39, $n = 12$; 40–49, $n = 8$; 50–65, $n = 16$). Maximal $\dot{V}O_2$ was evaluated by a progressive bicycle ergometer test. On the average, blood lactate did not rise as much in the older subjects. Maximum $\dot{V}O_2$ (ml/kg/min) was 29% lower in the oldest age group than the youngest. The relationship of heart rate to oxygen uptake at submaximal loads was similar in all age groups. Maximum heart rate, however, fell from 187 in the youngest group to 170 in the oldest. In a second experiment, the same women and additional men aged 27–45 performed bicycle ergometer tests to determine whether age affected mechanical efficiency. At lower work loads, mechanical efficiency was slightly inhibited by age; women had lower oxygen uptakes at a given workload. In a third experiment, ECG was recorded at rest and during exercise. The two most common findings were arrhythmias and S–T junction changes. Only two subjects (ages 30–39) had suspected pathological ECG's at rest. Exercise ECG's were suspected to be or definitely pathological in two subjects aged 30–39, two aged 40–49, and seven aged 50–59. Forty-two of these women also participated in a treadmill test at a constant speed. Oxygen uptake (ml/kg/min) did not differ significantly between younger and older age groups on the submaximal load.

Japanese women ages 20–62 (102 sedentary and 46 active) were evaluated by Atomi and Miyashita (1974) for performance on a maximal bicycle ergometer test. The active subjects, who had participated in a regular physical conditioning program for 2 hours/day, 2–3 times a week for 4 years, had greater maximum $\dot{V}O_2$ than sedentary subjects in each age group. The decline in work capacity (ml/kg/min) per year of age did not differ significantly between the sedentary (.283) and active (.386) women. Maximum $\dot{V}O_2$ (ml/kg/min) was approximately 25% lower in those 50–62 years ($n = 18$) than in those 20–29 years ($n = 73$) for both sedentary and active women.

Bailey et al. (1974) evaluated 1,230 men and women ages 15–69 in a survey of Canadian fitness. Subjects with evidence of heart disease were excluded from the analysis (11% of the initial random sample). Work capacity was evaluated on a submaximal bicycle ergometer test. Maximum $\dot{V}O_2$ (ml/kg/min) fell approximately 38% between the third and seventh decades in both men and women. Subjects who exercised frequently, according to self-reports, (14% of men and 9% of women) were approximately 25% higher in aerobic capacity than subjects who exercised infrequently.

Beall et al. (1985a) compared the physical fitness of elderly men in traditionally sedentary and active castes in Nepal. Forty-three men 60 and over (ages 60–88, mean age 67) participated in the study. The 9 Sarkis, members of the laborer caste, had lower rest and submaximal heart rates and blood pressures than men in the other, less active castes. Beall et al. (1985b) also compared the fitness of farmers ages 50–79 living in mountainous (n = 47) and flat terrains (n = 39). Heart rate and blood pressure at rest, submaximal, and maximal loads did not differ between the two groups.

Bottiger (1973) studied the performance of approximately 10,000 Swedish men ages 11–70 participating in cross-country foot and ski races. Performance decreased approximately 5–10% for every 10 years of age.

De Backer et al. (1981) studied the relationship between work and leisure physical activity and physical fitness in 1,513 men aged 40–55. Physical activity was evaluated by interview and physical fitness by the workload on a bicycle ergometer producing a heart rate of 150. Workload at this submaximal heart rate did not differ significantly among the three 5-year age groups. Both leisure-time and job activity were significantly related to physical fitness, but together they accounted for only 2% of the variance.

Both cross-sectional and longitudinal rates of change in maximal $\dot{V}O_2$ were calculated by Dehn and Bruce (1972). For 86 healthy men ages 40–72, a cross-sectional regression analysis indicated maximum $\dot{V}O_2$ (ml/kg/min) = 49.93 − 0.278 × age. Active subjects had significantly higher maximal work capacity than inactive subjects. Longitudinally, maximum $\dot{V}O_2$ declined −0.94 ml/kg/min per year over 2.3 years in 40 men who repeated the tests. Active men in the longitudinal study (n = 24) lost 0.56 ml/kg/min in maximum $\dot{V}O_2$ per year; inactive men (n = 8) lost 1.62 ml/kg/min, almost three times the rate.

A comparison of urbanites (n = 400) and communal agricultural members (n = 302) indicated that a higher level of daily activity contributed to greater work capacity (Epstein et al., 1981). Kibbutz members in the age groups 22–30 years and 31–40 years had significantly greater maximum $\dot{V}O_2$ (ml/kg/min) than the less active city dwellers in the same age ranges.

Espenschade (1969) compared aspects of physical and psychological well-being of two groups of women: Subjects who had majored in physical education from 1913–53 (n = 125) and controls (n = 87). Subjects ranged in age from 35 to 79 years at the followup. Scores on a self-administered step test were significantly better for former physical education majors than controls. No trend with age in the self-reported step test score was detected.

The relation of habitual activity to functional capacity in old age was studied by Fischer et al. (1965). A progressive bicycle ergometer test was performed by 171 men aged 53–90. Sixty-nine of the men exercised regularly and were classified as physically active. Maximal work capacity (watts) declined significantly with age in both active and inactive groups. Active men had greater work capacity than inactive men in the same decade. Maximum $\dot{V}O_2$ (ml/kg/min) did not differ significantly from the sixth to eighth decade in active men but declined in untrained men between the seventh and eighth decades. Similarly, maximum heart rate declined with age in the sedentary men, but there was no significant change in the active men.

Froelicher et al. (1974) evaluated work capacity in U.S. Air Force members: 519 men ages 20–53 were referred as part of medical evaluation, and 191 men ages 25–44 were evaluated prior to assignment on special projects. Based on analysis of 5-year age groups, maximum $\dot{V}O_2$ (16%, ml/kg/min) and heart rate (20 beats) decreased significantly and percentage body fat and serum cholesterol increased over 30 years in the first group of subjects. In the subjects referred for special projects, presumed to be more active, maximum $\dot{V}O_2$ (ml/kg/min), percentage body fat, and serum cholesterol did not differ between the age groups 25–29 and 40–44 years. Maximum heart rate was identical in the more active groups aged 25–29 (n = 72) and 40–44 (n = 6). These more active subjects had higher maximal work capacities and heart rates and lower percentage body fat than the less active airmen of the same age.

The work capacity of 33 middle-aged and old athletes (ages 42–68) was investigated by Grimby and Saltin (1966). Subjects were active competitors in long-distance cross-country running (orienteering). Oxygen uptake was determined at two submaximal and one maximal load on a bicycle ergometer. There was no significant difference in oxygen uptake at submaximal workloads among age groups (42–49, n = 14; 50–59, n = 15; 61–68, n = 4). Mean values for pulmonary ventilation per liter O_2 increased with age, and heart rate tended to be higher in the older age groups at the same workloads. Maximal heart rate had large individual variations. Maximal oxygen uptake ranged from 57 ml/kg/min in the youngest age group to 53 and 43 ml/kg/min in the two older age groups. Resting and maximal heart rates were highly correlated (r = .81). ST depressions during exercise were noted in 24% of subjects, ectopic heart beats were fairly common (27%), and two subjects (both ages 50–59) had Wencheback heart-block during the first minutes of exercise.

Saltin and Grimby (1968) later compared the work capacities of former athletes with those of the still active athletes of the previous study. Subjects were 29 men ages 44–67 who had participated in orienteering, cross-country running, or skiing for at least 8 years but had been sedentary for at least 10 years prior to the study. Maximal oxygen uptake, measured on a maximal bicycle ergometer test, was 20% higher than in sedentary men but 25% lower than in the continuing athletes.

Heath et al. (1981) matched masters endurance athletes (n = 16, ages 50–72) to 16 young athletes (n = 16, ages 18–27) on the basis of training intensity and distance. A control group of 18 middle-aged untrained men, divided into two equal groups on the basis of percent body fat, was also studied. Work capacity was measured on a maximal treadmill test for the 30 runners and on a maximal bicycle ergometer for the two cyclists. Maximum work capacity (ml/kg/min) was 15% lower in the masters than young athletes, and maximum heart rate was 14% lower. The oxygen pulse at maximum work was not significantly different between young and older athletes. Masters athletes had a maximum $\dot{V}O_2$ divided by lean body mass 70% higher than that of overweight untrained men and 54% higher than that of lean untrained men. The difference between masters and young athletes represented approximately 5%/decade decline in maximum $\dot{V}O_2$, which the authors hypothesize may be the lower limit of age-related changes.

Hermansen and Andersen (1965) evaluated work capacity in young adult Norwegian men and women. The 14 top male athletes (mean age 28) had a work capacity 25% greater than that of sedentary men (mean age 23), and the five top women athletes (mean age 25) averaged 23% greater in work capacity than sedentary women (mean age 23).

Pollock (1974) reviewed information on the physiological characteristics of athletes age 40 and over. American track athletes were lower in body weight and percentage fat than the average population. Older athletes had significantly lower work capacities than young athletes, and maximum work capacity declined with age within the group of older athletes, particularly between the two oldest age groups. Part of this decline might have resulted from the decline in training and the low numbers of subjects. Not only was maximum $\dot{V}O_2$ (ml/kg/min) of the older athletes greater than in age-matched sedentary men, but the values for these athletes aged 40–69 exceeded values for sedentary men aged 18–28.

Maximal short-term exercise declined by approximately 6% per decade of age in a study by Makrides et al. (1985). Fifty men and 50 women aged 15–71 exercised maximally for 30 seconds on an isokinetic ergometer. Maximum $\dot{V}O_2$ (ml/kg/min), evaluated on a progressive bicycle ergometer test, was correlated with short-term exercise (r = 0.89), vital capacity, and leisure activity. Leisure activity was not, however, significantly related to maximum short-term exercise.

McDonough et al. (1970) measured work capacity in 86 middle-aged men free of heart disease ages 40–69. Maximal work capacity declined fairly linearly with 5-year age groups for ages 40–44 to 55–59 and appeared to decline more steeply after the age of 60. Regression analysis yielded coefficients of 0.41 and 0.39 ml/kg/min decline in $\dot{V}O_2$ per year of age in sedentary and active men, respectively. A group of men participating in moderate or heavy activity were significantly higher than the sedentary and light activity group in maximum $\dot{V}O_2$ and oxygen pulse in all age groups.

Profant et al. (1972) tested 144 healthy women ages 29–70 for maximal oxygen uptake on a multistage treadmill test. The 42 women who were active had significantly higher work capacities than the sedentary women. Cross-sectionally, maximum $\dot{V}O_2$ (ml/kg/min) declined 0.8% per year. There was little change in maximal heart rate from the fourth to fifth decades (184 to 181, $p < 0.05$), but a greater change from the sixth to seventh decades (176 to 159, $p < 0.01$).

Tuxworth et al. (1986) evaluated fitness and health in 1,394 male factory workers aged 35–60. Maximum $\dot{V}O_2$, estimated by a submaximal bicycle ergometer test, was 23% lower in the oldest age group (55–60, n = 350) than in the youngest (n = 160, ages 35–39). Subjects who were more active (rode bicycles, participated regularly in physical activity, would run up stairs or for a bus) had a higher fitness level than the less active.

Vodak et al. (1980) studied 25 male (mean age 42) and female (mean age 39) tennis players ages of 31–55. Maximal oxygen uptake was evaluated on a graded treadmill test to exhaustion. Compared with other studies, mean maximum $\dot{V}O_2$ (ml/kg/min) was higher in both men (50.2) and women (44.2) than in sedentary subjects but lower than in dedicated runners.

Wessel et al. (1966) measured energy metabolism in 98 healthy women aged 20–69 during and after a 10-minute treadmill walk at 1.75 mph, 4% grade. Test-retest reliability, evaluated for 19 subjects aged 35–49, was .85 for $\dot{V}O_2$ in l/min and .65 for $\dot{V}O_2$ in ml/kg/min. Exercise $\dot{V}O_2$ (l/min) and exercise heart rate, but not $\dot{V}O_2$ in ml/kg/min, were significantly correlated with age.

In a later study (Wessel et al., 1968; Wessel and Van Huss, 1969), 47 women aged 20–69 (9 or 10 subjects per decade) were evaluated on a 10-minute walk at 3 mph, 0% grade. Subjects were accepted into the study only if they had no diseases or conditions that would affect their ability to take the treadmill test. Physical activity level was estimated by a questionnaire or interview. Age was significantly positively associated with $\dot{V}O_2$ (l/min) and oxygen pulse at this submaximal load. Physical activity was negatively correlated with $\dot{V}O_2$ (l/min), oxygen pulse, and heart rate. Neither age nor physical activity level was significantly correlated with O_2 in ml/kg/min. Although groups based on physical activity level differed in physiological response, this was confounded by the lower age and body weight of the more active subjects.

Longitudinal studies

General population

In a longitudinal study, 36 women (initial ages 18–68) were retested for work capacity after an average of 6.1 years (Plowman et al., 1979). Aerobic work capacity was evaluated on a modified Balke treadmill protocol. The time between tests ranged from 3.5 to 9.0 years. Subjects in their twenties (*n* = 5, mean age 23.0) did not change in aerobic work capacity. In all other age groups, work capacity declined at a rate similar to that predicted by the original cross-sectional study. Maximum heart rate declined significantly in all groups. For subjects under 50 (*n* = 25) and 60 and over (*n* = 3), maximum heart rate declined only 2–4 beats between tests. In the age group 50–59 years, however, it declined by an average of 7.6 beats/minute between the two tests.

Ericsson and Irnell (1969) evaluated longitudinal changes in work capacity in 42 men and 42 women aged 57–71. In 5 years, maximal work capacity declined an average of 12% in men and 19% in women. Within 5-year age groups, men aged 62–66 years and 67–71 years and women aged 67–71 years did not decline significantly in work capacity, although the six women aged 67–71 had the greatest percentage decline of all age and sex groups (27%, *n* = 6).

Former college students were retested after 22 and 31 years by Robinson et al. (1975). Work capacity was evaluated by a maximal graded treadmill test. During the first 22 years, maximal heart rate declined 15 beats/minute and maximal $\dot{V}O_2$ declined 25%. Eight men increased their activity level and improved by an average of 11% in maximum $\dot{V}O_2$ during the following 9 years; the rest of the 37 subjects continued to decline.

Active subjects

Astrand et al. (1973) tested 35 women and 31 men on a bicycle ergometer in 1949 (ages 20–33) and again in 1970. The subjects, physical education teachers, were well trained in 1949, and most were still relatively active in 1970. Maximal oxygen uptake decreased approximately 20% in both women and men. Maximum heart rate declined by 15 and 12 beats/minute, respectively, but there were wide individual differences. Although submaximal heart rates increased, because of the variation in the magnitude of change of maximum heart rate, the amount of reduction in maximal oxygen uptake could not be predicted from submaximal tests.

In a further followup, Asmussen et al. (1975) reevaluated 19 of these men and 6 women from this study after 40 years. Maximum $\dot{V}O_2$ (ml/kg/min), estimated from a submaximal bicycle ergometer test, declined significantly from the initial test, by approximately 50% in women and 33% in men. Resting systolic blood pressure increased significantly, from 117 to 138 in men and 121 to 135 in women. Heart rate at a submaximal load did not change significantly.

Dill et al. (1967) assessed the physiological status of 16 former championship runners after 20 years. Subjects were tested on a 94 m/min treadmill walk and a treadmill run. Changes in maximum $\dot{V}O_2$ (ml/min per kg LBM per year) ranged from 0.24 to 1.59.

Kasch (1976) followed two groups of formerly sedentary men for 6 years (*n* = 17, mean age 48) and 7 years (*n* = 10, mean age 47) and one group of habitually active men for 10 years (*n* = 16, mean age 45). The training method for the first two groups consisted of 20 minutes of calisthenics followed by 30–35 minutes of interval and continuous running. Group III ran an average of 59 minutes. Six men used 35–60 minutes of swimming as their training method. Subjects averaged two sessions per week in groups I and II and three times per week for group III, at intensities of 65–92% in the first two groups and 60–93% in the third. Maximum work capacity was evaluated on a treadmill, bicycle ergometer, or step test. On the post tests, group I had improved by approximately 23% (29.4 to 38.5), group II by approximately 10% (32.9 to 37.3), and group III, nonsignificantly, by 1% (43.9–44.4). Resting heart rates declined 15 and 17 beats in the first two groups and 2 beats in the third. Maximum heart rate decreased 6 and 7 beats in groups II and III and increased 2 beats in group I. Resting blood pressure declined substantially in group I and remained essentially the same in the other two groups. After 1 year, maximal oxygen uptake reached a plateau in groups I and II.

Kasch et al. (1985) followed 15 men in group III of the previous study for 15 years and 13 of these men for 18 years. The men exercised approximately 3.3 days/week for 45 minutes at 77% of heart-rate reserve. Over the 18 years, $\dot{V}O_2$max changed very little, from 44.6 to 43.1 ml/kg/min. There was a slight increase in the first 10 years (to 45.2) and a decrease in the next 5 years (to 40.2). Weight decreased from 76.1 kg to 71.2 kg over the 18 years.

Former champion runners were retested 25–43 years after their competitive track careers. Their physiological response to exercise was evaluated on a standard treadmill walk and exhausting work on the treadmill. Maximum $\dot{V}O_2$ declined from 71.4 ml/kg/min in youth to 41.8 at the mean age of 56.6. Although former athletes had greater work capacities than controls at similar ages, the decline in maximum $\dot{V}O_2$ from the peak training years was approximately twice as steep. Maximum heart rate declined 6 beats (186 to 180) in athletes and 13 beats (199 to 186) in nonathletes (Robinson et al., 1976).

Exercise intervention

Age influence on training effect

Kilbom (1971) studied the effects of training on three age groups of women: ages 19–31 (*n* = 12), ages 34–48 (*n* = 8), and ages 51–64 (*n* = 13). Prior to training, maximal oxygen uptake was considerably lower in the oldest subjects. Subjects trained for 7 weeks at approximately 70% of their maximum capacity on a bicycle ergometer, 18 minutes per session. After training, submaximal heart rates had decreased, particularly in the two younger groups. Maximal heart rates were not significantly affected by training. Maximal oxygen uptake (ml/kg/min), evaluated by a progressive bicycle ergometer test, increased with training by 11% in the youngest group, 13% in the group aged 34–48, and 9% in the group aged 51–64. Maximum cardiac output increased by 11%. Resting and submaximal cardiac outputs were lower before and after training for the oldest subjects. Submaximal stroke volume improved significantly with training in the two younger groups, but not the oldest group.

In another study analyzing the age influence on effects of training, 12 subjects in each of two age groups (mean ages 36.8 and 52.9) were matched for physical fitness before undertaking a 4-month physical fitness program. Training took place 3 days each week for 90 minutes and consisted of a variety of activities, including calisthenics, jogging, and sports activities. Both groups improved significantly on time run on a progressive treadmill exercise test, but improvements in maximal $\dot{V}O_2$ (ml/kg/min) were not significant. Resting blood pressure and submaximal heart rates decreased significantly in both groups. There were no significant changes with training in total lipids, cholesterol, triglycerides, or glucose. Increases in percent grade and time run were similar in younger and older age groups, although these factors tended to be lower in the older group, as did the change in maximum $\dot{V}O_2$. The authors noted that subjects initially had a high work capacity, which may account for the lack of significant improvement.

Naughton and Nagle (1965) trained 18 men, mean age 41, for 30 weeks, 3 times/week, 30 minutes/session. Maximum $\dot{V}O_2$ increased an average of 18%. The response to exercise was not significantly affected by age.

After 8 weeks of training, 27 previously sedentary men age 56–70 significantly increased maximal oxygen uptake, by 11% (Suominen et al., 1977a). Subjects were excluded if they had participated in regular physical activity during the previous 20 years. The subjects trained 3–5 times per week for 45–60 minutes. Intensity of the training was gradually increased, and by the third week pulse rates commonly reached 130–140 beats/minute. Increases in work capacity were similar in the three 5-year age groups.

Roskamm (1967) tested 18 men (6 per age group: 16–18, 20–30, and 50–60), who trained on the bicycle ergometer for 4 weeks. Each group improved significantly in maximum watt/pulse, but the oldest age group had a significantly smaller training effect (13%) than younger subjects (20%).

In another study reporting reduced training effects in older subjects, Wilmore et al. (1970) trained 55 men ages 17–59 for 10 weeks, 3 times/week, 12 or 24 minutes/session. The greatest improvement in maximum $\dot{V}O_2$ (ml/kg/min) occurred in the age group 35–44 (10%, n = 16), followed by subjects 25–34 (n = 15, 6.5%), subjects 17–24 (5%), and finally the subjects aged 45–59 (n = 7, 4%). Initially, subjects in the oldest age group were similar in maximum $\dot{V}O_2$ to the subjects aged 35–44, and a higher than average fitness in this group may account for part of the decrement in training effect. Resting heart rate and blood pressure declined in all age groups.

Effects of exercise on fitness in the elderly

An early study of training of the elderly (Benestad, 1965) reported that work capacity was unchanged by training in elderly subjects (n = 13, mean age 75.5). These men trained three times a week for 6 weeks on a treadmill inclined at 10% for 5 minutes at each of three loads (warmup, 70 m/min, and >80% maximum capacity). Work capacity was evaluated on a maximal bicycle ergometer test. As a result of the training program, submaximal heart rate declined significantly, but maximal oxygen intake was identical to the pretraining value. The results may reflect the short length of the program and the low intensity and duration of training sessions.

DeVries (1970) trained 68 men (mean age 70) for 6 weeks. Subjects trained 45–60 minutes per session, including 15–20 minutes of run-walk. Fitness was evaluated on a test terminated at a heart rate of 145. Heart rates were set under 145 beats/minute during the aerobic portion of the program. Oxygen pulse (29%) and physical work capacity (9%) at heart rate 145 improved in exercise subjects and was not significantly different in controls (n = 32) after 6 weeks. Changes in maximal $\dot{V}O_2$ (ml/kg/min) did not differ significantly between groups. Resting systolic and diastolic blood pressures were reduced in the exercise group and unchanged in the control group.

Twenty-three women from a retirement community (ages 52–79) participated in a study of the effects of exercise: 17 exercise subjects and 6 controls (Adams and deVries, 1973). A submaximal graded bicycle ergometer test was used to estimate 85% of maximum $\dot{V}O_2$. Exercise subjects trained for 3 months, 40–60 minutes per session. Of each session, 15 to 20 minutes consisted of walk-jog programs at 60% or greater of predicted heart rate reserve. As a result of the program, experimental subjects increased significantly in physical work capacity (37%, watt-min), oxygen uptake at 85% max (21%, ml/kg/min), and oxygen pulse at 85% max (19%). The control subjects did not change significantly on any of these variables. Exercise subjects differed significantly from control subjects in changes in resting heart rate, submaximal heart rates, physical work capacity (watt-min), and weight. Changes in oxygen uptake (ml/kg/min) and oxygen pulse did not differ significantly between groups.

Barry et al. (1966) studied five men and three women (ages 55–78, mean 70) who undertook an exercise program and two men and three women (ages 58–83, mean age 72) who served as controls. Exercise subjects trained 3 days/week in 40-minute sessions consisting of 2–3 minute work periods with a 30-second rest interval. Work capacity was evaluated at the maximal load tolerable for 6 minutes on a bicycle ergometer.

In the exercise subjects, submaximal pulse rate, systolic blood pressure, and blood lactate decreased significantly. Maximal oxygen uptake, pulmonary ventilation, heart rate, oxygen pulse, systolic pressure, and blood lactate increased significantly in the exercise group, whereas maximal work capacity and heart rate declined in the control group. In eight cases, initial work capacity evaluations were terminated because of abnormalities in the ECG and muscular insufficiency during the initial test; this accounts for some of the 38% increase in work capacity. Maximum heart rate and lactate values indicate that subjects were closer to exhaustion on the posttest.

Buccola and Stone (1975) compared the effects of walk-jogging ($n = 16$) and cycling ($n = 20$) on the work capacity of men aged 60–79. Subjects were pre- and posttested on a submaximal bicycle ergometer test. The training program, held three times a week for 14 weeks, consisted of 10 minutes of warmup and 10–40 minutes of cycling or jogging, followed by a 5-minute cooldown each session. Both groups decreased significantly in resting systolic and diastolic blood pressure and weight, and increased significantly in estimated maximum $\dot{V}O_2$ (ml/kg/min), by approximately 12%.

Cunningham et al. (1987) studied the effects of 1-year exercise training on fitness and blood lipids. Two hundred men at retirement (ages 55–65) were randomly assigned to exercise and control groups. Work capacity was evaluated on a maximal graded treadmill test. The exercise subjects trained three times per week at 60–70% of their maximum work capacities. Each session consisted of 10–15 minutes of warmup, 30 minutes of walking or jogging, and a 10-minute cooldown. The subjects attended an average of 2.5 sessions/week. Maximum $\dot{V}O_2$ increased in the exercise group by 10.9%, which was significantly better than the control group. Maximum heart rate did not change significantly. The exercise and control groups did not differ significantly in cholesterol or HDL levels at the start or finish of the program.

Kiessling et al. (1974) compared the work capacities of 10 men ages 46–62 after 10–13 weeks endurance training and 9 chronic runners ages 43–66. Work capacity (l/min) increased 8% with training in the previously sedentary men but was still significantly lower (by 23%) than in the runners.

Twenty-four elderly men and women (mean age 72) participated in three exercise sessions weekly for 12 weeks (Schocken et al., 1983). Subjects were screened for evidence of cardiovascular disease. The aerobic portion of the program consisted of 25–30 minutes of exercise on a bicycle ergometer at 70–85% of the maximal heart rate observed on a maximal bicycle ergometer test. Maximum workload increased from 690 kpm/min on the pretest to 758 kpm/min on the posttest (approximately 10%), and the rate-pressure product at submaximal workload decreased. The cardiac index response to exercise improved, and end diastolic volume index produced by exercise increased. The decrease in left ventricular ejection fraction and increase in left ventricular end systolic volume index that were induced by exercise, however, were unaltered by training. The authors concluded that deconditioning was not a significant contributor to left ventricular contractile performance decline in the elderly.

Stamford (1972a) compared the work capacity and trainability of chronically institutionalized (CI) men ($n = 7$, mean age 68.2) and recently hospitalized (RH) men ($n = 7$, mean age 66.7). A control group of five subjects (mean age 70) was also formed. Exercise subjects performed three loads (150, 300, and 450 kpm) on a bicycle ergometer stress test initially and every 6 weeks throughout the 18-week program. Training sessions took place 5 days per week and progressed from 15 minutes at 50% heart rate reserve for the first 6 weeks, to 30 minutes the middle 6 weeks and 15 minutes at 60% heart rate reserve the final 6 weeks. During the first 6 weeks, submaximal exercise heart rate, O_2 pulse, and blood pressure declined in the CI group. No significant changes occurred in the CI group between weeks 6 and 12. The RH group improved significantly in the final 6 weeks. During the final 6 weeks, both groups decreased significantly in exercise heart rate, exercise and resting systolic blood pressure, and oxygen pulse.

In a similar study, Stamford (1972b) randomly assigned 17 geriatric mental patients to exercise ($n = 9$, mean age 71.5) and control ($n = 8$, mean age 65.2) groups. Work capacity was evaluated by a treadmill test to 80% of predicted maximal heart rate. Exercise subjects trained on a treadmill 5 days per week at 70% maximum heart rate for 12 weeks. The length of exercise was gradually increased from 6 to 20 minutes. Submaximal heart rates and blood pressures were significantly lower at the end of the program in exercise subjects but did not change in control subjects.

Fourteen men and 12 women (age 69) selected from an age cohort participated in an 8-week exercise program 5 days per week for 1 hour. Maximal oxygen uptake (ml/kg/min), estimated by a submaximal bicycle protocol, improved by 11% in men and 12% in women (Suominen et al., 1977b).

Thomas et al. (1985) recruited 224 recently retired men and assigned them randomly to control (mean age 62.6) and exercise (mean age 62.9) groups. The exercise program took place 3 days/week for 1 year. Subjects trained at 60% or more of maximum heart rate reserve (based on the pretest). Sessions included 10–15 minutes of warmup, approximately 30 minutes of walking or jogging, and 10 minutes of cooldown. The 88 subjects who completed the training program increased significantly in maximum $\dot{V}O_2$, measured on a maximal treadmill test (by 12%). There was no significant change in maximum $\dot{V}O_2$ in the 100 control subjects. In the exercise group, the amount of improvement was negatively correlated with the initial maximum $\dot{V}O_2$ and pretraining activity level and positively correlated with the speed of walking/running during training.

Fifteen men aged 44–66 participated in a 25-week training program. Subjects averaged 2.3 sessions per week of 55 minutes each and performed a variety of exercise activities. Their physiological response to exercise was evaluated by a standard 10-minute walk and exhausting work on the treadmill. Heart rate on the 10-minute walk decreased significantly with training, and maximum work capacity increased significantly (21.6%). Maximal heart rate did not change significantly with training. Serum cholesterol declined in both groups (Tzankoff et al., 1972).

The combined effects of dietary and exercise intervention were studied at the Pritikin Longevity Center (Weber et al., 1983). Seventy subjects (mean age 78.7) attended a 26-day program. A progessive treadmill test to 90% of age-predicted maximum heart rate or symptoms was performed by each subject. The training program consisted of walking 15–40 minutes twice daily and 5 days/week of 1-hour exercise classes. Training heart rates were set at 10–15 beats below the maximum attained on the treadmill test. Serum cholesterol and triglycerides decreased significantly, and treadmill performance increased significantly, from 3.7 to 5.5 METS. Peak systolic blood pressure was significantly reduced. The 46 subjects classified as hypertensive had nonsignificant reductions in resting blood pressure.

Parameters of training

Atomi and Miyashita (1980) studied the effects of training intensity on maximal aerobic power in sedentary women 23–40 years of age. Subjects trained on a bicycle ergometer at progressive intensities of 60%, 75%, and 90% of maximum for 13, 18, and 13 consecutive weeks, with a frequency of 2–4 days a week. Maximum $\dot{V}O_2$ (ml/kg/min) increased significantly during 60% and 90% training but not during 75% training.

Another study on the effects of training intensity was performed by Badenhop et al. (1983). Thirty-two subjects (mean age 67.8) were assigned randomly to high-intensity (60–75% heart rate reserve, $n = 14$), low-intensity (30–45% heart rate reserve, $n = 14$) and control groups ($n = 4$). Subjects with evidence of organic heart disease detected by history, physical exam, resting electrocardiogram, 24-hour ambulatory ECG, or ventricular tachycardia or ischemic ST-segment changes on the initial exercise test were excluded from the study. Work capacity was evaluated by a progressive bicycle ergometer test to maximum effort. The exercise groups trained 25 minutes/day, 3 days/week for 9 weeks. Both groups increased significantly in maximum $\dot{V}O_2$ (ml/kg/min), by 16% and 14.8%. Maximal ventilation increased, and submaximal ventilation, $\dot{V}O_2$ (l/min), and heart rate decreased. The two exercise groups did not differ significantly in improvement.

The exercise intensity threshold for older men was studied by deVries (1971). Fifty-two men aged 60–79 performed a bicycle ergometer test to a heart rate of 120 before and after a 6-week exercise program. Training consisted of a structured walk-jog program during which subjects recorded their heart rates. Intensity was defined as percentage of heart rate reserve. The correlation between intensity and improvement in work capacity was poor and was significant only for the age group 60–69 years. Because improvement was negatively correlated with initial work capacity, the data were normalized for fitness, and the correlation between intensity-fitness and improvement was raised to 0.229 (ages 70–79) and .383 (ages 60–69). The authors concluded that the best estimate of training intensity was Astrand score/100 for the threshold level; i.e., maximum $\dot{V}O_2$ (ml/kg/min) divided by 100 gives the percent heart rate reserve.

Niinimaa and Shephard (1978a, 1986b) recruited 14 men and 20 women aged 60–76 (mean age 65.4) for an 11-week exercise program. Fifteen of the original 34 recruits dropped out before the completion of training. Training sessions of approximately 1 hour were held four times/week. Maximum oxygen uptake, evaluated on a maximal progressive treadmill test, increased 10% in the 9 subjects who exercised at heart rates of 145 to 155 beats/minute ($p < 0.07$, $n = 9$) but did not change significantly in the 10 subjects who exercised at heart rates of 125–145 beats/minute.

Seals et al. (1984) studied the effects of training on 11 men and women (mean age 63). Maximal $\dot{V}O_2$ was evaluated on a progressive maximal treadmill test before the start of training, after 6 months of low-intensity (LI) training, and after an additional 6 months of higher intensity (HI) training. LI training consisted of 20–30 minutes of walking (heart rate <120) at least 3 times/week, plus an increase in overall daily physical activity. HI training consisted of 10–15 minutes of warmup and 30–45 minutes of aerobic exercise at 75–85% heart rate reserve. Maximum $\dot{V}O_2$ (ml/kg/min) increased significantly, from 25.4 before training to 28.2 after LI training and 32.9 after HI training, for a total improvement of 30%. Maximal cardiac output did not change significantly; thus the authors hypothesized that the increase in work capacity resulted from the greater arteriovenous O_2 difference ($p < 0.01$) following training. At the same absolute work rates, stroke volume was higher; heart rate, blood pressure, and systemic vascular resistance were lower; and cardiac output and arteriovenous O_2 difference were unchanged. At the same relative work rates (% of maximum) arteriovenous O_2 difference was higher, and blood pressure and systemic vascular resistance were lower; heart rate, cardiac output, and stroke volume did not change significantly.

Sidney and Shephard (1978) recruited 14 men and 28 women aged 60–83 for participation in an exercise program. Four sessions of supervised exercise were conducted each week, and subjects were assigned to four categories based on attendance and exercise heart rates—low and high frequency (LF, HF), low and high intensity (LI, HI). On the initial test, maximum $\dot{V}O_2$ (ml/kg/min) was estimated from a progressive submaximal bicycle ergometer test, with the final load at 75–85% of the predicted maximal heart rate. At 7, 14, 21, and 52 weeks, maximum $\dot{V}O_2$ was evaluated on a progressive maximal treadmill test. After the first 7 weeks, the maximum $\dot{V}O_2$ was lower in the LF LI group than the other three groups. Maximum $\dot{V}O_2$ increased significantly in all other groups, with the HF and HI groups significantly more improved than the respective LF and LI groups. Maximum $\dot{V}O_2$ increased nonsignificantly from the 7th to the 14th week. The two HI groups did not change significantly in the following 7 weeks, but the HF LI group improved. Subjects who trained for a year ($n = 22$) gained a total of 24% in maximum $\dot{V}O_2$, with most of the improvement occurring during the first 7 weeks.

Testing

Bruce (1977) reviewed means of exercise testing, with the aim of defining electrocardiographic abnormalities, functional aerobic capacity, and the cardiovascular mechanisms of any impairment. He recommends a multistage treadmill test preceded by a medical examination and resting electrocardiogram. In 20,000 tests, he had observed less than 0.1% morbidity and no mortality. A 1971 survey of Rochmis and Blackburn, cited by Bruce, reported morbidity and mortality of 2.4 and 1, respectively, per 10,000 tests. In some cases, however, subjects with coronary heart disease, prior myocardial infarction, or angina pectoris had postexercise myocardial infarctions and cardiac arrest.

Cumming and Borysyk (1972) studied the criteria for maximal exercise in 65 men aged 40–65. Subjects performed two submaximal and one supramaximal load on a bicycle ergometer. Mean maximal $\dot{V}O_2$ (ml/kg/min) declined 27% from the youngest to oldest 5-year age groups, and maximum heart rate declined 15 beats/minute. The frequency of subjects failing to meet criteria for maximum was 22% for lactate, 46% for respiratory exchange ratio, 57% for $\dot{V}O_2$ plateau, and 26% for heart rate. Lactate was not significantly correlated with respiratory exchange ratio or percent of predicted $\dot{V}O_2$. Over 90% of subjects met two or more of the criteria for maximum.

Fairshter et al. (1983) compared 15-second and 1-minute incremental exercise protocols on the bicycle or treadmill. Maximum oxygen uptake was higher on the treadmill than on the bicycle, but there was no difference between 15-second and 1-minute tests using the same protocol.

Kamon and Pandolf (1972) compared the maximal aerobic power achieved on laddermill, treadmill, and cycling tests. Maximum $\dot{V}O_2$ was highest on the treadmill, followed by climbing and then cycling in males, and climbing, treadmill, and cycling for the females. Subjects who participated in competitive sports ($n = 9$) had significantly greater work capacities than subjects ($n = 11$) who exercised recreationally 1–2 times a week.

Katch et al. (1982) studied the reliability of maximal treadmill tests. Four trained females and one trained male repeated 8–20 tests over 2–4 weeks. Subject variability plus equipment variability amounted to approximately 5.6%, of which 90% was caused by biological variability.

Michael and Horvath (1965) tested 30 women ages 17–22 on a progressive maximal bicycle ergometer test. Maximal work capacity could not be predicted from submaximal measurements for any individual subject. Twenty-one subjects given repeat tests did not differ significantly at submaximal or maximal levels of work.

Barry et al. (1969) investigated the validity and reliability of multistage treadmill testing for 19 elderly men (ages 61–80) and 14 women (ages 60–77). Ten subjects repeated the test on two occasions 1 week apart. Each subject performed a multistage test (150 kpm/min increase every 2 minutes) and a single-stage test (at the highest load of the multistage test) on a bicycle ergometer. Test-retest reliability on the multistage test was 0.82 for work capacity and 0.97 for maximal heart rate. Maximum heart rate on the single-stage test had a test-retest reliability of 0.77.

Reproducibility of maximum $\dot{V}O_2$ and ventilatory and lactate thresholds in elderly women was studied by Foster et al. (1986) for both maximal and submaximal measures. Eight women aged 73–86 (mean age 80.6) repeated three maximal graded treadmill tests. Maximum $\dot{V}O_2$ varied little between tests. Intrasubject correlation coefficients were 0.76 for maximum $\dot{V}O_2$ (ml/kg/min), 0.86 for lactate concentration, and 0.71 for heart rate. The authors concluded that ventilatory or lactate thresholds were not sufficiently reproducible to be applicable in submaximal measures of fitness for elderly women.

Notelovitz et al. (1986) compared predicted maximal oxygen uptake based on a submaximal bicycle ergometer test and maximal uptake on a progressive treadmill test. The correlation between tests for 29 women was 0.789. Menopausal status had no effect on cardiorespiratory fitness of age-matched subjects. For 163 subjects who took a maximal treadmill test, maximum $\dot{V}O_2$ was approximately 16% lower in the group aged 65–75 ($n = 20$) than in the group aged 35–44. A similar change with age was observed in the estimated maximal $\dot{V}O_2$ of 111 women performing the bicycle ergometer test.

Sidney and Shephard (1977a) studied the responses of 26 men and women aged 60–83 to work on a maximal treadmill test. The $\dot{V}O_2$ plateau was reached by 69% of men and 66% of women. Blood lactate values were significantly different for those making "good" and "fair" maximal efforts. Maximal heart rates were different for "good" and "fair" performance by men. Submaximal bicycle ergometer tests showed significant habituation in women but not men. Heart rates at a given oxygen consumption were substantially higher on the bicycle than on the treadmill. Astrand predictions had a substantial systematic error relative to direct $\dot{V}O_2$max. Cooper 12-minute walk-run tests showed a correlation with directly measured $\dot{V}O_2$max of .83 in men and .51 in women.

Recommendations for work capacity testing of older adults

Aerobic work capacity is a major factor in an older adult's continuing independence. We recommend that, whenever possible, a maximal (symptom-limited) work capacity test be performed for older adults. Although there is a good correlation of maximal heart rate with age, there is a large standard deviation, and with submaximal tests work capacity can be severely underestimated or overestimated. Estimation of maximal capacity is often complicated if the subject is on blood pressure medication. Maximal individual evaluation allows for the best determination of cardiovascular limitations during exercise.

Many older adults have orthopedic, coordination, aerobic, balance, or psychological factors that prevent them from participating fully in a work capacity test conducted on a treadmill using the Bruce or Balke protocol. Some are unable to walk comfortably at 3 mph, and some have a work capacity of less than five METS (3 mph, 5% grade). An orientation period, without gas sampling (mouthpiece and noseclips) so that the subject can communicate with evaluators, is helpful in determining the appropriate protocol and treadmill speed for the individual. If there is online interactive computer capability, a branching protocol should be considered based on the heart rate at 2-minute measurement intervals. If this is not possible, a limited number of protocols should be available to provide an 8–12 minute test based on the subject's heart rate during the orientation period. During orientation, the treadmill should start at 0 mph and gradually progress to 2, 2.5, and 3 mph as long as the subject is comfortable. Based on the heart rate, stability, and reaction of the subject, a protocol can be chosen. (Example protocols based on heart rates are shown in table 2.) Another alternative is to use a protocol branching at 75% of predicted heart rate reserve. For example, on a 2-mph test, the grade would be increased by 3.5% (1 MET) until the subject exceeded 75% of heart rate reserve. Thereafter, the estimated MET level would be increased by ½ MET (table 3). Tests given on this basis tend to be longer, however, for the more fit subjects.

The bicycle ergometer is an option for subjects who cannot ambulate sufficiently for a treadmill test. Because older subjects, particularly, have difficulty maintaining a constant cadence, the electronic ergometer should be used. As in the treadmill test, work increases should be based on the subject's projected work capacity, obtained during orientation. An example protocol based on an initial level of 2 METS with .55 MET increments is shown in table 4. The problems with bicycle ergometer given by Wilmore (chapter 4) apply to all ages. Some older adults have never ridden a bicycle and are uncomfortable with the seat, motion, and cadence of pedaling. Older adults may, however, feel safer on a bicycle because they are supported, can hold onto the bars, and can stop at any time just by not pedaling, whereas on a treadmill they must first communicate the wish to stop.

Table 2. Sample protocols based on heart rate during orientation

2 mph protocols	
Up to 30%	2.5 or 3 mph protocol
30–40%	Grades 0%, 6%, 12%, 16%, 18%, 20%
40–50%	Grades 0%, 4%, 8%, 10%, 12%
50–100%	Grades 0%, 2%, 4%, 6%, 8%
2.5 mph protocols	
Up to 30%	3 mph protocol
30–35%	Grades 0%, 6%, 12%, 15%, 18%, 21%
40–45%	Grades 0%, 3%, 6%, 9%, 12%
50–55%	Grades 0%, 3%, 6%, 7.5%, 9%
60–100%	Grades 0%, 1.5%, 3%, 4.5%, 6%
3.0 mph protocols	
Up to 25%	Use protocol suggested for younger adults
25–35%	Grades 0%, 5%, 10%, 15%, 17.5%
35–50%	Grades 0%, 5%, 7.5%, 10%, 12.5%
50–100%	Grades 0%, 2.5%, 5%, 7.5%

Notes: % heart rate =

$$\frac{(\text{orientation heart rate} - \text{resting heart rate})}{(\text{predicted max heart rate} - \text{resting heart rate})} \times 100.$$

These protocols are based on the predicted maximal MET level based on the orientation % heart rate, and are chosen to provide an 8–12 minute evaluation for that maximum. The protocols taper off as the subject approaches the predicted maximum. This is important both for safety and so that the protocol does not overshoot the subject's maximum. The percentage heart rate will often underestimate the maximum; if the subject does not reach maximum capacity at the end of the protocol listed, continue to raise the grades by the last increment. Approximations of predicted maximal heart rate have been estimated as 220 − age, but it has been found that this equation generally produces an underestimate for older adults. The prediction for maximal heart rate should be adjusted if the subject is on medications that affect heart rate.

During work capacity testing of older adults, it is especially important to consider both the safety and comfort of subjects. Subjects should be instructed to stretch their calves before treadmill tests, as leg cramps from walking uphill could terminate the test before maximal capacity is reached. Subjects should be offered water before and after the test. In our laboratory, we instruct subjects to ask for a drink of water during the test rather than discontinue it because they feel too dry. Gas samples should be taken during the last 30 seconds of each testing stage; during the first minute, noseclips should be removed to allow the subjects to swallow more easily and alleviate feelings of discomfort and claustrophobia. Evaluators should be constantly aware of the subject's gait, centering on the treadmill, signs of dizziness, and ECG.

We developed a chair step test for use at home or when treadmill or bicycle ergometry equipment is not available (Smith and Gilligan, 1983). This is a submaximal test of four stages of difficulty, done seated. Subjects "step" up and down once per second for 2 minutes each at step heights of 6, 12, and 18 inches. At the final stage, arm movements are added (table 5). We do not recommend this test for large-scale screening, as it is a submaximal test and has not been validated. As the highest workload is less than 4 METS, it is not appropri-

Table 3. Speed and grade at approximately 1/2 MET increments

	Speed (mph)		
METS	2.0[1]	2.5[2]	3.0[3]
2.50	0.0		
3.00	1.5	0.0	
3.50	3.5	1.5	0.5
4.00	5.0	3.0	1.5
4.50	7.0	4.5	2.5
5.00	9.0	6.0	4.0
5.50	10.5	7.5	5.0
6.00	12.5	9.0	6.5
6.50	14.0	10.0	7.5
7.00	16.0	11.5	9.0
7.50	18.0	13.0	10.0
8.00	19.5	14.5	11.0
8.50	21.5	16.0	12.5
9.00	23.5	17.5	13.5
9.50	25.0	19.0	15.0
10.00		20.5	16.0
10.50		22.0	17.5
11.00		23.5	18.5
11.50		25.0	19.5
12.00			21.0
12.50			22.0
13.00			23.5
13.50			24.5

[1] 1.0 MET increments to 75% heart rate reserve, followed by 0.5 MET increases.
[2] 1.5 MET increments to 75% heart rate reserve, followed by 0.5 MET increases.
[3] 2.0 MET increments to 75% heart rate reserve, followed by 1.0 MET increases.
Note: Grades computed to nearest 0.5 for given MET level.

ate for subjects with work capacities of approximately 8 METS or higher.

Flexibility and range of motion

Changes with age

Adrian (1981) reviewed changes with age in components of the joint. Changes in connective tissue with age result in more resistance to mechanical stress. In addition to connective tissue changes, over 80% of people aged 55–64 years have signs of osteoarthrosis (Kellgren and Lawrence, 1957). Adrian noted that there was little evidence on the extent to which biological aging causes the decrease in flexibility with age.

Table 4. Bicycle ergometer protocol with .55 MET increments

Weight		Initial load[1]		Load increment[1]	
Kg	Lb	Kg/min	Watts	Kg/min	Watts
50	110	25	4.25	50	8.5
60	132	50	8.5	50	8.5
70	154	100	16.6	75	12.5
80	176	125	21	75	12.5
90	198	175	27	75	12.5
100	220	200	33.3	100	16.6

[1] Values for the MET levels, based on weight and load, were determined by the formula VO_2 (l/min) = kg/min × 2 + 300 (ACSM, 1980, p. 146). The initial load is approximately 2 METS, and the load increment is approximately .55 METS. Values of kg/min are rounded to the nearest 25 because of the limitations of dial settings on bicycle ergometers.
Source: Reprinted by permission of the publisher from Smith, 1984.

Johns and Wright (1962) evaluated the relative contributions of tissue components to joint stiffness in cats' wrists. The torque required to move the joint in its midrange was 47% attributable to the joint capsule, 41% to passive motion of the muscles, 10% to tendons, and 2% to skin. Tendons contributed a greater proportion at the extreme range of motion.

Ahlback and Lindahl (1964) obtained goniometric measurements of hip flexion plus extension for 53 men aged 20–79. Mean flexibility was 161 degrees for those 20–29 years (*n* = 10) and decreased to 122 degrees for those 70–79 years (*n* = 7). Significance levels were not reported.

Range of motion (ROM) of the shoulder, hip, wrist, and thumb in two populations was studied by Allander et al. (1974). They measured 309 women from Iceland, 208 women from Sweden, and 203 men from Sweden, ages 33–70. Metacarpophalangeal joint I (MCP 1, flexion and extension) ROM was measured using a goniometer. Special devices were utilized to measure wrist (flexion and extension), shoulder (inward and outward rotation), and hip (internal and external rotation) ROM's. In the Icelandic survey, ROM decreased with age in both wrists, both hips, and the right shoulder. The ROM's for left shoulder and MCP I did not decline significantly with age. The Swedish survey also showed significant declines with age for the wrist, hip, and shoulders. Men had significantly smaller ROM than women in the wrist and shoulder. The right wrist had less ROM than the left, but the left hip rotation was greater than the right.

Boone and Azen (1979) measured ROM in male subjects 18 months to 54 years of age using guidelines for measurement of flexibility from the American Academy of Orthopaedic Surgeons (AAOS). Sites measured were the shoulder (horizontal extension, forward flexion, backward extension, inward and outward rotation); elbow (flexion and extension); forearm (pronation and supination); wrist (flexion, extension, radial deviation, ulnar deviation); hip (flexion, abduction, adduction, inward and outward rotation); knee (flexion); ankle (flexion and extension); and foot (inversion and eversion). Subjects aged 40–54 years had less elbow extension than younger age groups, and subjects aged 30–54 had less wrist flexion than younger age groups. Subjects 20–54 years (*n* = 56, mean age 34.9) had significantly less ROM than subjects under 19 (*n* = 53, mean age 9.2) on all motions listed here except shoulder horizontal flexion, shoulder neutral abduction, elbow extension, forearm pronation, radial and ulnar deviation, hip flexion, and ankle extension. No consistent pattern of differences between left and right motions was detected.

Spine mobility (left and right lateral flexion, anterior flexion and extension) in women aged 20–84 was evaluated by Einkauf et al. (1987). Volunteers were chosen to meet the criteria of no history of back surgery, no lower back pain that required medical treatment, no rheumatoid arthritis or osteoarthritis of the spine or hips, and

Table 5. Comparison of 3-mph Balke treadmill test, the modified Balke test, and the chair step test

3-mph Balke treadmill test [1]			2-mph modified Balke test [1]			Chair step test [2]		
Grade (%)	$\dot{V}O_2$ (ml·kg^{-1}·min^{-1})	METS	Grade (%)	$\dot{V}O_2$ (ml·kg^{-1}·min^{-1})	METS	Step-ht (in.)	$\dot{V}O_2$ (ml·kg^{-1}·min^{-1})	METS
0.0	10.5	3.0	0.0	7.0	2.0	6	8.0	2.3
2.5	14.0	4.0	2.0	8.9	2.5	12	10.0	2.9
5.0	17.5	5.0	4.0	10.8	3.1	18	12.3	3.5
7.5	21.0	6.0	6.0	12.7	3.6	18	13.7	3.9
10.0	24.5	7.0	8.0	14.6	4.2			
			10.0	16.5	4.7			

[1] Energy costs calculated from formulas provided in Balke B., Ware RW: An experimental study of physical fitness of Air Force personnel. *US Armed Forces Med J* 10:675–688, 1959.

[2] Energy costs obtained from unpublished data from our laboratory tests of oxygen consumption during the chair step test.

Source: Reprinted by permission of the publisher from Smith and Gilligan, 1983.

no scoliosis. A total of 109 women were assigned to six 10-year age groups, each containing at least 15 subjects. Right and left lateral flexion and spinal extension were measured following the guidelines of the AAOS. Anterior spinal flexion was measured by the modified Schober method. Left and right lateral flexion was approximately 33% lower in the women aged 70–79 than in those aged 20–29; anterior flexion was approximately 29% lower and extension 50% lower.

Germain and Blair (1983) studied shoulder flexion variation with age in 159 men and women aged 1–103. Subjects had no history of injury or of neuromuscular or skeletal lesions of the right shoulder. Shoulder flexion was measured with a standard goniometer. Subjects were classified as active if they participated in an upper extremity occupation or upper extremity sport 3 or more hours per week for at least the last 3 months. Shoulder flexion declined with age ($R = -0.70$). There was a 10-degree (6%) difference between subjects aged 20–30 ($n = 39$, 180 degrees) and subjects aged 50–70 ($n = 15$, 170 degrees), and a further 18-degree difference for subjects over 70 ($n = 13$, 152 degrees). Their final regression model was: Shoulder flexion (degrees) = 190.5 − 0.43 × age + 2.43 (if active). There were no significant differences between males and females.

In another cross-sectional study that considered physical activity level, Rikli and Busch (1986) measured 60 (15 per group) inactive and active young (mean age 22) and old (mean age 69) women. They also reported on 15 old golfers (mean age 70). Subjects in the active groups exercised vigorously at least 3 times/week. Trunk flexibility was measured by a sit-and-reach test. For determination of shoulder flexibility, subjects put a hand over one shoulder and attempted to touch or overlap fingers of the other hand behind the back. Both age and activity level had significant main effects on shoulder flexibility and trunk flexibility. The age × activity interaction was significant for shoulder flexibility, for which there was a greater difference between active and inactive older women than between the two groups of younger women. The golfers were intermediate in flexibility compared to old active and inactive groups. Scores of golfers were not significantly lower than those of the active older women.

Hageman and Blanke (1986) studied free-speed gait patterns in young ($n = 13$, ages 20–35) and elderly ($n = 13$, ages 60–84). In addition to decreased velocity, older women had significantly lower stride length and ankle ROM (24.62 versus 31.31 degrees) during free walking. In a similar study, Murray et al. (1964) reported a reduction of ankle movement during walking in elderly men compared to young men.

Joseph (1954) studied motion of the metatarsophalangeal (MP) and interphalangeal (IP) joints of the great toe in three age groups of men: under 30 ($n = 17$), 30–44 ($n = 17$), and 45 and over ($n = 16$). Active flexion and active and passive dorsiflexion were evaluated by lateral radiographs. Active and passive dorsiflexion did not differ among age groups. Active MP plantar flexion was lower in the group 45 and over than in the younger groups. Active plantar flexion of the IP was also reduced in this group. The youngest group had greater IP plantar flexion in the right foot than other groups. Right and left sides did not differ significantly.

Hung et al. (1985) measured ankle (plantar flexion, dorsiflexion, inversion, and eversion) and toe (metatarsophalangeal and interphalangeal flexion and extension) ROM as part of a survey of foot deformities in 166 geriatric inpatients (ages 65–98). MP joint extension was limited (<40 degrees) in 15% of patients, and flexion was limited (<20 degrees) in more than 50% of patients. Eversion of the whole foot was under 10 degrees for 30–40% of patients, and only 5% of patients had less than 20 degrees inversion. Dorsiflexion was under 10 degrees in over 70% of patients, and plantar flexion was less than 20 degrees in 20% of patients.

In an attempt to provide age- and sex-related standards for shoulder motion, Murray (1985) measured active ranges of motion in 40 men and women of two age groups, 25–36 and 55–66, 10 subjects per group. None of the subjects had a history of shoulder joint pain or dysfunction. Shoulder flexion, extension, abduction, and inward and outward rotation were evaluated according to AAOS guidelines. Men had lower ROM than women in the same age group on a number of evaluations: Glenohumeral abduction and total inward and outward rotation in the young subjects, and flexion, outward rotation, and total rotation in the older subjects.

Younger men had significantly more flexion (170 versus 165) and outward extension (101 versus 94), but less inward rotation (45 versus 59) than older men. Younger women had significantly greater glenohumeral abduction than older women (129 versus 123).

A later study from the same laboratory (Sepic et al., 1986) reported on ankle (plantar flexion and dorsiflexion) and subtalar (inversion and eversion) ROM measured by goniometer. Forty subjects were divided equally by age and sex into four groups. None of the subjects had a history of musculoskeletal problems in the lower extremities, and none were runners or participated in a program of strengthening exercises for the lower limbs. On the average, women and younger subjects (ages 25–35) had greater plantar flexion than men and older subjects (ages 50–60). Plantar flexion was approximately 55 degrees in younger women, 50 degrees in older women and younger men, and 45 degrees in older men. No other significant differences among age or sex groups were found. There were no significant differences between left and right limbs.

In a study to determine whether flexibility declined in persons 55 and over, Smith and Walker (1983) studied 60 healthy older men and women in the age groups 55–64, 65–74, and 75–84. Subjects were recruited from mobile home parks (*n* = 37), community centers (*n* = 15), and health clubs (*n* = 3) and were screened for musculoskeletal disorders. Knee and elbow ranges of motion, active and passive, were measured by goniometer according to AAOS guidelines. Elbow flexion and extension and knee extension did not differ among age groups. Knee flexion, both active and passive, was slightly lower in the oldest group of women than in the younger groups but did not differ with age in the men. Women had slightly but significantly greater elbow flexion than men. Mean flexibility in this study was similar to that in the Boone and Azen study of men 20–54 years. Activity level, measured by questionnaire, was not correlated with range of motion, and no difference between left and right limbs was detected.

Walker (1984) also reported that age did not affect flexibility in the age range 60–84. Thirty men and 30 women were equally divided into two age groups: 60–69 and 75–84. All subjects lived independently and were screened for pathological conditions and ROM limitations on daily activity. Subjects with mild arthritis were not excluded. Motions of the shoulder (abduction, flexion, extension, and medial and lateral rotation); elbow (flexion); forearm (pronation, supination); wrist (flexion, extension, radial and ulnar deviation); hip (flexion, abduction, adduction, internal and external rotation); knee (flexion); ankle (plantar flexion, dorsiflexion); foot (subtalar inversion and eversion); and first metatarsal (metatarso-phalangeal extension and flexion) were assessed using AAOS guidelines. Differences between the two age groups were greater than intertester error only for subtalar eversion and metatarsophalangeal flexion. With age groups combined, women generally had greater flexibility than men and had significantly higher shoulder (abduction, flexion, extension, and lateral rotation), wrist (extension), hip (medial rotation), knee (flexion), and ankle (plantar flexion) measurements. Men had greater ROM for radial-ulnar supination, wrist extension, ulnar deviation, and hip adduction. The authors noted that differences between this study and AAOS measures exceeded one standard deviation and intertester error for a number of motions, with the largest difference in metatarsophalangeal flexion (31 degrees lower) and hip abduction (25 degrees lower). Again, there was no consistent relationship between ROM and leisuretime activities, assessed by questionnaire.

West (1945) reported on shoulder, elbow, and wrist flexibility in the contralateral joint of patients at the Minneapolis General Hospital fracture clinic and physical therapy departments. Subjects were divided into 10-year age groups. Shoulder flexion declined from 155 in the group aged 20–29 (*n* = 30) to 140 in the group 60–69 years (*n* = 93) but improved slightly in the group 70–79 years (145, *n* = 48). Shoulder extension also declined slightly, from 75 degrees for ages 20–49 (*n* = 98) to 70 degrees for ages 50–79 (*n* = 234). Elbow flexion and extension and wrist flexion and extension did not differ by more than 5 degrees between age groups. Significance levels were not reported.

Effect of exercise

Buccola and Stone (1975) enlisted 36 men ages 60–79 to participate in a 14-week walk-jog (*n* = 16) or cycling program (*n* = 20). Subjects trained 25–50 minutes/day, 3 days/week. Trunk and leg flexibility was measured by the sit-and-reach test. Joggers improved significantly in flexibility, from 7.31 to 8.75 inches, but cyclers did not change significantly.

Chapman et al. (1972) studied the effects of exercise on joint stiffness in young (*n* = 20, ages 15–19) and old (*n* = 20, ages 63–88) men. Torque and energy during passive movement of the index fingers were evaluated. Subjects trained for 6 weeks, with 10 lifts of one-half, three-quarters, and maximal loads on the experimental finger 3 times/weeks. Initial joint stiffness was significantly greater, by approximately 30%, in old men than in the young. The experimental finger decreased significantly in torque with exercise, but there was no significant change in the control finger. There was no significant effect of age on improvement in flexibility, and old subjects still had greater stiffness than young subjects at the end of the study.

Fifteen women aged 71–90 recruited by Frekany and Leslie (1975) exercised 30 minutes/session, 2 times/week for 7 months. Subjects were recruited from nursing homes or a Golden Age Club. Ankle (plantar flexion plus dorsiflexion) and hamstring-lower back (sit-and-reach test) were measured before and after the exercise program. Left ankle flexibility improved significantly, from 38 to 47.5 degrees, and right ankle flexibility improved from 36.3 to 50.7 degrees. Sit-and-reach flexibility also improved significantly, from 11.3 to 12.3 inches. Ankle ROM was measured by a different method than is used in most studies, and measurements are much lower than reported elsewhere, even for old subjects. Conversely, sit-and-reach measurements are very high compared with other sources.

Germain and Blair (1983) selected 30 volunteers ages 20–60 to participate in an experiment on the effect of exercise training on shoulder flexion. Fifteen subjects (mean age 34.4) were assigned to an exercise program, and 15 subjects (mean age 34.8) served as controls. Training consisted of seven repetitions of a shoulder flexion stretching exercise daily for 4 weeks. At the conclusion of the program, control subjects had not changed significantly in shoulder flexion. Exercise subjects had a mean increase of 5.61 degrees in shoulder flexion, significantly greater than that in the control group.

Gutman et al. (1977) recruited male and female subjects from nursing homes and divided them into four groups: Conventional exercise (*n* = 13), Feldenkrais exercise (*n* = 19), and two control groups (*n* = 16 and *n* = 19). Mean ages were 71–72 years. The exercise programs were conducted for 1 hour three times per week for 6 weeks. Rotational flexibility was measured by having subjects turn as far to the left and right as possible without foot movements. Flexibility was greater at the post-test than the pretest, but there were no significant differences between groups in amount of improvement.

Lesser (1978) divided 60 male and female volunteers (mean ages 61–79 and mean age 75, respectively) from nutritional centers into exercise and control groups. Exercise subjects participated in a 30-minute session twice per week for 10 weeks. A goniometer was used to assess shoulder, elbow, wrist, hip, knee, and ankle flexion and extension. The author reported that exercise but not control subjects improved on 8 of 12 measures, but no mean measurements or statistical significance levels were presented.

Munns (1981) assigned 40 subjects (ages 65–88, mean age 72) randomly into exercise and control groups. Subjects were recruited from a senior citizens' community recreation center. A Leighton flexometer was used to assess neck (flexion and extension), shoulder (abduction and adduction), wrist (flexion and extension), knee (flexion and extension), hip-back (flexion and extension) and ankle (flexion and extension) ranges of motion. The two groups did not differ significantly at the beginning of the study, but at the end, exercise group ROM was significantly greater than control group ROM for all sites.

Raab et al. (in press) evaluated the effect of exercise with and without weights on elderly women. Subjects were recruited from the community and formed into three groups: control (*n* = 13, mean age 73), exercise (*n* = 16, mean age 70) and exercise with weights (*n* = 17, mean age 70). The exercise groups participated in 1-hour sessions thrice weekly for 25 weeks. Subjects in the group exercising with weights were gradually introduced to wrist and ankle weights that increased in weight over the course of the study. Flexibility of the shoulder (flexion and abduction), ankle (plantar flexion and dorsiflexion), hip (flexion), wrist (flexion and extension), and cervical rotation (left and right) were measured by goniometer, with some variation from AAOS guidelines. Of the initial measures, control and exercise groups differed only in ankle plantar flexion. The exercise groups improved significantly compared with controls in ankle plantar flexion, shoulder flexion, shoulder abduction, and left cervical rotation. Hip flexion improved in all groups. The two exercise groups differed only in improvement of shoulder abduction.

Summary

Table 6 details the measurements obtained for a variety of joints on different age ranges. Between-study comparisons should be made with caution because of differences in measurement technique and subject screening methods.

Shoulder—Shoulder forward flexion was lower in older age groups in the studies by Boone and Azen (1979) and Germain and Blair (1983). Older men had lower shoulder flexion than younger men in the study of Murray (1985), but younger and older women did not differ. Walker (1984) found no difference between groups aged 60–69 and 75–84 years. The lowest values were reported by Raab et al. for women over 63; subjects were not excluded for musculoskeletal problems in that study.

Shoulder extension was lower in men 20–54 years than in the age group 1–19 years in the study by Boone and Azen (1979), but no age difference was detected by Murray (1985) or Walker (1984). Values for subjects 60–84 years studied by Walker were 12–17 degrees lower than those of the older groups in Murray. This may be partly caused by more stringent screening in the Murray study. Shoulder flexion plus extension, measured by Munns (1981) using the Leighton flexometer, was lower in women (mean age 72) than flexion alone in studies utilizing a goniometer, and probably reflects differences in measurement technique.Boone and Azen (1979) were the only researchers to measure shoulder horizontal flexion and extension. Extension, but not flexion, was lower in the older age group.

Shoulder abduction was lower in the age group 20–54 years than ages 1–19 (Boone and Azen, 1979) but did not differ significantly between age groups 25–35 and 56–66 (Murray, 1985) or between ages 60–69 and 75–84 (Walker, 1984). Men 60–84 years in the Walker study had shoulder abductions 23 degrees lower than men 56–66 years in the Murray study. Measurements of women in these two studies differed by only 3 degrees. Shoulder abduction in women over 63 measured by Raab et al. (in press) was 9 degrees lower than the measurement for women 60–64 years in the Murray study (1985). Again, these differences between studies may reflect screening differences.

Table 6. Study findings on joint measurements

		Population			Age		
Source and method	Side	Sex	Age	N	Mean and SD	Mean ± SD (degrees)	Range
Shoulder forward flexion							
AAOS, (a)			Four sources			158	(130–180)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	168.4 ± 3.7	
		M	20–54	56	34.9 ± 3.4	165.0 ± 5.0	
Germain and Blair, 1983 (a)		B	10–20	19		183.8 ± 6.9	
		B	20–30	39		179.3 ± 9.8	
		B	30–40	24		177.3 ± 7.9	
		B	40–50	10		176.0 ± 7.2	
		B	50–70	15		169.6 ± 11.1	
		B	70+	13		152.0 ± 16.9	
Murray, 1985 (a)		M	26–36	10	31	170 ± 6	
		M	56–66	10	62	165 ± 6	
		F	25–35	10	29	172 ± 3	
		F	60–64	10	62	170 ± 3	
Raab et al., in press (a)		F	63+	46	70.8 ± 5.3	148.4 ± 13.3	
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	160 ± 11	
		F	60–84	30	72.7 ± 8.2	169 ± 9	
West, 1945 (b)		B	20–29	30		155	(135–175)
		B	30–39	28		150	(130–175)
		B	40–49	43		150	(115–175)
		B	50–59	88		145	(115–175)
		B	60–69	93		140	(115–175)
		B	70–79	48		145	(120–170)
Shoulder extension							
AAOS, (a)			Four sources			53	(30–80)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	67.5 ± 8.0	
		M	20–54	56	34.9 ± 3.4	57.3 ± 8.1	
Murray, 1985 (a)		M	26–36	10	31	57 ± 9	
		M	56–66	10	62	55 ± 9	
		F	25–35	10	29	58 ± 9	
		F	60–64	10	62	61 ± 9	
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	38 ± 11	
		F	60–84	30	72.7 ± 8.2	49 ± 13	
West, 1945 (b)		B	20–29	30		75	(60–90)
		B	30–39	26		75	(50–90)
		B	40–49	42		75	(55–90)
		B	50–59	88		70	(45–85)
		B	60–69	95		70	(55–90)
		B	70–79	51		70	(50–80)
Shoulder flexion and extension							
Munns, 1981 (c)		F	65–88	40	72	138.5 ± 9.3	
Shoulder horizontal flexion							
AAOS, (a)			Four sources			135	
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	140.8 ± 6.8	
		M	20–54	56	34.9 ± 3.4	140.7 ± 4.9	
Shoulder horizontal extension							
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	47.3 ± 6.1	
		M	20–54	56	34.9 ± 3.4	43.7 ± 5.8	

Table 6. Study findings on joint measurements—Continued

		Population			Age		
Source and method	Side	Sex	Age	N	Mean and SD	Mean ± SD (degrees)	Range
Shoulder abduction							
AAOS, (a)			Four sources			170	(150–180)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	185.4±3.6	
		M	20–54	56	34.9±3.4	182.7±9.0	
Murray, 1985 (a)		M	26–36	10	31	178±3	
		M	56–66	10	62	178±3	
		F	25–35	10	29	180±3	
		F	60–64	10	62	178±3	
Raab et al., in press (a)		F	63+	46	70.8±5.3	169.4±11.7	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	155±22	
		F	60–84	30	72.7±8.2	175±16	
Shoulder adduction							
AAOS, (a)			Four sources			50	(30–75)
Shoulder glenohumeral abduction							
Murray, 1985 (a)		M	26–36	10	31	122±6	
		M	56–66	10	62	122±6	
		F	25–35	10	29	129±6	
		F	60–64	10	62	123±6	
Shoulder abduction + adduction							
Munns, 1981 (c)		F	65–88	40	72	141.1±10.5	
Shoulder internal rotation (arm at side)							
AAOS, 1965 (a)			Four sources			68	(40–90)
Shoulder external rotation (arm at side)							
AAOS, (a)			Four sources			68	(40–90)
Shoulder internal rotation (at 90° abduction)							
AAOS, (a)			Four sources			70	
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	70.5± 4.5	
		M	20–54	56	34.9±3.4	67.1±4.1	
Murray, 1985 (a)		M	26–36	10	31	49±9	
		M	56–66	10	62	59±6	
		F	25–35	10	29	53±9	
		F	60–64	10	62	56±6	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	59±16	
		F	60–84	30	72.7±8.2	66±13	
Shoulder external rotation (at 90° abduction)							
AAOS, (a)			Four sources			90	
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	108.0±7.2	
		M	20–54	56	34.9±3.4	99.6±7.6	
Murray, 1985 (a)		M	26–36	10	31	94±6	
		M	56–66	10	62	82±13	
		F	25–35	10	29	101±6	
		F	60–64	10	62	94±6	

Table 6. Study findings on joint measurements—Continued

		Population			Age	Mean ± SD	
Source and method	Side	Sex	Age	N	Mean and SD	(degrees)	Range
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	76 ± 13	
		F	60–84	30	72.7 ± 8.2	85 ± 16	
Shoulder internal + external rotation (at 90° abduction)							
Murray, 1985 (a)		M	26–36	10	31	143 ± 9	
		M	56–66	10	62	141 ± 13	
		F	25–35	10	29	154 ± 9	
		F	60–64	10	62	150 ± 9	
Allander et al., 1974 (d)		F	33–39	50	35	119/119	(L/R)
(Iceland)		F	41–45	57	43	121/124	
		F	46–49	77	47	120/120	
		F	50–53	67	51	118/119	
		F	55–60	58	56	114/114	
(Sweden)		F	45–	18		119/117	
		F	50–	43		124/122	
		F	55–	45		120/115	
		F	60–	43		115/116	
		F	65–	44		120/112	
		F	70–	12		104/104	
		M	45–	18		122/121	
		M	50–	45		116/117	
		M	55–	45		111/112	
		M	60–	41		106/109	
		M	65–	40		108/112	
		M	70–	14		108/115	
Elbow flexion							
AAOS, (a)			Four sources			146	(135–150)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	145.4 ± 5.3	
		M	20–54	56	34.9 ± 3.4	140.5 ± 4.9	
Smith and Walker, 1983 (a)		M	55–84	30	69.4 ± 7.9	146 ± 4	(Active)
		M	55–84	30	69.4 ± 7.9	150 ± 4	(Passive)
		F	55–84	30	69.7 ± 8.7	150 ± 4	(Active)
		F	55–84	30	69.7 ± 8.7	154 ± 4	(Passive)
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	139 ± 14	
		F	60–84	30	72.7 ± 8.2	148 ± 5	
West, 1945 (b)		B	20–29	12		145	(125–155)
		B	30–39	15		145	(140–155)
		B	40–49	19		140	(135–150)
		B	50–59	34		145	(135–155)
		B	60–69	24		140	(130–150)
		B	70–79	18		145	(140–145)
Elbow extension							
AAOS, 1965 (a)			Four sources			0	(0–0)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	0.8 ± 3.5	
		M	20–54	56	34.9 ± 3.4	0.3 ± 2.7	
Smith and Walker, 1983 (a)		M	55–84	30	69.4 ± 7.9	–1 ± 3	
		M	55–84	30	69.4 ± 7.9	0 ± 2	
		F	55–84	30	69.7 ± 8.7	0 ± 1	
		F	55–84	30	69.7 ± 8.7	–1 ± 2	

Table 6. Study findings on joint measurements—Continued

Source and method	Side	Population			Age	Mean ± SD (degrees)	Range
		Sex	Age	N	Mean and SD		
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	–6 ± 5	(start flexion)
		F	60–84	30	72.7 ± 8.2	–1 ± 3	(start flexion)
West, 1945 (b)		B	20–29	12		0	(–5 ± 10)
		B	30–39	16		0	(0 ± 5)
		B	40–49	16		0	(–20 ± 10)
		B	50–59	34		0	(–5 ± 5)
		B	60–69	23		–5	(–10 ± 5)
		B	70–79	17		0	(–20 ± 15)
Forearm pronation							
AAOS, (a)			Four sources			71	(50–80)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	76.7 ± 4.8	
		M	20–54	56	34.9 ± 3.4	75.0 ± 5.3	
Forearm supination							
AAOS, (a)			Four sources			84	(80–90)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	83.1 ± 3.4	
		M	20–54	56	34.9 ± 3.4	81.1 ± 4.0	
Wrist flexion							
AAOS, (a)			Four sources			73	(70–80)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	78.2 ± 5.5	
		M	20–54	56	34.9 ± 3.4	74.8 ± 6.6	
Cobe, 1928 (b)	R	M	<30	100	23.5	80.4	(52–99)
	L	M	<30	100	23.5	81.3	(61–95)
	R	F	<30	15	22.8	84.6	(72–96)
	L	F	<30	15	22.8	80.2	(67–94)
Raab et al., in press (a)		F	63+	46	70.8 ± 5.3	79.2 ± 6.7	
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	62 ± 12	
		F	60–84	30	72.7 ± 8.2	65 ± 8	
West, 1945 (b)		B	20–29	42		60	(40–85)
		B	30–39	83		65	(40–85)
		B	40–49	123		60	(40–85)
		B	50–59	279		65	(35–85)
		B	60–69	190		60	(35–90)
		B	70–79	110		60	(35–75)
Wrist extension							
AAOS, (a)			Four sources			71	(60–90)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	75.8 ± 6.1	
		M	20–54	56	34.9 ± 3.4	74.0 ± 6.6	
Cobe, 1928 (b)	R	M	<30	100	23.5	66.4	(38–92)
	L	M	<30	100	23.5	70.0	(40–90)
	R	F	<30	15	22.8	71.0	(31–95)
	L	F	<30	15	22.8	76.6	(50–95)
Raab et al., in press (a)		F	63+	46	70.8 ± 5.3	74.9 ± 7.6	
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	61 ± 6	
		F	60–84	30	72.7 ± 8.2	65 ± 10	
West, 1945 (b)		B	20–29	43		80	(60–105)
		B	30–39	83		85	(60–100)

Table 6. Study findings on joint measurements—Continued

		Population			Age		
Source and method	Side	Sex	Age	N	Mean and SD	Mean ± SD (degrees)	Range
		B	40–49	129		80	(60–105)
		B	50–59	284		80	(50–105)
		B	60–69	198		80	(50–105)
		B	70–79	111		80	(50–100)
Wrist flexion + extension							
Allander et al., 1974 (d)		F	33–39	50	35	141/145	(L/R)
(Iceland)		F	41–45	57	43	138/143	
		F	46–49	77	47	133/139	
		F	50–53	67	51	132/135	
		F	55–60	58	56	127/133	
(Sweden)		F	45–	18		127/136	
		F	50–	43		122/129	
		F	55–	45		121/127	
		F	60–	43		114/122	
		F	65–	44		117/124	
		F	70–	12		104/114	
		M	45–	18		121/126	
		M	50–	45		116/122	
		M	55–	45		115/120	
		M	60–	41		109/116	
		M	65–	40		109/115	
		M	70–	14		111/116	
Cobe, 1928 (b)	R	M	<30	100	23.5	133.0	(107–176)
	L	M	<30	100	23.5	154.5	(123–177)
	R	F	<30	15	22.8	155.0	(127–178)
	L	F	<30	15	22.8	160.0	(126–184)
Munns, 1981 (c)		F	65–88		72	87.3±16.8	
Wrist–radial deviation							
AAOS, (a)			Four sources			19	(15–20)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	21.8±4.0	
		M	20–54	56	34.9±3.4	21.1±4.0	
Cobe, 1928 (b) (in pronation)	R	M	<30	100	23.5	43.4	(22–77)
	L	M	<30	100	23.5	44.4	(21–71)
	R	F	<30	15	22.8	50.9	(36–75)
	L	F	<30	15	22.8	48.8	(24–59)
Cobe, 1928 (b) (in supination)	R	M	<30	100	23.5	26.9	(7–68)
	L	M	<30	100	23.5	32.2	(12–62)
	R	F	<30	15	22.8	30.8	(12–54)
	L	F	<30	15	22.8	35.9	(14–49)
Walker, 1984 (a)		M	60–84	30	72.4±6.8	20±6	
		F	60–84	30	72.7±8.2	17±6	
Wrist–ulnar deviation							
AAOS (a)			Four sources			33	(30–40)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	36.7±3.7	
		M	20–54	56	34.9±3.4	35.3±3.8	
Cobe, 1928 (b) (in pronation)	R	M	<30	100	23.5	32.8	(10–65)
	L	M	<30	100	23.5	33.6	(18–58)
	R	F	<30	15	22.8	37.2	(22–78)
	L	F	<30	15	22.8	34.8	(24–58)

Table 6. Study findings on joint measurements—Continued

		Population			Age		
Source and method	Side	Sex	Age	N	Mean and SD	Mean ± SD (degrees)	Range
Cobe, 1928 (b)	R	M	<30	100	23.5	61.5	(43–88)
(in supination)	L	M	<30	100	23.5	56.3	(20–88)
	R	F	<30	15	22.8	66.3	(52–82)
	L	F	<30	15	22.8	60.7	(43–83)
Walker, 1984 (a)		M	60–84	30	72.4±6.8	28±7	
		F	60–84	30	72.7±8.2	23±7	
Thumb (MCP I flexion + extension)							
Allander et al., 1974 (d)		F	33–39	50	35	52/54	(L/R)
(Iceland)		F	41–45	57	43	50/52	
		F	46–49	77	47	51/53	
		F	50–53	67	51	47/49	
		F	55–60	58	56	50/52	
(Sweden)		F	45–	18		51/56	
		F	50–	43		50/52	
		F	55–	45		45/52	
		F	60–	43		52/52	
		F	65–	44		50/54	
		F	70–	12		43/45	
		M	45–	18		48/52	
		M	50–	45		47/52	
		M	55–	45		45/49	
		M	60–	41		44/46	
		M	65–	40		46/50	
		M	70–	14		51/54	
Hip flexion							
AAOS, (a)			Four sources			113	(112–120)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	123.4±5.6	
		M	20–54	56	34.9±3.4	121.3±6.4	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	110±11	
		F	60–84	30	72.7±8.2	111±12	
Hip extension							
AAOS (a)			Four sources			28	(20–30)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	7.4±7.3	
		M	20–54	56	34.9±3.4	12.1±5.4	
Hip flexion + extension							
Ahlback and Lindahl, 1964 (b)		M	20–29	10		161	(141–178)
		M	30–39	8		153	(128–165)
		M	40–49	9		151	(129–167)
		M	50–59	10		136	(115–143)
		M	60–69	9		132	(116–141)
		M	70–79	7		122	(107–143)
Munns, 1981 (c)		F	65–88	40	72	97.2±18.0	
Hip abduction							
AAOS (a)			Four sources			48	(40–55)
AAOS (a) (at 90° flexion)			Four sources				(45–60)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	51.7±8.8	
		M	20–54	56	34.9±3.4	40.5±6.0	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	23±9	
		F	60–84	30	72.7±8.2	24±6	

Table 6. Study findings on joint measurements—Continued

		Population			Age		
Source and method	Side	Sex	Age	N	Mean and SD	Mean ± SD (degrees)	Range
Hip adduction							
AAOS (a)			Four sources			31	(20–45)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	28.3 ± 4.1	
		M	20–54	56	34.9 ± 3.4	25.6 ± 3.6	
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	18 ± 4	
		F	60–84	30	72.7 ± 8.2	11 ± 4	
Hip internal rotation							
AAOS (a) (in flexion)			Four sources			45	
AAOS (a) (in extension)			Four sources			35	(20–45)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	50.3 ± 6.1	
		M	20–54	56	34.9 ± 3.4	44.4 ± 4.3	
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	22 ± 6	
		F	60–84	30	72.7 ± 8.2	36 ± 7	
Hip external rotation							
AAOS (a)			Four sources			45	
AAOS (a)			Four sources			48	(45–50)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	50.5 ± 6.1	
		M	20–54	56	34.9 ± 3.4	44.2 ± 4.8	
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	32 ± 6	
		F	60–84	30	72.7 ± 8.2	30 ± 7	
Hip internal + external rotation							
Allander et al., 1974 (d)		F	33–39	50	35	77/75	(L/R)
(Iceland)		F	41–45	57	43	77/74	
		F	46–49	77	47	76/72	
		F	50–53	67	51	72/68	
		F	55–60	58	56	71/70	
(Sweden)		F	45–	18		68/69	
		F	50–	43		64/68	
		F	55–	45		59/65	
		F	60–	43		59/62	
		F	65–	44		58/59	
		F	70–	12		49/53	
		M	45–	18		68/64	
		M	50–	45		66/53	
		M	55–	45		61/58	
		M	60–	41		60/56	
		M	65–	40		63/55	
		M	70–	14		60/56	
Knee flexion							
AAOS (a)			Four sources			134	(120–145)
Boone and Azen, 1979 (a)		M	1–19	53	9.2 ± 1.7	143.8 ± 5.1	
		M	20–54	56	34.9 ± 3.4	141.2 ± 5.3	
Smith & Walker, 1983 (a)		M	55–84	30	69.4 ± 7.9	141 ± 5	(active)
		F	55–64	10	59.3 ± 2.5	143 ± 3	(active)
		F	65–74	10	70.4 ± 2.5	144 ± 3	(active)
		F	75–84	10	79.5 ± 2.5	140 ± 4	(active)
		M	55–84	30	69.4 ± 7.9	146 ± 5	(passive)

Table 6. Study findings on joint measurements—Continued

		Population			Age		
Source and method	Side	Sex	Age	N	Mean and SD	Mean ± SD (degrees)	Range
		F	55–64	10	59.3±2.5	147±3	(passive)
		F	65–74	10	70.4±2.5	149±4	(passive)
		F	75–84	10	79.5±2.5	145±5	(passive)
Walker, 1984 (a)		M	60–84	30	72.4±6.8	131±4	
		F	60–84	30	72.7±8.2	135±4	
Knee extension							
AAOS (a)			Four sources			10	
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	2.1±3.2	(begin flex)
		M	20–54	56	34.9±3.4	1.1±2.0	(begin flex)
Smith and Walker, 1983 (a)		M	55–84	30	69.4±7.9	1±1	(active)
		M	55–84	30	69.4±7.9	0±1	(passive)
		F	55–84	30	69.7±8.7	0±1	(active)
		F	55–84	30	69.7±8.7	-1±1	(passive)
Walker, 1984 (a)		M	60–84	30	72.4±6.8	2±2	(begin flex)
		F	60–84	30	72.7±8.2	0±1	(begin flex)
Knee flexion + extension							
Munns, 1981 (c)	F	65–88	40	72	110.0±15.4		
Ankle plantar flexion							
AAOS (a)			Four sources			48	(40–50)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	58.2±6.1	
		M	20–54	56	34.9±3.4	54.3±5.9	
Hung et al., 1985 (b)		B	65–98	166	(ger inpat)	30–50	(median)
Raab et al., in press (b)		F	63+	46	70.8±5.3	68.4±6.6	
Sepic et al., 1986 (b)		M	25–35	10	32.0±2.6	50	
		M	50–60	10	56.5±3.2	55	
		F	25–35	10	29.1±3.2	45	
		F	50–60	10	54.8±4.0	50	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	29±7	
		F	60–84	30	72.7±8.2	40±6	
Ankle dorsiflexion (extension)							
AAOS (a)			Four sources			18	(15–20)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	13.0±4.7	
		M	20–54	56	34.9±3.4	12.2±4.1	
Hung et al., 1985 (b)		B	65–98	166	(ger inpat)	0–10	(median)
Raab et al., in press (b)		F	63+	46	70.8±5.3	10.3±3.9	
Sepic et al., 1986 (b)		M	25–35	10	32.0±2.6	23	
		M	50–60	10	56.5±3.2	24	
		F	25–35	10	29.1±3.2	25	
		F	50–60	10	54.8±4.0	23	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	9±5	
		F	60–84	30	72.7±8.2	10±5	
Ankle flexion + extension							
Frekany and Leslie, 1975 (d)	R	F	71–90	15		38.0	
	L	F	71–90	15		36.3	

Table 6. Study findings on joint measurements—Continued

Source and method	Side	Population			Age	Mean ± SD (degrees)	Range
		Sex	Age	N	Mean and SD		
Hageman and Blanke, 1986 (d)		F	20–33	13	23.9±3.6	31.3±5.2	(normal walking)
Munns, 1981 (c)		F	60–84	13	66.9±7.6	24.6±4.6	
		F	65–68	40	72	24.8±6.0	
Hindfoot (subtalar) inversion							
AAOS (a)			Four sources			5	
Sepic et al., 1986 (b)		M	25–35	10	32.0±2.6	13	
		F	25–35	10	29.1±3.2	18	
		M	50–60	10	56.5±3.2	18	
		F	50–60	10	54.8±4.0	18	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	31±11	
		F	60–84	30	72.7±8.2	29±10	
Hindfoot (subtalar) eversion							
AAOS (a)			Four sources			5	
Sepic et al., 1986 (b)		M	25–35	10	32.0±2.6	10	
		M	50–60	10	56.5±3.2	11	
		F	25–35	10	29.1±3.2	11	
		F	50–60	10	54.8±4.0	10	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	13±6	
		F	60–84	30	72.7±8.2	12±5	
Forefoot inversion							
AAOS (a)			Four sources			33	(30–35)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	37.5±4.7	
		M	20–54	56	34.9±3.4	36.2±4.2	
Forefoot eversion							
AAOS (a)			Four sources			18	(15–20)
Boone and Azen, 1979 (a)		M	1–19	53	9.2±1.7	22.3±4.6	
		M	20–54	56	34.9±3.4	19.2±4.9	
Foot inversion							
Hung et al., 1985 (b)		B	65–98	166	(ger inpat)	21–30	(median)
Foot eversion							
Hung et al., 1985 (b)		B	65–98	166	(ger inpat)	11–20	(median)
Great toe–interphalangeal joint flexion							
AAOS (a)			Four sources			60	(30–90)
Hung et al., 1985 (b)		B	65–98	166	(ger inpat)	21–40	(median)
Joseph, 1954 (d)	R	M	<30	17		55.4±11.5	
	L	M	<30	17		51.1±14.8	
	R	M	30–44	17		41.7±15.3	
	L	M	30–44	17		48.5±14.4	
	R	M	45+	16		42.6±17.2	
	L	M	45+	16		37.4±16.8	
Great toe–interphalangeal joint extension							
AAOS (a)			Four sources			0	
Hung et al., 1985 (b)		B	65–98	166	(ger inpat)	0–20	(median)
Joseph, 1954 (d)	R	M	<30	17		16.0±13.2	(active)

Table 6. Study findings on joint measurements—Continued

		Population			Age	Mean ± SD (degrees)	
Source and method	Side	Sex	Age	N	Mean and SD		Range
	L	M	<30	17		15.7±14.4	(active)
	R	M	30–44	17		12.8±13.6	(active)
	L	M	30–44	17		10.2±9.9	(active)
	R	M	45+	16		7.3±14.4	(active)
	L	M	45+	16		9.8±18.0	(active)
	R	M	<30	17		16.0±8.7	(additional passive)
	L	M	<30	17		18.6±9.1	(additional passive)
	R	M	30–44	17		18.4±6.6	(additional passive)
	L	M	30–44	17		18.1±6.6	(additional passive)
	R	M	45+	16		22.4±12.8	(additional passive)
	L	M	45+	16		20.6±10.4	(additional passive)
Great toe–metatarsophalangeal (proximal) joint flexion							
AAOS (a)			Four sources			37	(30–45)
Hung et al., 1985 (b)		B	65–98	166	(ger inpat)	0–40	(median)
Joseph, 1954 (d)		M	<30	17		25.6±7.4	
		M	<30	17		24.5±7.8	
		M	30–44	17		23.9±9.1	
		M	30–44	17		23.3±7.8	
		M	45+	16		17.6±9.2	
		M	45+	16		21.1±9.6	
Walker, 1984 (a)		M	60–84	30	72.4±6.8	5±7	
		F	60–84	30	72.7±8.2	8±16	
Great toe—metatarsophalangeal joint extension							
AAOS (a)			Four sources				
Hung et al., 1985 (b)		B	65–98	166	(ger inpat)	41–60	(median)
Joseph, 1954 (d)		M	<30	17		54.2±15.7	(active)
		M	<30	17		56.0±14.0	(active)
		M	30–44	17		51.6±17.7	(active
		M	30–44	17		51.6±16.1	(active)
		M	45+	16		46.3±16.0	(active)
		M	45+	16		44.1±22.8	(active)
		M	<30	17		23.7±11.5	(additional passive)
		M	<30	17		21.1±9.1	(additional passive)
		M	30–44	17		24.7±7.4	(additional passive)
		M	30–44	17		22.9±9.8	(additional passive)
		M	45+	16		24.6±16.0	(additional passive)
		M	45+	16		18.6±10.0	(additional passive)
Walker, 1984 (a)		M	60–84	30	72.4±6.8	62±17	
		F	60–84	30	72.7±8.2	59±8	

Table 6. Study findings on joint measurements—Continued

		Population			Age	Mean ± SD (degrees)	Range
Source and method	Side	Sex	Age	N	Mean and SD		
Great toe beginning flexion							
Walker, 1984 (a)		M	60–84	30	72.4 ± 6.8	3 ± 7	
		F	60–84	30	72.7 ± 8.2	1 ± 4	
Spine lateral flexion							
AAOS (a)			Four sources			28	(20–35)
Einkauf et al., 1987 (a)	R	F	20–29	>14		36.0 ± 5.7	
	L	F	20–29	>14		32.7 ± 7.1	
	R	F	30–39	>14		32.1 ± 6.2	
	L	F	30–39	>14		29.3 ± 5.8	
	R	F	40–49	>14		29.4 ± 5.4	
	L	F	40–49	>14		27.9 ± 6.9	
	R	F	50–59	>14		31.1 ± 5.8	
	L	F	50–59	>14		27.8 ± 5.1	
	R	F	60–69	>14		24.4 ± 7.5	
	L	F	60–69	>14		22.2 ± 6.2	
	R	F	70–84	>14		23.7 ± 3.5	
	L	F	70–84	>14		20.0 ± 4.4	
Spine anterior flexion							
Einkauf et al., 1987 (d)		F	20–29	>14		6.96 ± 1.21	
		F	30–39	>14		6.02 ± 1.07	
		F	40–49	>14		6.11 ± 1.41	
		F	50–59	>14		6.34 ± 1.44	
		F	60–69	>14		5.05 ± 1.01	
		F	70–84	>14		5.02 ± 1.02	
Spine extension							
AAOS (a)			Four sources			30	(20–30)
Einkauf et al., 1987 (a)		F	20–29	>14		36.3 ± 12.1	
		F	30–39	>14		28.3 ± 7.8	
		F	40–49	>14		21.4 ± 8.3	
		F	50–59	>14		21.6 ± 6.9	
		F	60–69	>14		19.4 ± 4.9	
		F	70–84	>14		17.7 ± 4.4	
Spine cervical rotation							
AAOS (a)			Four sources			38	(30–45)
Raab et al., in press (a)	R	F	63+	46	70.8 ± 5.3	69.7 ± 5.8	
	L	F	63+	46	70.8 ± 5.3	65.4 ± 8.4	
Sit and reach							
Buccola and Stone, (d)		M	60–79	20	67.6	8.0	(inches)
		M	60–79	16	65.4	7.3	(inches)
Frekany and Leslie, 1975 (d)		F	71–90	15		11.3	(inches)
Rikli and Busch, 1986 (d)		F	YAct	15	22.2	12.5 ± 9.3	(cm)
		F	YIn	15	21.1	3.4 ± 6.1	(cm)
		F	OAct	15	68.7	4.3 ± 2.6	(cm)
		F	OGolf	15	69.5	2.6 ± 4.5	(cm)
		F	OIn	15	68.9	−8.3 ± 10.9	(cm)

NOTE: Measurement methods are coded as follows:

(a) Goniometric measurements following AAOS guidelines.
(b) Other goniometric measurements.
(c) Leighton flexometer.
(d) Other.

Shoulder internal rotation decreased with age in the study of Boone and Azen (1979) but increased with age in men in the study of Murray (1985). Walker (1984) reported no difference between the groups 60–69 and 75–84 years. Shoulder external rotation was lower in the older age groups of Boone and Azen (1979) and Murray (1985) than in younger age groups. Subjects aged 60–84 in the study of Walker (1984) had external rotations 6 to 9 degrees lower than those of subjects 56–66 years in the Murray study. Shoulder internal plus external rotation did not differ between the age groups 26–36 and 56–66 years in the Murray study. Allander et al. (1974) reported a decline with age for shoulder internal plus external rotation. Differences in measurement techniques in the Allander study disallow comparison of values with other studies.

Elbow—Elbow flexion was lower in males 20–54 years than those 1–19 years (Boone and Azen, 1979) but did not differ among the older age groups studied by Smith and Walker (1983) and Walker (1984).

Wrist—Wrist flexion declined with age in the study of Boone and Azen (1979), but not in Walker (1984). Mean wrist flexion for the men aged 60–84 years in Walker was 12 degrees lower than that of men 20–54 years in Boone and Azen. Wrist extension did not differ with age within any of the studies. Wrist flexion plus extension decreased with age in the study of Allander et al. (1974).

Wrist radial and ulnar deviation did not differ with age within either study (Boone and Azen, 1979; Walker, 1984) reporting on different age ranges.

Thumb—Metacarpophalangeal I flexion plus extension did not decline with age in the study by Allander et al. (1974).

Hip—Hip flexion was not lower in older age groups in either of the two studies reporting on age. Hip extension was greater in subjects 20–54 years than those 1–19 years. Hip flexion plus extension, measured by Ahlback and Lindahl (1964), decreased with age.

Hip abduction and adduction both declined with age in the study of Boone and Azen (1979). Values for subjects 60–69 and 75–84 (Walker, 1984), although not differing by age group within the study, are 7–17 degrees lower than those for subjects 20–54 years in the Boone and Azen study.

Hip internal and external rotation was lower in subjects 20–54 years than those 1–19 years. Men aged 60–84 in the study of Walker (1984) had only one-half the internal rotation and three-quarters the external rotation of subjects 20–54 years in the study of Boone and Azen (1979). Hip internal plus external rotation declined with age in the study of Allander et al. (1974).

Knee—Knee flexion was slightly lower in males 20–54 years than in males 1–19 years (Boone and Azen, 1979). In women, but not men, studied by Smith and Walker (1983), knee flexion decreased with age in subjects 55–84 years. Walker (1984) found no difference between the age groups 60–69 and 75–84.

Ankle—Ankle plantar flexion was lower in the subjects aged 20–54 in the Boone and Azen (1979) study. Sepic et al. (1986) reported no significant difference between subjects 25–35 and 50–60, and Walker (1984) reported no differences between subjects 60–69 and 75–84. Mean ankle plantar flexion for male subjects in the Walker study was approximately 25 degrees lower than the values for subjects 20–54 years in the Boone and Azen study and those 50–60 years in the Sepic et al. study. The high value for ankle plantar flexion reported by Raab et al. (in press) may result from their measurement with the knee extended rather than at 90 degree flexion, as in other studies. Ankle dorsiflexion did not differ with age within any study.

Foot—Subtalar inversion and eversion did not differ with age between age groups 25–35 and 50–60 (Sepic et al., 1986) or between groups 60–69 and 75–84 years. Forefoot inversion and eversion were slightly lower in subjects aged 20–54 than in those aged 1–19 (Boone and Azen, 1979).

Toe—Active IP and MP joint flexion declined with age in the study of Joseph (1954), but extension was not significantly different among age groups. Neither MP joint flexion nor extension differed between the groups 60–69 and 75–84 years in the study of Walker (1984).

Spine—Spine lateral flexion, anterior flexion, and extension declined significantly with age in women aged 20–84 (Einkauf et al., 1987).

Measurement methods

Goniometer—A goniometer is an instrument that measures the angle between its two arms. The two arms are aligned with bony landmarks to measure range of motion. The American Academy of Orthopaedic Surgeons (1965) provided a handbook on measurement of joint motion using the goniometer. Motions of a joint are measured from zero starting positions. Three types of joints were described—those with 1-degree freedom of motion (hinge joints), such as the knee and elbow; those with 2-degree freedom of motion, with natural motion in two planes (e.g., flexion, extension, abduction, and adduction), as seen in the wrist; and ball-and-socket joints, whose motion is three dimensional, such as the shoulder and hips. The AAOS described positioning and measurement guidelines for all movements of the joints, and this reference is widely used in studies of flexibility. Average ranges of motion were compiled from four sources (The Committee on Medical Rating of Physical Impairment, 1958; The Committee of the California Medical Association and Industrial Accident Commission of the State of California, 1960; Clarke, 1920; and the Committee on Joint Motion, American Academy of Orthopaedic Surgeons). The AAOS stated that the subject's opposite extremity is perhaps the best normal standard when joint motion is limited. Mayerson and Milano (1984) evaluated the reliability of goniometric measurements of 22 separate joint positions. Both interobserver (r = .97) and intraobserver (r = .98) reliability coefficients were high. Agreement within 5 degrees was 64% between observers and 73% and 77% within observers. Based on 95% confidence intervals of average differences on all joints combined, the authors reported that repeated measures can be expected to fall within four angular degrees under controlled conditions.

Leighton flexometer—Leighton (1955) developed an instrument consisting of a weighted 360-degree dial and weighted pointer mounted in a case. This instrument, known as the Leighton flexometer, evaluates the total active range of motion in one plane. Measurements using the Leighton flexometer consider active ranges of motion as a whole, that is, flexion plus extension or abduction plus adduction. Leighton provided guidelines for measurement of the neck (flexion plus extension, left plus right rotation); shoulder (flexion plus extension, adduction plus abduction, internal plus external rotation); elbow (flexion plus extension); wrist (flexion plus extension, ulnar and radial flexion); hip-back (flexion plus extension; abduction plus adduction); knee (flexion plus extension); ankle (flexion plus extension, inversion plus eversion); trunk (flexion plus extension, lateral flexion). Leighton reported that test-retest readings on 120 boys were correlated at .913 to .996. The position from which measurements are taken differ in many cases from those recommended by the AAOS, and this may affect the range of motion.

Other measurements—The sit-and-reach test of the American Alliance for Health, Physical Education, Recreation and Dance (1980) has been used to evaluate trunk and hamstring flexibility. This test is quite general because components of ankle flexibility, hip flexibility, spine flexibility, and arm and leg length may all contribute to the results. Minor differences in procedure may make it risky to compare different studies. Standards for sit-and-reach flexibility for adults are not available at this time.

Recommendations

We recommend that flexibility be measured according to the guidelines of the AAOS, because this method is widespread and well defined. Joint motions should include shoulder forward flexion and extension, shoulder abduction, wrist flexion and extension, ankle plantar flexion and dorsiflexion, hip flexion, spine flexion, and the straight-leg raising test because these motions are important in the activities of daily living. If a subject has limited movement in one side because of injury or disease, both sides should be measured; otherwise measurement of the dominant side is sufficient. The presence of any pathological process affecting flexibility should be noted.

Flexibility measures

Table 6 details the mean values of flexibility reported by various investigators for the following joint motions:

Shoulder:
  Forward flexion/backward extension
  Horizontal flexion/horizontal extension
  Abduction/adduction
  Glenohumeral abduction
  Internal rotation/external rotationElbow:
  Flexion/extension
Forearm:
  Pronation/supination
Wrist:
  Flexion/extension
  Radial deviation/ulnar deviation
Thumb:
  Flexion/extension
Hip:
  Flexion/extension
  Abduction/adduction
  Internal rotation/external rotation
Knee:
  Flexion/extension
Ankle:
  Plantar flexion/dorsiflexion
Foot:
  Inversion/eversion
Toe:
  Flexion/extension
Spine:
  Right lateral flexion/left lateral flexion
  Anterior flexion/extension
  Left cervical rotation/right cervical rotation
Sit and reach

Bone

Introduction

Bone involution, which results in osteoporosis in the older adult, is a major public health problem in the United States. Osteoporosis contributes to at least 1.2 million fractures per year. The most common sites of fracture are the distal forearm, vertebrae, and neck of the femur. About 227,000 hip fractures occur each year, with 12–20% mortality and up to 50% of the survivors requiring long-term care (Riggs and Melton, 1986). The cost, direct and indirect, of fractures resulting from osteoporosis has been estimated at 6.1 billion dollars per year in the United States (Holbrook et al., 1984).

Osteoporosis is common in aging women, who start to lose bone at a rate of approximately 0.75–1% per year after about age 35. Cortical bone loss is accelerated to 2–3%/year for the first 4–5 years after menopause (Mazess, 1982; Lindsay et al., 1980), and some investigators have reported an accelerated loss of trabecular bone as well (Krolner and Nielsen, 1982; Meunier et al., 1973). The onset and rate of bone loss may vary with the skeletal site (Mazess, 1982; Riggs et al., 1981). Riggs and Melton (1986) stated that, over a lifetime, women lose 35% of their cortical bone and 50% of their trabecular bone. Bone loss in men begins at about age 40 at a rate of 0.4–0.5%/year.

Osteoporosis is a major problem in older women, with 33% of women over 65 having vertebral fractures (Riggs and Melton, 1986). On the other hand, most men do not incur osteoporotic fractures until they are over 80 years old. Cummings et al. (1985) reported that, by extreme old age, one out of every three women and one out of every six men will have fractured a hip.

Bone mineral content at maturity and at any point in time is affected by a variety of factors, such as genetics, lifetime calcium intake and absorption, hormonal homeostasis, and physical activity. Inadequate calcium intake or absorption (common after menopause) can cause significant reductions in bone mineral content.

The skeletal system is dynamic, modulated by the local and systemic control mechanisms that maintain mineral homeostasis. The two control mechanisms reflect the dual functions of the skeleton, i.e., mechanical support and mineral reservoir. Systemic control consists primarily of hormones regulating the absorption and excretion of calcium by acting on bone, kidney, and intestinal function. The major hormones involved are parathyroid hormone, calcitonin, vitamin D, growth hormone, and gonadotropins. Local control mechanisms are effected by the mechanical forces of gravity and muscle contraction, in which exercise is an important component.

Exercise, estrogen replacement therapy, and calcium supplements are currently the most widely suggested measures for deterring bone involution. Although exercise is advised, little is known as to the type of exercise best suited to reversing or preventing osteopenia.

Bone responds to a proportionate degree when the stimulant of either weight bearing or muscle contraction is changed, as in disuse or exercise. Because of this response, exercise is invaluable in reversing or preventing the age-related osteopenia leading to fractures in the hip, spine, femur, and wrist in millions of women. Little is known about the means by which bone is affected by mechanical stress. However, researchers suggest that local stress by strain-related electrical potential and production of skeletal growth factor stimulates bone response. Strain-related charges or currents are produced when the bone is bent, and external applications on the bone with these currents promote bone deposition (Singh and Saha, 1984). Because it is hypothesized that strain-related potential increases osteogenesis, it is also hypothesized that the degree of the strain-related potential is proportionate to the intensity of exercise.

Bone mineral content formation and removal are balanced in healthy young adults. However, in extreme conditions, such as prolonged bed rest, osteoblastic formation is less than osteoclastic removal, and the result is a net loss of bone. However, with increased physical activity or body weight, bone mass is increased as bone formation surpasses removal.

Although strength is a major parameter related to fracture resistance, it has been an aspect of research largely neglected as an effective intervention on bone strength. Presently there is not a comprehensive method of measuring the strength in vivo, and it is difficult to simulate normal loading patterns in vitro. Bone mineral content is not the only determinant of bone strength, although most studies involving exercise have focused solely on exercise's effect on bone mineral content. The correlation of bone strength to the mineral phase has been high, particularly in static strength loading. Additional contributing factors to bone strength include size and shape, porosity, osteon size, geometric configuration of calcified tissue, number of cement lines, and collagen location (Burr, 1980).

Weightlessness and immobilization

Humans

When compressive gravitational and muscular forces on the legs are deficient, as in bed rest, injury, or weightlessness, bone mass in humans is lost rapidly. Astronauts have lost as much as 2%/month in calcaneous bone mineral content (BMC) (Hattner and McMillan, 1968), as well as having a negative calcium balance. Also, an increase in urinary excretion of calcium hydroxyproline and nitrogen has been documented (Whedon et al., 1976). The alteration in calcium steady states (0.3–0.4%/month) with astronauts is similar to that in bed rest studies (Whedon et al., 1976).

When healthy subjects were placed at bed rest for periods of 4–36 weeks, they lost approximately 1%/week in the BMC of the os calcis (Hulley et al., 1971; Donaldson et al., 1970; Schneider and McDonald, 1984), iliac crest (Minaire et al., 1974), and lumbar spine (Krolner and Toft, 1983; Hansson et al., 1975). With bed rest, calcium excretion increases initially (Donaldson et al., 1970; Krolner and Toft, 1983; Whedon, 1984; Abramson and Delagi, 1961), reaches a plateau, and moderately declines, although it does not fall to prebed-rest levels. Similarly, calcium balance rapidly declines, followed by a leveling off at a highly negative value (Schneider and McDonald, 1984). Nitrogen (Whedon, 1984), hydroxyproline, and pyrophosphate excretion also increase (Donaldson et al., 1970).

Although studies of bone loss following bed rest and weightlessness prove to be excellent examples for sudden marked loss of stress, the models are not entirely applicable to the gradual deterioration in activity as seen in the aging healthy adult. In addition, other factors, such as physiological-metabolic parameters involved in bed rest and weightlessness, may bias results. However, the studies do substantiate the significance of muscle contraction and gravitational compression in preserving bone.

Animals

Disuse osteoporosis can be induced in animals by numerous methods, such as immobilization of bony segments, restraint, and denervation. Bone loss models may differ according to the type of animal studied or method of causing disuse. Generally, in regard to the area of bone immobilization, there is an increase in intracortical porosity and endosteal diameter and a decrease in cortical cross-sectional area, mass, and bone mineral content (Mosekilde et al., 1985; Young et al., 1983; Young and Schneider, 1981). These variations correlate with the pattern of age-related osteopenia. Uhthoff and associates (Uhthoff and Jaworski, 1978; Jaworski et al., 1980) documented a three-staged reaction to immobilization in the limbs of adult dogs: rapid loss, a restoration phase, and a moderate loss. In canines who were older in age, the loss was primarily endosteal, with evidence of intracortical porosity playing a larger role; in younger dogs, the loss was primarily periosteal. Conflicts regarding evidence of reversibility of imposed disuse osteoporosis have arisen, in that some researchers have documented almost complete rebuilding of bone but other researchers report little response to remobilization (Young et al., 1983; Young and Schneider, 1981).

Exercise and stress

In general, when age-matched normals have been compared with athletes (Dalen and Olsson, 1974; Nilsson et al., 1978; Brewer et al., 1983; Lane et al., 1986), it has been found that the athletes have greater bone density in the sites measured, with the quantity of hypertrophy correlated to the stress placed by the sports activity (Nilsson and Westlin, 1971). In activities that favor one arm over the other, like baseball or tennis, the bone hypertrophies in the playing arm. This has been documented in older adult (Montoye et al., 1980; Skrobak-Kaczynski and Andersen, 1974), adult (Buskirk et al., 1956; Jones et al., 1977), and young (Watson, 1974) men. However, in one study involving fencers, no variance was found in the bone density of the dominant and nondominant hands (Sosna, 1970). In a study of college-aged women, Jacobson et al. (1984) reported that both swimmers and tennis players had greater metatarsus and radius BMC than controls. In addition, the tennis players had a greater spine bone mineral density.

There are wide degrees of conflict in reports regarding cross-sectional studies relating physical activity and bone in nonathletes. Variance occurs for many reasons. Population characteristics differ, as do bone measurement sites and measurement methods. Some studies assessing physical activity through questionnaires find little correlation between activity and BMC of the limbs (Black-Sandler et al., 1982), the metacarpals (Montoye et al., 1976) and indexes of skeletal mass (Zanzi et al., 1981). In contrast, however, numerous studies have documented high correlations between BMC and physical activity. Sidney et al. (1977) and Oyster et al. (1984) reported correlations between metacarpal cortical diameter and physical activity. Stillman et al. (1986) showed a relation between assessed physical activity and BMC of the third distal radius. Jacobson et al. (1984) documented that less active postmenopausal women had a lower bone density in the lumbar spine than their more active postmenopausal counterparts. Several studies have examined fitness levels rather than physical activity alone. Work capacity was correlated significantly with lumbar spine BMC, but not with forearm BMC, in women who had a history of Colles' fractures (Krolner et al., 1982). Pocock et al. (1986) found a relationship between femoral neck and lumbar bone mineral density and fitness levels in 84 healthy postmenopausal women. Mashkara (1969) and Prives (1960) found that the structures and shapes of bone were directly related to occupation and that shape and structure were altered with changes in tenure. Other studies, although not specifically focused on exercise, support the hypothesis of the role of muscular stress and gravitational pull in maintaining bone. Many have reported that osteoporotic women weighed less (Saville and Nilsson, 1966), were weaker (Ellis et al., 1974), and were more sedentary (Krolner and Nielsen, 1982) than nonosteoporotics. Significant correlations have been documented between BMC and muscle strength in the spine (Sinaki et al., 1986; Pogrund et al., 1986) or skeleton (Harrison, 1984) and between grip strength and phalangeal bone mineral density, but it does appear that radial bone mineral content and grip strength have little correlation (Sinaki et al., 1974). Cross-sectional studies, as stated before, are significant because they aid in determining the variety of factors that subscribe to BMC, but they are confounded by variations in lifestyle and genetics so that they cannot fully clarify the relationship between skeletal changes and physical activity.

Exercise intervention

The crucial question is if, by increasing physical activity, age-related bone loss can be deterred. Several investigators have examined the effect of exercise intervention on bone loss. Dalen and Olsson (1974), in a study involving 19 male office workers, reported no increase in BMC with exercise, but the study was only 3 months in length. In contract, Margulies et al. (1986) documented tibial BMC increases in 268 young adult men following 14 weeks of intense physical activity. Because the training was so strenuous, however, stress fractures resulted for many of the participants. Chow et al. (1986) compared active healthy postmenopausal subjects ages 50–62 who exercised for 1 year with controls, and they reported greater bone mineral mass of the central third of the skeleton in the exercising group. In other studies involving postmenopausal women, variations were also found. Aloia et al. (1978) documented increases in total body calcium, as compared with a control group, after 1 year of exercise. Krolner et al. (1983) reported variations of change in the lumbar spine BMC when they compared subjects who exercised for 8 months and a control group. However, in both studies, BMC of the radius was not affected significantly by physical activity. In two other studies involving postmenopausal women (6 months and 4 years long), loss of radial BMC and BMC/W was deterred when exercise was used as an intervention modality (White et al., 1984; Smith et al., 1984). Similarly, Smith et al. (1984) showed that bone loss in both the humerus and ulna was reduced by exercise. Simkin et al. (1986) studied osteoporotic women (mean age 63) who participated in a 1-year exercise program emphasizing forearm use or served as controls. Bone density of the distal radius was increased by 3.8% in the exercising subjects while the control group decreased by 1.9% (Simkin et al., 1986). Exercise can deter bone mineral loss in elderly subjects also. Smith et al. (1981) documented nonsignificant increases in radius BMC and BMC/W in elderly women following 3 years of exercise; the control group BMC and BMC/W deteriorated, and the difference between the groups was significant. Likewise, women who trained for 9 months did not improve significantly the BMC of the heel; however, when compared with a control group, the BMC was substantially higher (Rundgren et al., 1984).

Animal studies

The bone mineral density is related to the amount of weight on an animal's limbs. Body mass is related to bone mineral density. It has been shown in experiments that bone hypertrophied when more weight was placed on bone by either surgical means or lead weights (Goodship et al., 1979; Martin et al., 1981 and 1982; Lanyon et al., 1982). A variety of exercise programs increased bone density, cortical thickness, and/or cross-sectional area in pigs, dogs, and small rodents. Although the limbs of the animals immediately affected by the exercise program hypertrophy the most, bone density of the axial portions may also increase (McDonald et al., 1986; Burr et al., 1983). Several studies experimenting on rats and mice report that, with older animals, bone response to stress was enhanced (McDonald et al., 1986; Ringe and Steinhagen-Thiessen, 1985; McDonald et al., 1985). For instance, in young rats, increases in mass occurred in only the cortical areas directly affected, but in 19-month-old rats the ribs and skull increased as well.

To explore bone's response to highly regulated levels and types of stress, a few researchers have examined the results of static and dynamic artifical loading. Dynamic loading is more effective than constant compression (Perren et al., 1969; Lanyon and Rubin, 1984), but static loading, when added to the already existing forces placed on the limbs by walking, may enhance bone mineral (Meade et al., 1984). With dynamic loading, bone hypertrophy has been linked with peak strain (Rubin and Lanyon, 1985). In research conducted by Rubin and Lanyon (1984), peak strain was found to be a primary determinant of bone response; in a turkey model, 36 and 1,800 cycles/day of the strain proved to be equally effective. Cross-sectional areas of bone increase with dynamic loading, yet there are variations among studies relative to endosteal and periosteal formation because of differences in research designs and loads. Investigators have noted periosteal and endosteal remodeling simultaneously (Burr et al., 1983), primarily periosteal remodeling (Lanyon and Rubin, 1984), solely endosteal remodeling (Woo et al., 1981), or intracortical and periosteal remodeling (Churches and Howlett, 1986).

Measurement methods

A number of methods are available for noninvasive measurements of skeletal bone mass. These include single- and dual-energy quantitative computed tomography (CT), single photon absorptiometry (SPA), dual photon absorptiometry (DPA), and x rays of the lateral lumbar spine or hands. Although x rays are readily available, they are relatively imprecise and cannot detect changes in bone mineral less than 15–20%. CT, SPA, and DPA are more precise and have been shown to be reliable and valid. SPA can be used for measurements of the radius, ulna, and humerus, and DPA for the spine, hip, and tibia. Currently, CT is used only in measurements of spine. CT and DPA measurements of the spine are well correlated (Reinbold et al., 1986).

The radiation exposures from SPA and DPA (5–20 mrem) are much less than from CT (300–1,000 mrem). For repeated measures or general screening programs, therefore, SPA and DPA are preferable.

Correlations of bone mineral content between sites such as the radius and spine or hip and spine are only moderate, and therefore measurement of one site cannot be used reliably to predict bone mineral content at another site in the same individual.

Recommendations for bone mineral analysis

The most common sites for osteoporosis-related fractures are the wrist, hip, and spine. Because it appears that the rate of decline in bone mineral mass is not constant throughout the skeleton, we recommend scanning each site—by single-photon absorptiometry for the radius and ulna and by dual-photon absorptiometry for the spine (L2–L4) and femur. Average values adjusted for weight, height, age, and sex are generally provided within the programs for bone scanning.

References

Abramson AS, EF Delagi. Influence of weight-bearing and muscle contraction on disuse osteoporosis. *Arch Phys Med Rehabil* 42:147–151, 1961.

Adams GM, HA deVries. Physiological effects of an exercise training regimen upon women aged 52 to 79. *J Gerontol* 28(1):50–55, 1973.

Adrian MJ. Flexibility in the aging adult. In *Exercise and Aging: The Scientific Basis,* Smith EL, RC Serfass (eds.). Hillside, NJ: Enslow, 1981, pp. 45–57.

Ahlback SO, O Lindahl. Sagittal mobility of the hip joint. *Acta Orthop Scand* 34:310–322, 1964.

Allander E, OJ Bjornsson, O Olafsson, N Sigfusson, and J Thorsteinsson. Normal range of joint movements in shoulder, hip, wrist and thumb with special reference to side: A comparison between two populations. *Int J Epid* 3(3):253–261, 1974.

Aloia JF, SH Cohn, J Ostuni, R Cane, and K Ellis. Prevention of involutional bone loss by exercise. *Ann Int Med* 89:356–358, 1978.

American Academy of Orthopaedic Surgeons. *Joint Motion. Method of Measuring and Recording.* Chicago: American Academy Orthopaed. Surg., 1965.

American Alliance for Health, Physical Education, Recreation and Dance. *AAHPERD Health Related Fitness Manual.* Reston, VA:, 1980.

Andersen KL, L Hermansen. Aerobic work capacity in middle-aged Norwegian men. *J Appl Physiol* 20(3):432–436, 1965.

Asano K, S Ogawa, and Y Furuta. Aerobic work capacity in middle and old-aged runners. In *Proceedings, International Congress of Physical Activity Sciences.* Quebec City, Canada, 1976.

Asmussen E, K Fruensgaard, and S Norgaard. A follow-up longitudinal study of selected physiologic functions in former physical education students—after forty years. *J Am Geriat Soc* 23(10):442–450, 1975.

Astrand I. The physical work capacity of workers 50–64 years old. *Acta Physiol Scand* 42:73–86, 1958.

Astrand I. Aerobic work capacity in men and women with special reference to age. *Acta Physiol Scand* Supplement 169 49:1–92, 1960.

Astrand I. ST depression, heart rate, and blood pressure during arm and leg work. *Scand J Lab Clin Invest* 30:411–414, 1972.

Astrand I, P-O Astrand, I Hallback, and A Kilbom. Reduction in maximal oxygen uptake with age. *J Appl Physiol* 35(5):649–654, 1973.

Astrand P-O. *Experimental Studies of Physical Working Capacity in Relation to Sex and Age.* Copenhagen: Munksgaard, 1952.

Astrand P-O, K Rodahl. *Textbook of Work Physiology.* New York: McGraw Hill, 1970.

Atomi Y, M Miyashita. Maximal aerobic power of Japanese active and sedentary adult females of different ages (20 to 62 years). *Med Sci Sports* 6(4):223–225, 1974.

Atomi Y, M Miyashita. Effect of training intensity in adult females on aerobic power, related to lean body mass. *Eur J Appl Physiol* 44:109–116, 1980.

Badenhop DT, PA Cleary, SF Schaal, EL Fox, and RL Bartels. Physiological adjustments to higher- or lower-intensity exercise in elders. *Med Sci Sports Exer* 15(6):496–502, 1983.

Bailey DA, RJ Shephard, RL Mirwald, and GA McBride. A current view of cardiorespiratory fitness levels of Canadians. *Canad Med Assoc J* 111:25–30, 1974.

Barry AJ, JW Daly, EDR Pruett, JR Steinmetz, HF Page, NC Birkhead, and K Rodahl. The effects of physical conditioning on older individuals. I. Work capacity, circulatory-respiratory functions and work electrocardiogram. *J Gerontol* 21:182–191, 1966.

Barry AJ, GW Webster, and JW Daly. Validity and reliability of a multistage exercise test for older men and women. *J Gerontol* 24:284–291, 1969.

Beall CM, MC Goldstein, and ES Feldman. Social structure and intracohort variation in physical fitness among elderly males in a traditional third world society. *J Am Geriatr Soc* 33:406, 1985a.

Beall CM, MC Goldstein, and ES Feldman. The physical fitness of elderly Nepalese farmers residing in rugged mountain and flat terrain. *J Gerontol* 40(5):529–535, 1985b.

Benestad AM. Trainability of old men. *Acta Med Scand* 178:321–327, 1965.

Benestad AM, J Havorsrud, and KL Andersen. The physical fitness of old Norwegian men and women. *Acta Med Scand* 183:73, 1968.

Berry MA, WG Squires, and AS Jackson. Fitness variables and the lipid profile in United States astronauts. *Aviat Space Environ Med* 51(11):1222–1226, 1980.

Black-Sandler R, RE LaPort, D Sashin, LH Kuller, E Sternglass, JA Cauley, and MM Link. Determinants of bone mass in menopause. *Preventive Medicine* 11:269–280, 1982.

Boone DC, SP Azen. Normal range of motion of joints in male subjects. *J Bone Joint Surg* 61A(5):756–759, 1979.

Bortz WM. Disuse and aging. *JAMA* 248(10):459–468, 1982.

Bottiger LE. Regular decline in physical working capacity with age. *Br Med J* 4:270–271, Aug. 1973.

Brand RJ, RS Paffenbarger, RI Sholtz, and JB Kampert. Work activity and fatal heart attack studied by multiple logistic risk analysis. *Am J Epidemiol* 110(1):52–62, 1979.

Brewer V, BM Meyer, MS Keele, SJ Upton, and RD Hagan. Role of exercise in prevention of involutional bone loss. *Med Sci Spts Exer* 15:445–449, 1983.

Brown JR, RJ Shephard. Some measurements of fitness in older female employees of a Toronto department store. *Can Med Assoc J* 97:1208–1213, 1967.

Bruce RA. Methods of exercise testing: Step test, bicycle, treadmill, isometrics. In *Exercise in Cardiovascular Health and Disease,* Amsterdam EA, JH Wilmore, and AN DeMaria (eds.). New York: Yorke Medical Books, 1977, pp. 150–160.

Bruce RA. Primary intervention against coronary atherosclerosis by exercise conditioning? *N Engl J Med* 305(25):1525–1526, 1981.

Bruce RA. Normal values for VO2 and the VO2-HR relationship. *Am Rev Respir Dis* 129:S41–S43, 1984.

Buccola VA, WJ Stone. Effects of jogging and cycling programs on physiological and personality variables in aged men. *Res Quart* 46(2):134–139, 1975.

Burr DB. The relationships among physical, geometrical and mechanical properties of bone, with a note on the properties of nonhuman primate bone. *Yearbook Phys Athropol* 23:109–146, 1980.

Burr DB, B Martin, and PA Martin. Lower extremity loads stimulate bone formation in the vertebral column: Implications for osteoporosis. *Spine* 8(7):681–686, 1983.

Buskirk ER, KL Anderson, and J Brozek. Unilateral activity and bone and muscular development in the forearm. *Res Q* 27:127–131, 1956.

Buskirk ER, JL Hodgson. Age and aerobic power: The rate of change in men and women. *Fed Proc* 46:1824–1829, 1987.

Chapman EA, HA deVries, and R Swezey. Joint stiffness: Effects of exercise on young and old men. *J Gerontol* 27(2):218–221, 1972.

Chiang BN, HJ Montoye, and DA Cunningham. Treadmill exercise study of healthy males in a total community—Tecumseh, Michigan: Clinical and electrocardiographic characteristics. *Am J Epidemiol* 91:368–377, 1970.

Chow RK, JE Harrison, CF Brown, and V Hajek. Physical fitness effect on bone mass in postmenopausal women. *Arch Phys Med Rehabil* 67:231–234, 1986.

Churches AE, CR Howlett. Functional adaptation of bone in response to sinusoidally varying controlled compressive loading of the ovine metacarpus. *Clin Orthop Rel Res* 168:265–280, 1986.

Clarke WA. A system of joint measurements. *J Orthop Surg* 2(12):687–700, 1920.

Committee of California Medical Association and the Industrial Accident Committee of the State of California: *Evaluation of Industrial Disability.* Oxford University Press, 1960.

Crow RS, PM Rautaharju, RJ Prineas, JE Connett, C Furberg, S Broste, and J Stamler. Risk factors, exercise fitness and electrocardiographic response to exercise in 12,866 men at high risk of symptomatic coronary heart disease. *Am J Cardiol* 57:1075–1082, 1986.

Cumming GR, LM Borysyk. Criteria for maximum oxygen intake in men over 40 in a population survey. *Med Sci Sports* 4:18–22, 1972.

Cumming GR, C Defresne, L Kich, and J Samm. Exercise electrocardiogram patterns in normal women. *Br Heart J* 35:1055, 1973.

Cummings SR, JL Kelsey, MC Nevitt, and KJ O'Dowd. Epidemiology of osteoporosis and osteoporotic fractures. *Epidemiol Rev* 7:178–208, 1985.

Cunningham DA, JS Hill. Effect of training on cardiovascular response to exercise in women. *J Appl Physiol* 39(6):891–895, 1975.

Cunningham DA, PA Rechnitzer, JH Howard, and AP Donner. Exercise training of men at retirement: A clinical trial. *J Gerontol* 42(1):17–23, 1987.

Dalen N, KE Olsson. Bone mineral content and physical activity. *Acta Orthop Scand* 45:170–174, 1974.

De Backer G, M Kornitzer, J Sobolski, M Dramaix, S Degre, M de Marneffe, and H Denolin. Physical activity and physical fitness levels of Belgian males aged 40–55. *Cardiology* 67:110–128, 1981.

Dehn MM, RA Bruce. Longitudinal variations in maximal oxygen intake with age and activity. *J Appl Physiol* 33(6):805–807, 1972.

deVries HA. Physiological effects of an exercise training regimen upon men aged 52 to 88. *J Gerontol* 25:325–336, 1970.

deVries HA. Exercise intensity threshold for improvement of cardiovascular-respiratory function in older men. *Geriatrics* 26:94–101, 1971.

deVries HA, GM Adams. Comparison of exercise responses in old and young men: I. The cardiac effort/total body effort relationship. *J Gerontol* 27(3):344–348, 1972a.

deVries HA, GM Adams. Comparison of exercise responses in old and young men: II. Ventilatory mechanics. *J Gerontol* 27(3):349–352, 1972b.

Dill DB, S Robinson, and JC Ross. A longitudinal study of 16 champion runners. *J Sports Med Phys Fitness* 4:4–27, 1967.

Donaldson CI, SB Hulley, JM Vogel, RS Hattner, JH Bayers, and E McMilan. Effect of prolonged bed rest on bone mineral. *Metabolism* 19:1071–1084, 1970.

Drinkwater BL. Physiological responses of women to exercise. In *Exercise and Sports Sciences Reviews* 1, Wilmore JH (ed.). New York: Academic Press, 1973, pp. 125–152.

Drinkwater BL, SM Horvath, and CL Wells. Aerobic power of females, ages 10 to 68. *J Gerontol* 30(4):385–394, 1975.

Durnin JVGA, V Mikulicic. The influence of graded exercises on the oxygen consumption, pulmonary ventilation and heart rate of young and elderly men. *Quart J Exp Physiol* 41:442–452, 1956.

Einkauf DK, ML Gohdes, GM Jensen, and MJ Jewewll. Changes in spinal mobility with increasing age in women. *Phys Ther* 67(3):370–375, 1987.

Ellis KJ, KK Shukla, SH Cohn, and RN Pierson. A predictor for total body potassium in men based on height, weight, sex and age: Applications in metabolic disorders. *J Lab Clin Med* 83:716–727, 1974.

Epstein Y, G Keren, R Udassin, and Y Shapiro. Way of life as a determinant of physical fitness. *Eur J Appl Physiol* 47:1–5, 1981.

Ericsson P, L Irnell. Effect of five years' ageing on ventilatory capacity and physical work capacity in elderly people. *Acta Med Scand* 185:193–199, 1969.

Erikssen J, J Jervell, and K Forfang. Blood pressure responses to bicycle exercise testing in apparently healthy middle-aged men. *Cardiology* 66:56–63, 1980.

Espenschade AS. Role of exercise in the well-being of women 35–80 years of age. *J Gerontol* 24:86–89, 1969.

Evans JG, IAM Prior, and WMG Tunbridge. Age-associated change in QRS Axis: Intrinsic or extrinsic ageing? *Gerontology* 28:132–137, 1982.

Fairshter RD, J Walters, K Salness, M Fox, V-D Minh, and AF Wilson. A comparison of incremental exercise tests during cycle and treadmill ergometry. *Med Sci Sports Exer* 15:549–554, 1983.

Fardy PS, CM Maresh, R Abbott, and T Kristiansen. An assessment of the influence of habitual physical activity, prior sport participation, smoking habits, and aging upon indices of cardiovascular fitness: Preliminary report of a cross-sectional and retrospective study. *J Sports Med Phys Fit Quart Rev* 16:77–90, 1976.

Fischer, A, J Parizkova, and Z Roth. The effect of systematic physical activity on maximal performance and functional capacity in senescent men. *Int Z Angew Physiol Einschl Arbeitsphysiol* 21:269–304, 1965.

Flint HM, BL Drinkwater, CL Wells, and SM Horvath. Validity of estimating body fat of females: Effect of age and fitness. *Human Biol* 49(4):559–572, 1977.

Flint MM, BL Drinkwater, and SM Horvath. Effects of training on women's response to submaximal exercise. *Med Sci Sports* 6(2):89–94, 1974.

Foster VL, GJE Hume, AL Dickinson, SJ Chatfield, and WC Byrnes. The reproducibility of VO2max, ventilatory, and lactate thresholds in elderly women. *Med Sci Sports Exer* 18(4):425–430, 1986.

Franks BD, HK Hellerstein, C Yoran, PS Fardym, and A Bram. Ten-year follow-up of coronary heart disease exercise program: Effects of different amounts of physical activity on cardiovascular function. *J Cardiac Rehab* 3:282–287, 1983.

Frekany GA, DK Leslie. Effects of an exercise program on selected flexibility measurements of senior citizens. *The Gerontologist* 4:182–183, 1975.

Froelicher VF, M Allen, and MC Lancaster. Maximal treadmill testing of normal USAF aircrewman. *Aerospace Med* 45:310–315, 1974.

Germain NW, SN Blair. Variability of shoulder flexion with age, activity and sex. *Amer Corr Ther J* 37(6):156–160, 1983.

Goodship AE, LE Lanyon, and H McFie. Functional adaptation of bone in increased stress. *J Bone Joint Surg* 61A:539, 1979.

Gordon TI, EW Banister, and BP Gordon. The caloric cost of competitive figure skating. *J Sports Med* 9:98–103, 1969.

Gozna ER, AE Marble, A Shaw, and JG Holland. Age-related changes in the mechanics of the aorta and pulmonary artery of man. *J Appl Physiol* 36(4):407–411, 1974.

Grimby G, N Nilsson, and B Saltin. Cardiac output during submaximal and maximal exercise in active middle-aged athletes. *J Appl Physiol* 21:1150, 1966.

Grimby G, B Saltin. Physiological analysis of physically well-trained middle-aged and old athletes. *Acta Med Scand* 179(5):513–526, 1966.

Grimby G, L Wilhelmsen, B Ekstrom-Jodal, M Aurell, J Bjure, and G Tibblin. Aerobic power and related factors in a population study of men aged 54. *Scand J Clin Lab Invest* 26:287–294, 1970.

Gutman GM, CP Herbert, and SR Brown. Feldenkrais versus conventional exercises for the elderly. *J Gerontol* 32(5):562–572, 1977.

Gyntelberg F, K Ohlsen. Physical fitness and serum cholesterol in Copenhagen males aged 40–59. *Scand J Clin Lab Invest* 32:211–216, 1973.

Hagan RD, LR Gettman. Maximal aerobic power, body fat, and serum lipoproteins in male distance runners. *J Cardiac Rehabil* 3:331–337, 1983.

Hageman PA, DJ Blanke. Comparison of gait of young women and elderly women. *Phys Ther* 66(9):1382–1387, 1986.

Hansson TH, BO Roos, and A Nachemson. Development of osteopenia in the fourth lumbar vertebra during prolonged bed rest after operation for scoliosis. *Acta Orthop Scand* 46:621–630, 1975.

Hanson JS, BS Tabakin, AM Levy, and W Nedde. Long-term physical training and cardiovascular dynamics in middle-aged men. *Circulation* 38:783, 1968.

Harris R. Cardiac Changes with Age. In *The Physiology and Pathology of Human Aging,* R Goldman and M Rockstein (ed.). New York: Academic Press, 1975, pp. 109–122.

Harrison JE. Neutron activation studies and the effect of exercise on osteoporosis. *J Med* 15(4):285–294, 1984.

Hartung GH. Responses of middle-aged women to maximal cycling exercise. *Amer Corr Ther J* 28(4):103–106, 1974.

Hattner RS, DE McMillan. Influence of weightlessness upon the skeleton: A review. *Aerospace Med* 39:849–855, 1968.

Havel V, A Tidjditi, J Hert, O Skranc, and K Bartak. La capacite physique des etudiantes en medecine algeriennes et tcheques. *Int Z Angew Physiol* 27:292–298, 1969.

Heath GW, JM Hagberg, AA Ehsani, and JO Holloszy. A physiological comparison of young and older endurance athletes. *J Appl Physiol: Respirat Environ Exer Physiol* 51(3):634–640, 1981.

Hermansen L, KL Andersen. Aerobic work capacity in young Norwegian men and women. *J Appl Physiol* 20:425–431, 1965.

Hodgson JL. *Age and aerobic capacity of urban midwestern males.* Ph.D. Dissertation, University of Minnesota, 1971.

Holbrook TL, K Grazier, JL Kelsey, and RN Stauffer. *The Frequency of Occurrence, Impact and Cost of Selected Musculoskeletal Conditions in the United States.* Chicago: Am. Academy Orthop. Surgeons, 1984.

Hollmann W, HW Knipping. Die Bestimmung der menschlichen Leistungfsfahigkeit. In *Health and Fitness in the Modern World,* Larson LA (ed.). Madison, WI: US Athletic Institute, 1961.

Hossack KF, RA Bruce. Maximal cardiac function in sedentary normal men and women: Comparison of age-related changes. *J Appl Physiol* 53(4):799–804, 1982.

Hossack KF, RA Bruce, B Green, F Kusumi, TA DeRouen, and S Trimble. Maximal cardiac output during upright exercise: Approximate normal standards and variations with coronary heart disease. *Am J Cardiol* 46:204–212, 1980.

Hossack KF, F Kusumi, and RA Bruce. Approximate normal standards of maximal cardiac output during upright exercise in women. *Am J Cardiol* 47:1080–1086, 1981.

Hulley SB, JM Vogel, CI Donaldson, JH Bayers, RJ Friedman, and SN Rosen. The effect of supplemental oral phosphate on the bone mineral changes during prolonged bed rest. *J Clin Invest* 50:2506–2518, 1971.

Hung LK, YF Ho, and PC Leung. Survey of foot deformities among 166 geriatric inpatients. *Foot and Ankle* 5(4):156–164, 1985.

Hurley BF, JM Hagberg, DR Seals, AA Ehsani, AP Goldberg, and JO Holloszy. Glucose tolerance and lipid-lipoprotein levels in middle-aged powerlifters. *Clin Physiol* 7:11–19, 1987.

Ikai M, K Kitagawa. Maximal oxygen uptake of Japanese related to sex and age. *Med Sci Sports* 4:127–131, 1972.

Ismail AH, DL Montgomery. The effect of a four-month physical fitness program on a young and an old group matched for physical fitness. *Europ J Appl Physiol* 40:137–144, 1979.

Jacobson PC, W Beaver, SA Grubb, TN Taft, and RV Talmage. Bone density in women: College athletes and older athletic women. *J Orthop Res* 2(4):328–332, 1984.

Jaworski ZFG, M Liskova-Kiar, and HK Uhthoff. Effect of long-term immobilisation on the pattern of bone loss in older dogs. *J Bone Joint Surg* 62B(1):104–110, 1980.

Johns RJ, U Wright. Relative importance of various tissues in joint stiffness. *J Appl Physiol* 17:824–828, 1962.

Jones HH, JD Priest, and WC Hayes. Humeral hypertrophy in respnse to exercise. *J Bone Joint Surg* 59A(2):204–208, 1977.

Joseph J. Range of movement of the great toe in men. *J Bone Joint Surg [Br]* 36:450–456, 1954.

Julius S, A Amery, LS Whitlock, and J Conway. Influence of age on the hemodynamic response to exercise. *Circulation* 36:222–230, 1967.

Kamon E, KB Pandolf. Maximal aerobic power during laddermill climbing, uphill running and cycling. *J Appl Physiol* 32:467–473, 1972.

Kannel WB, P Wilson, and SN Blair. Epidemiological assessment of the role of physical activity and fitness in development of cardiovascular disease. *Am Heart J* 109(4):876–885, 1985.

Kasch FW. The effects of exercise on the aging process. *Phys Sportsmed* 4(6):64–68, 1976.

Kasch FW, WH Phillips, JEL Carter, and JL Boyer. Cardiovascular changes in middle-aged men during two years of training. *J Appl Physiol* 34(1):53–57, 1973.

Kasch FW, W Phillips, J Carter, W Ross, and J Boyer. Maximum work capacity in middle-aged males by a step-test method. *J Sports Med* 5:198, 1965.

Kasch FW, JP Wallace, and SP Van Camp. Effects of 18 years of endurance exercise on the physical work capacity of older men. *J Cardiopulmonary Rehabil* 5:308–312, 1985.

Katch VL, SS Sady, and P Freedson. Biological variability in maximum aerobic power. *Med Sci Sports Exer* 14(1):21–25, 1982.

Kavanaugh T, RJ Shephard. The effects of continued training on the aging process. *Ann NY Acad Sci* 301:656–670, 1977.

Kellgren JH, JS Lawrence. Radiological assessment of osteoarthrosis. *Annals Rheum Dis* 16:494–502, 1957.

Kiens B, I Jorgensen, S Lewis, G Jensen, H Lithell, B Vessby, S Hoe, and P Schnohr. Increased plasma HDL-cholesterol and apo A-1 in sedentary middle-aged men after physical conditioning. *Eur J Clin Invest* 10:203–209, 1980.

Kiessling KH, L Pilstrom, AC Bylund, B Saltin, and K Piehl. Enzyme activities and morphometry in skeletal muscle of middle-aged men after training. *Scand J Clin Lab Invest* 33:63–69, 1974.

Kilbom A. Physical training in women. *Scand J Clin Lab Invest* 28(119):S1–S34, 1971.

Klissouras V, F Pirnay, F Petit, and JM Petit. Adaptation to maximal effort: Genetics and age. *J Appl Physiol* 35:288–293, 1973.

Kostis JB, AE Moreyra, MT Amendo, JD Pietro, N Cosgrove, and PT Kuo. The effect of age on heart rate in subjects free of heart disease. *Circulation* 65:141–145, 1982.

Kramsch DM, AJ Aspen, BM Abramowitz, T Kreimendahl, & WB Hood. Reduction of coronary atherosclerosis by moderate conditioning and exercise in monkeys on an atherogenic diet. *N Engl J Med* 305(25):1483–1489, 1981.

Krolner B, SP Nielsen. Bone mineral content of the lumbar spine in normal and osteoporotic women: Cross-sectional and longitudinal studies. *Clin Sci* 62:329–336, 1982.

Krolner B, B Toft. Vertebral bone loss: An unheeded side effect of therapeutic bed rest. *Clin Sci* 64:537–540, 1983.

Krolner B, B Toft, SP Nielson, and E Tondevold. Physical exercise as prophylaxis against involutional vertebral bone loss: A controlled trial. *Clin Sci* 64:541–546, 1983.

Krolner B, E Tondevold, B Toft, B Berthelson, and SP Nielsen. Bone mass of the axial and the appendicular skeleton in women with Colles' fracture: Its relation to physical activity. *Clin Physiol* 2:147–157, 1982.

Lakatta EG. Hemodynamic adaptations to stress with advancing age. In *Physical Activity in Health and Disease,* Astrand P-O, G Grimby (eds.). Uppsala, Sweden: Almqvist & Wiksell Tryckeri, 1986, pp. 39–52.

Lane NE, DA Bloch, HH Jones, WH Marshall, PD Wood, and JF Fries. Long-distance running, bone density, and osteoarthritis. *JAMA* 255(9):1147–1151, 1986.

Lanyon LE, AE Goodship, CJ Pye, and H MacFie. Mechanically adaptive bone remodelling. A quantitative study on functional adaptation in the radius following ulna osteotomy in sheep. *J Biomechanics* 15:141–154, 1982.

Lanyon LE, CT Rubin. Static vs. dynamic loads as an influence on bone remodelling. *J Biomechanics* 17(12):897–905, 1984.

Leighton JR. An instrument and technic for the measurement of range of joint motion. *Arch Phys Med Rehabil* 571–578, 1955.

Lesser M. The effects of rhythmic exercise on the range of motion in older adults. *Amer Corr Ther J* 32(4):118–122, 1978.

Lester FM, LT Sheffield, and TJ Reeves. Electrocardiographic changes in clinically normal older men following near maximal and maximal exercise. *Circulation* 36:5–14, 1967.

Lester M, LT Sheffield, P Trammell, and TJ Reeves. The effect of age and athletic training on the maximal heart rate during muscular exercise. *Am Heart J* 76(3):370–376, 1968.

Lie H, R Mundal, and J Erikssen. Coronary risk factors and incidence of coronary death in relation to physical fitness. Seven-year follow-up study of middle-aged and elderly men. *Eur Heart J* 6:147–157, 1985.

Lindsay R, DM Hart, C Forrest, and C Baird. Prevention of spinal osteoporosis in oophorectomised women. *Lancet* 2:1151–1153, 1980.

Luchi RJ, E Snow, JM Luchi, CL Nelson, and FJ Pircher. Left ventricular function in hospitalized geriatric patients. *J Am Geriatr Soc* 30(11):700–705, 1982.

MacNab RB, RP Conger, and PS Taylor. Differences in maximal and submaximal work capacity in men and women. *J Appl Physiol* 47:644–648, 1969.

Makrides L, GJF Heigenhauser, N McCartney, and NL Jones. Maximal short term exercise capacity in healthy subjects aged 15–70 years. *Clin Sci* 69:197–205, 1985.

Maksud MG, RL Wiley, LH Hamilton, and B Lockhart. Maximal VO2, ventilation, and heart rate of Olympic speed skating candidates. *J Appl Physiol* 29:186–190, 1970.

Mann GV, HL Garrett, A Rahri, H Murrah, and FT Billings. Exercise to prevent coronary heart disease. An experimental study of the effects of training on risk factors for coronary disease in men. *Amer J Med* 46:12–27, 1969.

Margulies JY, A Simkin, I Leichter, A Bivas, R Steinberg, M Giladi, M Stein, et al. Effect of intense physical activity on the bone-mineral content in the lower limbs of young adults. *J Bone Joint Surg* 68A(7):1090–1093, 1986.

Martin RK, JP Albright, and HK Huang. Vertebral body change due to exercise as assessed by computerized tomography. *Trans Orthop Res Soc* 7:104, 1982.

Martin RK, JP Albright, WSS Jee, GN Taylor, and WR Clarke. Bone loss in the beagle tibia: Influence of age, weight and sex. *Calcif Tissue Int* 33:233–238, 1981.

Mashkara K. Effect of physical labor on the structure of the bones of the upper extremities. *Arkansas Anat Gistol Embryol* 56:7–15, 1969.

Massicotte DR, G Avon, and G Corriveau. Comparative effects of aerobic training on men and women. *J Sports Med* 19:23–32, 1979.

Mayerson NH, RA Milano. Goniometric measurement reliability in physical medicine. *Arch Phys Med Rehabil* 65:92–94, 1984.

Mazess RB. On aging bone loss. *Clin Orthop Rel Res* 165:239–252, 1982.

McArdle WD, JR Magel, and LC Kyvallos. Aerobic capacity, heart rate, and estimated energy cost during women's competitive basketball. *Res Quart* 42:178–186, 1971.

McDonald R, J Hegenauer, and P Saltman. Calcium metabolism and bone mineralization in aged rats following exercise. *The Gerontologist* 25:56, 1985.

McDonald R, J Hegenauer, and P Saltman. Age-related differences in the bone mineralization pattern of rats following exercise. *J Gerontol* 41(4):445–452, 1986.

McDonough JR, F Kusumi, and RA Bruce. Variations in maximal oxygen intake with physical activity in middle-aged men. *Circulation* 41:743–751, 1970.

Meade JB, SC Cowin, JJ Klawitter, WC Buskirk, and HB Skinner. Bone remodeling due to continuously applied loads. *Calcif Tissue Int* 36:S25–S30, 1984.

Metheny E, L Brouha, RE Johnson, and WH Forbes. Some physiologic responses of women and men to moderate and strenuous exercise: A comparative study. *Am J Physiol* 137:318, 1942.

Meunier P, P Courpron, C Edouard, J Bernard, J Bringuier, and C Vignon. Physiological senile involution and pathological rarefaction of bone quantitative and comparative histological data. *Clin Endocrinol Metab* 2:239–256, 1973.

Michael ED, SM Horvath. Physical working capacity of college women. *J Appl Physiol* 20:263–266, 1965.

Minaire P, P Menier, C Edouard, J Bernard, P Courpron, and J Bourret. Quantitative histological data on disuse osteoporosis. *Calcif Tiss Res* 17:57–73, 1974.

Montoye HJ. Age and oxygen utilization during submaximal treadmill exercise in males. *J Gerontol* 37(4):396–402, 1982.

Montoye HJ, JF McCabe, HL Metzner, and SM Garn. Physical activity and bone density. *Hum Biol* 48:599–610, 1976.

Montoye HJ, EL Smith, DF Fardon, and ET Howley. Bone mineral in senior tennis players. *Scand J Sports Sci* 2(1):26–32, 1980.

Montoye HJ, WD Van Huss, H Olson, A Hudec, and E Mahoney. Study of the longevity and morbidity of college athletes. *JAMA* 162:1132–1134, 1956.

Mosekilde L, A Viidik, and L Mosekilde. Correlation between the compressive strength of iliac and vertebral trabecular bone in normal individuals. *Bone* 6:291–295, 1985.

Munns K. Effects of exercise on the range of joint motion in elderly subjects. In *Exercise and Aging: The Scientific Basis*, Smith EL, RC Serfass (eds.). Hillside, NJ: Enslow, 1981, pp. 167–178.

Murray MP. Shoulder motion and muscle strength of normal men and women in two age groups. *Clin Orthop Rel Res* 192:268–273, 1985.

Myhre L, S Robinson, A Brown, et al. Changes observed in the pulmonary function of older men following six months of physical training. In *American College of Sports Medicine Meeting,* Albuquerque, NM, 1970.

Narayanan N. Differential alterations in ATP-supported calcium transport activities of sarcoplasmic reticulum and sarcolemma of aging myocardium. *Biochimica et Biophysica Acta* 678:442–459, 1981.

Naughton J, F Nagle. Peak oxygen intake during physical fitness program for middle-aged men. *JAMA* 191:103–109, 1965.

Niinimaa V, RJ Shephard. Training and oxygen conductance in the elderly. I. The respiratory system. *J Gerontol* 33:354–361. 1978a.

Niinimaa V, RJ Shephard. Training and oxygen conductance in the elderly. II. The cardiovascular system. *J Gerontol* 33:362–367, 1978b.

Nilsson BE, SM Andersson, T Havdrup, et al. Ballet-dancing and weightlifting—effects on BMC. *Amer J Roentgen* 13:541–542, 1978.

Nilsson BE, NE Westlin. Bone density in athletes. *Clin Orthop Rel Res* 77:177–182, 1971.

Notelovitz M, C Fields, K Caramelli, M Doughert, and AL Schwartz. Cardiorespiratory fitness evaluation in climacteric women: Comparison of two methods. *Am J Obstet Gyn* 154:1009–1013, 1986.

Olree H, WC Stevens. Evaluation of physical fitness with special reference to girls and women.

Oscai LB, BT Williams, and BA Hertig. Effects of exercise on blood volume. *J Appl Physiol* 24:622–624, 1968.

Oyster N, M Morton, and S Linnell. Physical activity and osteoporosis in post-menopausal women. *Med Sci Spts Exer* 16(1):44–50, 1984.

Paffenbarger RS, AL Wing, and RT Hyde. Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol* 108:161–175, 1978.

Patton JF, JA Vogel, J Bedynek, D Alexander, and R Albright. Response of age forty and over military personnel to an unsupervised, self-administered aerobic training program. *Aviat Space Environ Med* 54(2):138–143, 1983.

Perren SM, A Huggler, M Russenburger, M Allgower, R Mathys, R Schenk, H Willenegger, and ME Muller. The reaction of cortical bone to compression. *Acta Orthop Scand.* Suppl:125, 1969.

Plowman SA, BL Drinkwater, and SM Horvath. Age and aerobic power in women: A longitudinal study. *J Gerontol* 34(4):512–520, 1979.

Pocock NA, JA Eisman, MG Yeates, PN Sambrook, and S Eberl. Physical fitness is a major determinant of femoral neck and lumbar spine bone mineral density. *J Clin Invest* 78:618–621, 1986.

Pogrund H, RA Bloom, and H Weinberg. Relationship of psoas width to osteoporosis. *Acta Orthop Scand* 57:208–210, 1986.

Polednak AP. Longevity and cause of death among Harvard College athletes and their classmates. *Geriatrics* 27:53–64, 1972.

Pollock ML. Physiological characteristics of older champion track athletes. *Res Quart* 45:363–373, 1974.

Pollock ML, J Broida, Z Kendrick, HS Miller, R Janeway, and AC Linnerud. Effects of training two days per week at different intensities on middle-aged men. *Med Sci Sports* 4:192–197, 1972.

Pollock ML, GW Dawson, HS Miller, A Ward, D Cooper, W Headley, AC Linnerud, and MM Nomeir. Physiologic responses of men 49 to 65 years of age to endurance training. *J Am Geriatr Soc* 24:97–104, 1976.

Pollock ML, HS Miller, R Janeway, AC Linnerud, B Robertson, and R Valentino. Effects of walking on body composition and cardiovascular function of middle-aged men. *J Appl Physiol* 30(1):126–130, 1971.

Port S, FR Cobb, RE Coleman, and RH Jones. Effect of age on the response of the left ventricular ejection fraction to exercise. *N Engl J Med* 303(20):1133–1137, 1980.

Prives M. Influence of labor and sport upon skeleton structure in man. *Anat Rec* 136:261, 1960.

Profant GR, RG Early, KL Nilson, F Kusumi, V Hofer, and RA Bruce. Responses to maximal exercise in healthy middle-aged women. *J Appl Physiol* 33(5):595–599, 1972.

Raab DM, JC Agre, M McAdam, and EL Smith. Light resistance and stretching exercise in elderly women: effect upon flexibility. In press.

Reinbold W-D, HK Genant, UJ Reiser, and ST Harrison. Bone mineral content in early-postmenopausal and postmenopausal osteoporotic women: Comparison of measurement methods. *Radiology* 160:469–478, 1986.

Ribisl PM. Effects of training upon the maximal oxygen uptake of middle-aged men. *Int Z Angew Physiol* 27:154–160, 1969.

Riggs BL, LJ Melton. Involutional osteoporosis. *New Engl J Med* 314:1676–1686, 1986.

Riggs BL, HW Wahner, WL Dunn, RB Mazess, KP Offord, and LJ Melton. Differential changes in bone mineral density of the appendicular and axial skeleton with aging: Relationship to spinal osteoporosis. *J Clin Invest* 67:328–335, 1981.

Rikli R, S Busch. Motor performance of women as a function of age and physical activity level. *J Gerontol* 41(5):645-649, 1986.

Ringe JD, E Steinhagen-Thiessen. Prevention of physiological age-dependent bone atrophy by controlled exercise in mice. *Age* 2:44–47, 1985.

Robinson S. Experimental studies of physical fitness in relation to age. *Arbeitsphysiol* 10:251–323, 1938.

Robinson S, DB Dill, RD Robinson, SP Tzankoff, and JA Wagner.

Physiological aging of champion runners. *J Appl Physiol* 41(1):46–51, 1976.

Rodeheffer RJ, G Gerstenblith, LC Becker, JL Fleg, ML Weisfeldt, and EG Lakatta. Exercise cardiac output is maintained with advancing age in healthy human subjects: Cardiac dilatation and increased stroke volume compensate for a diminished heart rate. *Circulation* 69(2):203–213, 1984.

Rook A. An investigation into the longevity of Cambridge sportsmen. *Br Med J* 3:774–777, Apr. 1954.

Roskamm H. Optimum patterns of exercise for healthy adults. *Can Med Assoc J* 96:895–900, 1967.

Rubin CT, LE Lanyon. Regulation of bone formation by applied dynamic loads. *J Bone Joint Surg* 66A:397–402, 1984.

Rubin CT, LE Lanyon. Regulation of bone mass by mechanical strain magnitude. *Calcif Tissue Int* 37:411–417, 1985.

Rundgren A, A Aniansson, P Ljungberg, and H Wetterqvist. Effects of a training programme for elderly people on mineral content of the heel bone. *Arch Gerontol Geriatr* 3:243–248, 1984.

Saltin B, G Grimby. Physiological analysis of middle-aged and old former athletes. Comparison with still active athletes of the same age. *Circulation* 38:1104–1115, 1968.

Saltin B, LH Hartley, A Kilbom, and I Astrand. Physical training in sedentary middle-aged men and older men. II. Oxygen uptake, heart rate and blood lactate concentration at submaximal and maximal exercise. *Scand J Clin Lab Invest* 24:323–334, 1969.

Sato I, Y Hasegawa, N Takahashi, Y Hirata, K Shimomura, and K Hotta. Age-related changes of cardiac control function in man. *J Gerontol* 36(5):1981, 1981.

Saville PD, BER Nilsson. Height and weight in symptomatic postmenopausal osteoporosis. *Clin Orthop* 45:49, 1966.

Schneider VS, J McDonald. Skeletal calcium homeostasis and countermeasures to prevent disuse osteoporosis. *Calcif Tissue Int* 36:S151–S154, 1984.

Schocken DD, JA Blumenthal, S Port, P Hindle, and RE Coleman. Physical conditioning and left ventricular performance in the elderly: Assessment by radionuclide angiocardiography. *Am J Cardiol* 52:359–364, 1983.

Seals DR, JM Hagberg, BJ Hurley, AA Ehsani, and JO Holloszy. Endurance training in older men and women. I. Cardiovascular responses to exercise. *J Appl Physiol* 57(4):1024–1029, 1984.

Sebban C, P Berthaux, H Lenoir, M Eugene, R Venet, Y Memin, X de la Fuente, and C Reisner. Arterial compliance, systolic pressure and heart rate in elderly women at rest and on exercise. *Gerontology* 27:271–280, 1981.

Sedgwick AW, AH Davidson, RE Taplin, and DW Thomas. Relationships between physical fitness and risk factors for coronary heart disease in men and women. *Aust NZ J Med* 14:208–214, 1984.

Sepic SB, MP Murray, LA Mollinger, GB Spurr, and GM Gardner. Strength and range of motion in the ankle in two age groups of men and women. *Am J Phys Med* 65(2):75–84, 1986.

Shaw DJ, DA Rothbaum, CS Angell, and NW Shock. The effects of age and blood pressure upon the systolic time intervals in males aged 20–89 years. *J Gerontol* 28(2):133–139, 1973.

Sheffield LT, JA Maloof, JA Sawyer, and D Roitman. Maximal heart rate and treadmill performance of healthy women in relation to age. *Circulation* 57(1):79–84, 1978.

Shephard RJ. World standards of cardiorespiratory performance. *Arch Environ Health* 13:664–672, 1966.

Shephard RJ. *Physical Activity and Aging*. London: Croom Helm, 1978.

Shephard RJ. Cardiovascular limitations in the aged. In *Exercise and Aging: The Scientific Basis*, Smith EL, RC Serfass (eds.). Hillside, NJ: Enslow, 1981.

Shephard RJ. The cardiovascular benefits of exercise in the elderly. *Topics Ger Rehabil* 1(1):1–10, 1985.

Siconolfi SF, TM Lasater, S McKinlay, P Boggia, and RA Carleton. Physical fitness and blood pressure: The role of age. *Am J Epidemiol* 122:452–457, 1985.

Sidney KH, RJ Shephard. Maximum and submaximum exercise tests in men and women in the seventh, eighth and ninth decades of life. *J Appl Physiol* 43(2):280–287, 1977[a].

Sidney KH, RJ Shephard. Perception of exertion in the elderly. Effects of aging, mode of exercise and physical training. *Percept Motor Skills* 44:999–1010, 1977[b].

Sidney KH, RJ Shephard. Frequency and intensity of exercise training for elderly subjects. *Med Sci Sports* 10(2):125–131, 1978.

Sidney KH, RJ Shephard, and JE Harrison. Endurance training and body composition of the elderly. *Am J Clin Nutr* 30:326–333, 1977.

Siegel MD, G Blomqvist, and JH Mitchell. Effects of quantitated physical training program on middle-aged sedentary men. *Circulation* 51:19–29, 1970.

Simkin A, J Ayalon, and I Leichter. Increased trabecular bone density due to bone-loading exercises in postmenopausal osteoporotic women. *Calcif Tissue Int* 39, 1986.

Sinaki M, MC McPhee, SF Hodgson, JM Merritt, and KP Offord. Relationship between bone mineral density of spine and strength of back extensors in healthy postmenopausal women. *Mayo Clinic Proc* 61:116–122, 1986.

Sinaki M, JL Opitz, and HW Wahner. Bone mineral content: Relationship to muscle strength in normal subjects. *Arch Phys Med Rehabil* 55(11):508–512, 1974.

Singh S, S Saha. Electrical properties of bone: A review. *Clin Orthop Rel Res* 186:249–271, 1984.

Sinning WE, MJ Adrian. Cardiorespiratory changes in college women due to a season of competitive basketball. *J Appl Physiol* 25:720–724, 1968.

Skinner JS, CM Tipton, and AC Vailas. Exercise, physical training and the ageing process. Chapter 13. In *Lectures on Gerontology, Volume I: On Biology of Ageing, Part B*, Vidik A (ed.). London: Academic Press, 1982, pp. 407–439.

Skrobak-Kaczynski J, KL Andersen. Age dependent osteoporosis among men habituated to a high level of physical activity. *Acta Morphol Neerl-Scand* 12:283, 1974.

Smith, EL. Special considerations in developing exercise programs for the older adult. In *Behavioral Health: A Handbook of Health Enhancement and Disease Prevention*, ed. Matarazzo JD, NE Miller, SM Weiss, JA Herd, and SM Weiss. New York: John Wiley, 1984, pp. 525–546.

Smith EL, C Gilligan. Physical activity prescription for the older adult. *Physician Sportsmed* 11:91–101, 1983.

Smith EL, C Gilligan, M McAdam, CP Ensign, and PE Smith. Bone involution decrease in exercising middle-aged women. *Calcif Tissue Int* 36:S129–S138, 1984.

Smith EL, W Reddan, and PE Smith. Physical activity and calcium modalities for bone mineral increase in aged women. *Med Sci Sports Exer* 13(1):60–64, 1981.

Smith J, JM Walker. Knee and elbow range of motion in healthy older individuals. *Physical and Occupational Therapy in Geriatrics* 2:31–38, 1983.

Sosna J. The influence of fencing on the limb musculature. In *Physical Fitness and Its Laboratory Assessment*, Kral J, V Novotny (eds.). Prague: Universita Karlova, 1970.

Sprynarova S, J Parizkova. Comparison of the functional, circulatory and respiratory capacity in girl gymnasts and swimmers. *J Sports Med* 13:165–172, 1969.

Stamford BA. Effects of chronic institutionalization on the physical working capacity and trainability of geriatric men. *J Appl Physiol* 33:346–350, 1972a.

Stamford BA. Physiological effects of training upon institutionalized geriatric men. *J Gerontol* 27(4):451–455, 1972b.

Stillman RJ, TG Lohman, MH Slaughter, and BH Massey. Physical activity and bone mineral content in women aged 30 to 85 years. *Med Sci Spts Exer* 18(5):576–580, 1986.

Sucec A. Oxygen uptake and heart rate in middle-aged males participating in an adult fitness program. *Am Corr Ther J* 23(4):98–103, 1969.

Suominen H, E Heikkinen, H Liesen, D Michel, and W Hollman. Effects of 8 weeks endurance training on skeletal muscle metabolism in

56–70 year old sedentary men. *Europ J Appl Physiol* 37:173–180, 1977a.

Suominen H, E Heikkinen, and T Parkatti. Effects of eight weeks' physical training on muscle and connective tissue of the m. vastus lateralis in 69-year-old men and women. *J Gerontol* 32:33–37, 1977b.

Taylor HL, ER Buskirk, and RD Remington. Exercise in controlled trials of the prevention of coronary heart disease. *Fed Proc* 32:1623, 1973.

Thomas SG, DA Cunningham, PA Rechnitzer, AP Donner, and JH Howard. Determinants of the training response in elderly men. *Med Sci Sports Exer* 17:667–672, 1985.

Tietz N, uH Valentin. Das Verhalten der Sauerstoffaufnahme, des Wirkinungsgrades und des Herzleistungsquotienten in den verschiedenen Belastungsstufen und Altersklassen. *Sportmedizin* 8:148–158, 1957.

Tlusty L. Physical fitness in old age. I. Aerobic capacity and the other parameters of physical fitness followed by means of graded exercise in ergometric examination of elderly individuals. *Respiration* 26:161–181, 1969.

Tuxworth W, AM Nevill, C White, and C Jenkins. Health, fitness, physical activity, and morbidity of middle aged male factory workers, I. *Br J Indust Med* 43:733–753, 1986.

Tzankoff SP, S Robinson, FS Pyke, and CA Brawn. Physiological adjustments to work in older men as affected by physical training. *J Appl Physiol* 33(3):346–350, 1972.

Uhthoff HK, ZFG Jaworski. Bone loss in response to long term immobilization. *J Bone Joint Surg* 60B:420, 1978.

Vodak PA, WM Savin, WL Haskell, and PD Wood. Physiological profile of middle-aged male and female tennis players. *Med Sci Sports Exer* 12(3):159–163, 1980.

Vogel JA, JF Patton, RP Mello, and WL Daniels. An analysis of aerobic capacity in a large United States population. *J Appl Physiol* 60:494–500, 1986.

von Dobeln W, I Astrand, and A Bergstrom. An analysis of age and other factors related to maximal oxygen uptake. *J Appl Physiol* 22:934–938, 1967.

Wahren J, B Saltin, L Jorfelt, and B Pernow. Influence of age on the local circulatory adaptation to leg exercise. *Scand J Clin Lab Invest* 33:79–86, 1974.

Walker JM. Active mobility of the extremities in older subjects. *Phys Ther* 64(6):919–923, 1984.

Watson RC. Bone growth and physical activity in young males. In *International Conference on Bone Mineral Measurements*. Washington, DC: DHEW #NIH 75-683, 1974, pp. 380–385.

Weber F, RJ Barnard, and D Roy. Effects of a high-complex-carbohydrate, low-fat diet and daily exercise on individuals 70 years of age and older. *J Gerontol* 38(2):155–161, 1983.

Wessel JA, DA Small, WD Van Huss, DJ Anderson, and DC Cederquist. Age and physiological responses to exercise in women 20–69 years of age. *J Gerontol* 23(3):269–278, 1968.

Wessel JA, DA Small, WD Van Huss, WW Heusner, and DC Cederquist. Functional responses to submaximal exercise in women 20–69 years. *J Gerontol* 21:168–181, 1966.

Wessell JA, WD Van Huss. The influence of physical activity and age on exercise adaptation of women, 20–69 years. *J Sports Med Phys Fitness* 9(3):173–180, 1969.

West CC. Measurement of joint motion. *Arch Phys Med* 26:414–425, 1945.

Whedon GD. Disuse osteoporosis: Physiological aspects. *Calcif Tiss Int* 36:S146–S150, 1984.

Whedon GD, L Lutwak, P Rambaut, M Whittle, C Leach, J Reid, and M Smith. Effect of weightlessness on mineral metabolism; metabolic studies on skylab orbital space flights. *Calcif Tissue* 21:S423–S430, 1976.

White MK, RB Martin, RA Yeater, RL Butcher, and EL Radin. The effects of exercise on the bones of postmenopausal women. *International Orthopaedics* 7:209–214, 1984.

Wilmore JH. Acute and chronic physiological responses to exercise. In *Exercise in Cardiovascular Health and Disease* 23(3), Amsterdam EA, JH Wilmore, and AN DeMaria (eds.). New York: Yorke Medical Books, 1968, pp. 269–278.

Wilmore JH, J Royce, R Girandola, F Katch, and V Katch. Physiological alterations resulting from a 10-week program of jogging. *Med Sci Sports* 2:7–14, 1970.

Woo SL, SC Kuei, D Amiel, MA Gomez, WC Hayes, FC White, and WH Akeson. The effect of prolonged physical training on the properties of long bone: A study of Wolff's law. *J Bone Jt Surg* 63A:780–786, 1981.

Wright TW, CW Zauner, and R Cade. Cardiac output in male middle aged runners. *J Sports Med* 22:17–22, 1982.

Young DR, WJ Niklowitz, and CR Steele. Tibial changes in experimental disuse osteoporosis in the monkey. *Calcif Tissue Int* 35:304–308, 1983.

Young DR, VS Schneider. Radiographic evidence of disuse osteoporosis in the monkey. *Calcif Tissue Int* 33:631–639, 1981.

Zanzi I, KJ Ellis, J Aloia, and SH Cohn. Effect of physical activity on body composition. In *Osteoporosis: Recent Advances in Pathogenesis and Treatment,* Deluca HF, HM Frost, CC Johnston, and AM Parfitt (eds.). Baltimore, MD: University Park Press, 1981, pp. 139–146.

Zauner CW. Physical fitness in aging men. *Maturitas* 7:267–271, 1985.

Zir LM, RH Rahe, RT Rubin, and RJ Arthur. Effects of strenuous swim competition in the older age group. *J Sports Med Phys Fitness* 12:180–185, 1972.

Part V

Lessons From Community, National, and International Studies

Chapter 15

Lessons from Tecumseh on the Assessment of Physical Activity and Fitness

Henry J. Montoye, Ph.D.
University of Wisconsin, Madison

Introduction

The Tecumseh Michigan Community Health Study (1, 2) has much in common with the National Health and Nutrition Examination Survey (NHANES). The Tecumseh Study is an ecologic investigation of an entire community—not just of the people living there but of the biological, physical, and social environment and how these factors interact to enable some individuals to maintain health and others to develop a proclivity to disease. In that sense, it is broader than the NHANES. Nevertheless, the main thrust of the Tecumseh Study is the examination and evaluation of the residents of the town and its rural fringe. As in the NHANES, the ultimate application lies in the field of prevention.

The study population in Tecumseh included about 11,000 people, smaller than the anticipated NHANES sample, but sufficiently close in size to require the same epidemiologic approach and dictating similar limitations. Like NHANES, the Tecumseh project was an observational investigation involving no intervention. But also, as in the NHANES, followup of the respondents is planned; both investigations include elements of cross-sectional and longitudinal studies. The last round of examinations in Tecumseh required 3½ years to complete, again similar to NHANES.

Even though the Tecumseh Study ultimately involved the entire population, some measurements were taken on only a sample. When this was necessary, usually a cluster sample of randomly selected households was employed. This again is similar to NHANES. The two studies also included both sexes and almost the entire age range for most measurements. In both studies, it is impractical to ask subjects to return if equipment breakdown or other problems occur.

In the Tecumseh investigation, a high response rate was alway a concern as it is in NHANES. The overall response rate in Tecumseh was 83 percent, not much different from the 74–91 percent in NHANES. There were important lessons learned in Tecumseh relative to the response rate in exercise and fitness tests that will be commented on later in this report. Data on nonresponders were studied just as was done in NHANES. Finally, Tecumseh was selected for study for a number of reasons, one of them being the diversification of occupations and the socioeconomic classes represented. This also is similar to NHANES. Other studies of physical activity and fitness measurements have included only a limited segment of the population—college graduates, for example. This has important implications for the methods employed.

There were differences in the two populations that might also be important. There likely will be differences in the distribution of ethnic backgrounds. At the time of the second round of examinations, there was only one black family in Tecumseh. The town is located in the north, which has a bearing on the physical activities engaged in by the participants, particularly in winter. The clinic and laboratory in Tecumseh were not mobile, and expertise of all kinds was available at the University of Michigan in Ann Arbor, only 45 miles distant. This made possible some measurements and observations that would not be possible in a mobile unit. These differences will be borne in mind during the discussion that follows.

Assessment of habitual physical activity

Most investigators who have tried to measure physical activity by questionnaire or interviews agree that the soft data generated by this approach leaves much to be desired. Hence, the search for more objective methods continues. For a detailed discussion of these and other techniques of assessing habitual physical activity in population studies, the reader is referred to a recent review (3). The use of mechanical or electronic devices to estimate movement or energy expenditure looks promising, but none has been adequately validated as yet. In any case, they would not be appropriate in surveys such as NHANES. As a result of a previous study (4), we were aware of the limitations of the questionnaire and interview approach, but there appeared to be no alternative. Indeed, this still seems to be the only practical method for the NHANES.

An assessment of the habitual physical activity of the respondents in the Tecumseh Study was included to investigate the relationship of the amount and intensity of physical activity in one's daily life to the maintenance of health, particularly of the cardiorespiratory system. It is not known which dimensions of physical activity might be related to health or disease, so the plan was to retain as much information as possible—such as frequency, duration, energy cost estimates, and specific sports.

Although an earlier attempt to estimate habitual physical activity by a self-administered questionnaire met with only limited success (4), we decided to revise the form used and try it once more. The 2,885 dwelling units in the study were randomly allocated to 10% samples. These samples were then canvassed successively. The first 10% probability sample of dwelling units in Tecumseh provided about 350 men and women for this preliminary trial. Despite field trials of preliminary forms and numerous revisions in the wording of questions, there was still considerable evidence of misinterpretation by the respondents. Furthermore, many questions were unanswered or answered incompletely. It was concluded that, with a total population containing people of varying occupations and education, a personal interview was necessary to obtain accurate data.

Questionnaire and interview forms were again revised and field tested. A second revision followed that was used on the final 80% of the male population (1962–65). This form was published (5) and was used a second time with all males 16 years and over in Tecumseh who were not in school (1965–69).

The interviewer, at first contact, delivered a self-administered questionnaire to the respondent. If the respondent was not employed at the time, a modified form was used. In the questionnaire, inquiries were made about occupation, hours worked, transportation to and from work, and participation in major home repairs and maintenance. Leisure-time sports, gardening, and other physical activities were checked on a list.

At the next contact, the trained interviewer spent from 30 minutes to 1 hour with each respondent inquiring about the physical activity involved in the respondent's occupation and leisure activities. A prepared supplementary interview form was used for occupations and each leisure-time activity checked. The questionnaire and interview forms were designed to estimate physical activity during the preceding year.

The supplements were designed to probe for differences in intensity of physical activity. For example, it was learned in preliminary work that the occupational title was frequently of little or no value in estimating the physical activity associated with the job. Therefore, questions were asked to determine as accurately as possible how much actual physical activity was involved. In only one general occupation, farming, this was not possible. The tasks on the farm were too varied. Therefore, the interviewer ascertained that the man actually worked on a farm and was not a gentleman farmer. She also determined approximately how many hours a week the respondent worked on the farm. It is clear that another questionnaire is needed to assess specifically the physical activity of farmers.

Questions about leisure-time activities on the supplements were designed to determine, for example, whether a man fished while sitting in a boat or walking in a stream, because the energy costs are quite different in these two methods of fishing. Similarly, if a man plays baseball on an organized team, he likely expends more energy than in playing catch with his young son. It is important from the standpoint of energy cost to make such distinctions.

Objective scoring

It has not been possible as yet to establish with certainty the validity of our procedures by employing external criteria such as direct assessment of energy cost, recording the heart rate, or making direct observations. However, it was recognized that an objective method had to be developed for interpreting and scoring these data. Therefore, the reliability of interpretation of the record was studied as follows. Two randomly selected samples from the data collected, each composed of 20 completed interview forms, were subjectively ranked by three judges independently, with the most active respondent at the top and the least active at the bottom. This was done for each of three components (namely, occupation, leisure, and occupation plus leisure) on the basis of responses recorded on the questionnaire and interview forms. After approximately 6 weeks, sufficient time for the first ranking to be forgotten, each judge ranked the two samples a second time. The rank order correlations illustrating each judge's consistency in ranking within sets ranged from 0.96 to 0.99. Because the two rankings by each judge within each sample correlated so highly, the first ranking was arbitrarily selected for the comparisons between judges. The correlations between judges for both samples ranged from 0.90 to 0.99 for occupation, from 0.66 to 0.82 for leisure, and from 0.86 to 0.94 for occupation plus leisure. The mean of the three judges' subjective ranks for each respondent was then calculated for occupation, leisure, and occupation plus leisure. This was done for both samples. These mean ranks formed the criterion for the development of an objective scoring system. A simple objective system was desirable to enable clerks to score the record with little subjective judgment required.

The first step in the preparation of the scoring system was the construction of a table of metabolic costs for 36 occupational and 63 leisure tasks, and these have been published (5). The work metabolic rate to basal metabolic rate ratio, or METS, was used in the metabolic cost index. This index was used to eliminate the necessity of considering the subject's body weight, of converting the work to calories, et cetera. The method assumes that a task performed by a heavy person raises his METS to the same extent as the same task performed by a person weighing less, even though the caloric expenditure might be different. Because much of the activity recorded involves moving one's own body weight, errors in making this assumption probably are not serious. There are other limitations to the scoring procedure. The occupational and leisure activities are those found in this particular area. Data on energy cost of many activities are available in the literature, so reasonably good estimates can be made. However, for other activities, data are lacking, and therefore our estimates may not be accurate. It is also well known that the energy cost for the same task varies from individual to individual depending upon skill, walking surface, equipment, clothing, environmental conditions, terrain, and perhaps many other factors. Finally, the activities themselves are carried out in a different manner in various parts of the world and even within areas of the United States. Hence, this table may have to be amended, changed, or both, when applied to another population or another culture. For these reasons, the average metabolic cost of occupational and/or leisure task as calculated is used primarily to classify people into one of several physical activity categories. We are not deluded into thinking these are precise quantitative values. However, using three or four groups of people of different habitual activity enables one to test hypotheses about the relationship of physical activity to biomedical measurements, risk factors, et cetera.

The procedure to categorize subjects was as follows. The work-to-basal ratio for a particular activity was multiplied by the number of hours during the year in which the subject engaged in that activity. This was repeated for all the activities requiring appreciable energy expenditures. The products were summed and divided by the total hours to arrive at a weighted average daily work-to-basal metabolic ratio. When estimating the overall activity classification, the leisure and occupational activities were averaged, weighted of course, by the hours in each. For our purpose, several assumptions were made. Because it was not feasible to determine how many hours each subject slept or how many hours were involved in eating meals, shaving, et cetera, it was assumed that each respondent slept 8 hours a day, at a work-to-basal ratio of one. When the subject was not sleeping or doing physical work, it was assumed that he was eating, reading, driving a car, watching television, or doing other quiet activities. An average work-to-basal metabolic ratio of 1.8 was assigned to the time spent in this way. By using data processing equipment, it is necessary only to punch in the hours spent in the various activities and, in some instances, the work-to-basal ratio. A program can then be written to do the actual calculations.

In order to assess the validity of this objective method of calculating the average energy expenditure, the subjective average ranking of the two sets of 20 respondents by the three judges was used as a criterion. The subjects were then ranked by average expenditure according to the objective method just described. The rank order correlation coefficients have been published (5), and they ranged from 0.86 to 0.97. It is clear that the two methods give approximately the same result.

Occupation activity

In Tecumseh, as would be true in the NHANES, occupational activities varied greatly. Because there was interest not only in physical activity (movement) but in energy expenditure, details about the jobs of the respondent were sought. The respondent filled out a brief questionnaire about the nature of his work, which was followed by more detailed questions by the interviewer (5). Some men were unemployed because of retirement, layoff, injury, or other reasons. A brief self-administered form was filled out by these respondents (5). The interviewer then made inquiries about the last job held. The occupational activity questionnaire/interview is being revised and simplified by Montoye and Jacobs. The latest version may be found in appendix A.

A significant number of men were engaged in more than one occupation. In these instances, information concerning second or third jobs was obtained with the same forms used for the first occupation. In the first round in which occupational energy expenditure was being sought, data were obtained on the occupational activities for the last 3 months as well as for the entire preceding year. The correlation coefficient between the two time periods, again using the objective scoring system, was 0.98. Although there were some individuals who changed jobs during the 3 months preceding the interview, for the most part the new job was fairly similar in physical demand to the preceding one.

For many occupations, objective estimates of energy expenditure could be obtained because the respondent was able to tell the interviewer the exact weight he lifted, for example, and the precise number of times it was lifted per day. In other cases, carpentry for instance, the tasks were so varied that an overall estimate of energy expenditure for this kind of work was assigned.

Leisure-time activities

The leisure activities included in the Tecumseh activity recall record are those found in this particular area and are necessarily limited by weather conditions, the accessibility of naturally occurring recreational sites such as ski hills and streams or lakes for fishing, and the availability of such other facilities as tennis courts and golf courses. The Tecumseh area is situated close to many recreational facilities; and within the city and the nearby surrounding area, there are 5 tennis courts, 19 bowling lanes, 3 swimming pools (including 1 heated year-round pool at Adrian, Michigan) and a 5-acre pond in a 22-acre park. A recreation center has space for dance, exercise, and instructional groups. Local schools are used for various city activities that include men's exercise groups and a judo club. Softball leagues are open to all men in the city, and adequate opportunity is available to participate on these teams. Within a 20-minute drive from Tecumseh, there is a large recreation area that includes many lakes suitable for swimming, boating, and fishing in the summer and hills for winter skiing. The lands surrounding Tecumseh can be used for small game and bird hunting.

The forms used in the assessment of nonoccupational activities have been published (5) and form the basis for a revision by Taylor and others (6).

Indexes of physical activity

An example of the calculations for a particular subject is shown in appendix B. This of course was done by computer using prepared software after the basic data were entered. Our present knowledge does not tell us what index of activity might be most closely related to health. Therefore, 15 indexes were investigated as follows:

Occupation only

a. Hours worked per week

b. Weighted mean WMR/BMR (or METS)
Each activity MET is weighted by the hours per day, on the average.

$$\text{Thus, } \overline{X}_w = \frac{\Sigma\, Wi\, Xi}{\Sigma\, Wi}$$

where $\overline{X}_w$ = weighted mean MET
Wi = hours at a particular MET
Xi = particular MET

c. Weighted standard deviation for METS

$$\text{Thus, } S_w = \sqrt{\frac{\Sigma\, Wi\, (Xi - \overline{X}_w)}{\Sigma\, Wi}}$$

where S_w = weighted standard deviation of METS

d. Weighted coefficient of variation CV_w

$$\text{Thus, } CV_w = \frac{S_w}{\overline{X}_w}$$

e. Peak METS. This was defined as the lowest MET in the most active 1½ hours per week. The highest MET attained by an individual was considered his peak if he averaged 1.5 hours or more per week at this level. If not, the next highest ratio was used as the peak provided the time spent at the two levels totaled 1.5 hours. The process was continued until a MET was reached such that the subject spent 1.5 hours a week at levels at least as high.

Leisure time

f. Hours of active leisure per week prorated for the year
g. Weighted mean METS
h. Weighted standard deviation for METS
i. Weighted coefficient of variation for METS
j. Peak METS
k. Estimate of energy expenditure during active leisure expressed in METS. This is simply the sum of each activity MET multiplied by the hours spent in that activity.

$$\text{Thus, Leisure index} = \Sigma\, Wi\, Xi,$$

where Wi = hours per week prorated for the year for a particular activity
Xi = estimated MET level for that activity

Occupation plus leisure time

l. Weighted mean METS (which is an estimate of average energy expenditure per week)
m. Weighted standard deviation for METS
n. Weighted coefficient of variation for METS
o. Peak METS.

The distribution of these indexes and their intercorrelations were studied within each of five age groups (16–29 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years). From this analysis the indexes were reduced to six, namely:

Occupation

1. The average number of hours per week of occupational work including all occupations of the respondent (a. above).
2. A weighted mean work-to-basal ratio utilizing the METS associated with each of the subject's occupational activities, weighted by the number

of hours spent at each. This reflects the rate of energy expended on the job (b. above).

Leisure time

3. The average number of hours per week in which the respondent was engaged in active leisure. This does not include hours spent at reading, playing cards, etc. (f. above).
4. A weighted mean work-to-basal ratio. This mean reflects the average rate of energy expenditure for active leisure and hence does not estimate the total energy expenditure during active leisure (g. above).
5. Active leisure index, the sum of products for each active leisure pursuit (METS × hours spent). This reflects the total energy expenditure during active leisure (k. above).

Occupation plus leisure time

6. A weighted mean work-to-basal ratio including both occupational and leisure activities. This index reflects the average 24-hour daily rate of energy expenditure (l. above).

Next, the respondents were grouped by decade of age. Distributions for each of the six indexes and for each of the age groups were examined. Our plan was to classify respondents who were in the lowest 20 percent (for a particular index in a particular age group) as least active, those in the middle 60 percent as intermediate, and those in the upper 20 percent as most active. In most cases, this was possible. However, in the case of hours worked (1. above), about one-half of the respondents were placed in the lower group, and about 10 percent were placed in the upper group. For each index, the proportions in each group (lower, middle, and upper) were the same for all age ranges.

Generally when studying physical activity as related to coronary risk factors and other variables, these were the groups used. Farmers, as indicated earlier, were treated as a separate group.

A careful validation of this method of assessing habitual physical activity was not done. However, during the third round of examinations in Tecumseh, a nutritionist made an independent assessment of caloric intake by dietary recall. It was then possible to correlate this estimate of caloric intake, divided by body weight, with the last index described above which is an estimate of caloric expenditure. The most sedentary 20 percent were successfully separated from the intermediate 60 percent and the most active 20 percent, but the latter two groups were not as well differentiated. The average intake was about 300 kcal greater for the most active compared with the least active group (7).

The six indexes were studied in relation to a number of disease risk factors. The most active group was leaner than the intermediate group, which was in turn leaner than the least active group. But this was true only when the mean METS for occupation or occupation plus leisure were employed, that is, the estimate of total energy expended in the occupation or total 24-hour estimate (2). In similar comparisons, the active group had significantly lower systolic and diastolic blood pressure (8), serum total cholesterol and triglycerides (7), serum uric acid (9), and improved glucose tolerance among the leaner subjects (10). However, in selected populations, other indexes may be important. For example, subjects who worked the longest hours (Index 1) had the highest blood pressure even though they were not overrepresented in the more sedentary group (8). In this group, leisure activity was inversely correlated with blood pressure.

As one might expect, the number of hours worked per week changes little with age, but the number of hours spent per week in active leisure decreases with age (2, 11, 12). The occupational mean METS decreases slightly with age, but the leisure time mean METS decreases sharply (2, 11–13). Energy expended and hours spent at work and leisure were correlated with occupational classification (12).

Taylor and colleagues (6) developed a modification of the Tecumseh questionnaire/interview for leisure activities, which is simpler and probably more useful.

Lessons from Tecumseh: Summary

1. It is practical to assess habitual physical activity in a large number of males, ages 16 and over, who are not in school, by a questionnaire/interview approach. However, highly precise results are not likely to be obtained.
2. The assessment of habitual physical activity in females who are responsible for maintaining a home is more difficult, requires more time, and therefore is potentially subject to a reduced response rate. A form that has been tried but not validated in Tecumseh is reproduced as appendix C.
3. The assessment of occupational activities during the past 3 months is highly correlated (r = .98) with the 12-month assessment, so only a 3-month or 12-month assessment is necessary.
4. In a population of varied socioeconomic status, occupational activity is very important in differentiating individuals by habitual physical activity.
5. Leisure time activity, however, may be the only source of significant physical activity for a large segment of the population who have sedentary occupations.
6. In a population of varied socioeconomic status, a meaningful estimate of habitual physical activity cannot be obtained from a self-administered questionnaire; an interview followup is necessary.
7. A careful training period is required for interviewers to develop skill in assessing habitual physical activity. Even so, a panel of judges is necessary to resolve problems of interpretation in a small percent of respondents.
8. It is possible to develop a computer program to calculate various physical activity indexes.

9. At a minimum, the following information has been found useful: hours worked, an estimate of energy expenditure in METS for occupational and leisure activity, specific leisure activities, and time spent on each.

Assessment of physical fitness

There is no generally accepted measure of total physical fitness. However, the single most popular measurement is the maximum amount of oxygen that can be taken in from air, delivered, and utilized in the tissues ($\dot{V}O_2$max). Sometimes it is referred to as aerobic capacity or metabolic capacity. It is popular because, of all the various measurements of fitness, it is the most closely related to maximum work capacity (endurance) and the one possibly related to such chronic diseases as coronary heart disease, diabetes, and emphysema. The measurement is expressed in liters per minute, or, more commonly, ml per kg body weight per minute, so that people of various sizes may be compared. The measurement of $\dot{V}O_2$max is routine in many laboratories and was measured using a motor-driven treadmill in the third round of exams in the Tecumseh study. This study clearly demonstrated that it is practical and safe to measure $\dot{V}O_2$max, or near max, in an unselected naive population of males and females, none of whom had ever walked on a treadmill before. Although the response rate in Tecumseh was 83 percent for males, it was only 59 percent in females. Above age 14, the response rate averaged only 52 percent for females. Girls and women beyond the age of 14 are reluctant to exercise maximally. The motivation to participate in the Tecumseh study was probably higher than it would be in NHANES. Also, the laboratory in Tecumseh was a somewhat permanent facility. A mobile laboratory might present more problems for assessing $\dot{V}O_2$max. A maximum exercise test also requires more careful medical screening than a submaximum test. Nevertheless, with the advancements that have been made in instrumentation, it should be possible to measure $\dot{V}O_2$max in the NHANES survey.

Estimation of $\dot{V}O_2$max

If the decision is made not to measure $\dot{V}O_2$max, it is possible to estimate $\dot{V}O_2$max. The best estimates of $\dot{V}O_2$max may be had by an all-out endurance performance, that is, a maximum work capacity test in which the large muscles of the body are employed. This has generally taken the form of one of two types of tests. In the first type, the subjects perform work of a gradually increasing intensity until they are no longer able to continue. Generally, either a motor-driven treadmill is employed in which the grade and/or speed is increased, or a stationary bicycle is used in which the resistance is increased. The test used in Tecumseh consisted of walking at 3 mph (4.84 $\text{km}\cdot\text{hr}^{-1}$) at 0-percent grade for 3 minutes; thereafter, the grade was increased 3% after 3 minutes, at each grade, with belt speed constant, until the subject indicated exhaustion. Oxygen uptake, $\dot{V}CO_2$, ventilation, and respiration rates were measured during the third minute of each grade using electronic gas analyzers and a recording dry gas meter. These measurements, together with the ECG and heart rate, were recorded on an E. for M.8-channel photographic recorder. Details of these procedures have been published (2, pp. 40–54). The correlation coefficients between treadmill time to exhaustion and $\dot{V}O_2$max in most of our population are shown in column 3 of table 1 for various age groups, the mean being .84.

The second test form consists of a run or run/walk on the open track where the subject runs either a fixed distance or runs for a fixed time. In the former, the score is the time required to run a fixed distance; and in the latter, it is the distance covered in a fixed time. The time or distance required should be great enough so that the source of energy is principally aerobic. Balke (14) originally proposed a 15-minute run, and later Cooper (15) modified the test to limit the time to 12 minutes. Correlation coefficients expressing the relationship of performance in these tests to measure $\dot{V}O_2$max range from about 0.85 to about 0.93, depending on the group studied (16). However, these tests require careful medical screening, cooperation, and motivation of the subject.

The practical alternative to a maximal performance test is a submaximal test for estimating $\dot{V}O_2$max. The remaining discussion in this section will be confined to such measurements. Tests of the physiological response to submaximal exercise have a long history (17). The earlier tests have not stood up well under critical evaluation, because they were not well standardized and/or the exercise was so mild as to differentiate only those subjects on opposite ends of the distribution. Currently, submaximal tests generally use a motor-driven treadmill, a stationary bicycle, or a fixed or variable step. In Europe, the bicycle ergometer found more favor than in the United States, where the treadmill or step test is more popular. The advantages and disadvantages of each will be discussed below together with lessons learned

Table 1. Correlation coefficients expressing the relationship between measured $\dot{V}O_2$max and an estimate from maximum treadmill times in 367 males

Age	*n*	Maximum treadmill time [1]
10–14 years	82	.84 (3.3)
15–19 years	99	.91 (2.9)
20–24 years	48	.75 (2.7)
25–29 years	32	.87 (2.3)
30–34 years	38	.76 (1.9)
35–39 years	68	.81 (3.0)
Weighted mean	367	.84 (2.8)

[1] Standard errors of estimate in ml/min/kg body weight.
Source: Reprinted by permission of the American Alliance for Health, Physical Education, Recreation and Dance, 1900 Association Drive, Reston, Va. 22091

from the Tecumseh experience. It should be clearly understood that submaximal tests, whatever their form, provide only estimates of $\dot{V}O_2$max.

Treadmill test

In Tecumseh, the treadmill was selected for measuring and estimating $\dot{V}O_2$max during the third round of examinations. The treadmill has several advantages over the bicycle ergometer or step test as follows:

a. In Tecumseh, as in the NHANES, subjects of both sexes, various body builds, and ages were to be tested. Some subjects weighed as much as 350 pounds, and it was our experience that subjects of this size have difficulty with a bicycle ergometer. For one thing, the usual saddle is not satisfactory. Generally if subjects can do any exercise, they can walk. There are exceptions to this, but in the Tecumseh study about 1,400 subjects who had never walked on a treadmill were taught to do so within 5–10 minutes. From age 18 on, efficiency of walking does not vary with age even at grade levels up to 15 percent (18).
b. Walking on a treadmill utilizes most of the large muscle groups, hence, there is a minimum of local fatigue. This is one reason why $\dot{V}O_2$max measured on the treadmill is usually slightly higher than when measured on a bicycle ergometer.
c. The work load is standardized because the belt speed and grade are fixed and cannot be altered by the subjects.
d. A steady state is reached within 2 minutes when walking on the treadmill, provided the grade increments are small (for example, 3 percent) (19).
e. An adjustment for weight is generally not necessary because the subjects lift their own body weight.

There are several disadvantages to the treadmill that are listed below:

a. A source of electrical power (at least 110 volts) is required.
b. The initial cost of a treadmill is greater than a bicycle ergometer or step.
c. The measurement of blood pressure during the test is more difficult than on the bicycle ergometer, especially at the higher work loads, because the subject has a tendency to swing the arms.
d. In some cases, the subjects experience pain in the lower legs for a few days following the test. This is especially true if the subjects walk at the higher grades, and it is more true for women who are used to wearing higher heels. This was avoided for the most part by providing wedge shoes in Tecumseh.
e. It is advisable for the subjects to wear rubber soled shoes or other footwear that minimizes slipping on the bed of the treadmill.
f. There is always a chance, albeit small, of a fall from a treadmill, which is much less likely with a bicycle ergometer.

Bicycle ergometer test

The advantages of a bicycle ergometer are well-known but may be worth repeating.

a. The equipment is simpler, especially the mechanical ones, which do not require a source of electrical power. In electrically powered bikes, the load may be increased by small precise units; however, mechanically operated bikes are sufficiently accurate for field studies. There are sometimes problems with their calibration, however.
b. There is less danger of a fall and less likelihood of post-exercise hypotension and its problems.
c. Manual recording of blood pressure and heart rate is more easily accomplished than when the subject is walking on a treadmill or performing on a step test.

On the other hand, there are disadvantages to the bicycle ergometer.

a. This is not a natural exercise for many people in the United States. In Tecumseh, for example, about 95 percent of the men, ages 35–65 years, had not ridden a bicycle in the previous year.
b. At a minimum, a seat height adjustment is necessary for each participant.
c. Local fatigue of the leg muscles is more of a problem than with the treadmill.
d. Maintenance of prescribed speed is a problem. Termination of the test when subject does not maintain the speed becomes subjective.
e. An adjustment for body weight becomes necessary.

Step test

There are some practical advantages to the step test.

a. Only a minimum of equipment is needed, namely, a fixed height bench, several of different heights, or a variable height platform and a metronome or tape-recorded cadence.
b. This is an activity, at least at the lower bench heights, that people are used to doing.
c. No adjustment for weight is needed because subjects lift their own weight.
d. Work may be easily calculated.

There are also disadvantages.

a. Maintaining a cadence is a more serious problem than in the case of the treadmill or bicycle ergometer.
b. Risk of injury is a concern, especially when the bench height is greater than 8 inches (the usual height of a stair).
c. The manually recording of heart rate and blood pressure is more difficult than with a bicycle ergometer or treadmill.
d. Incorrect performances of the step can be a problem;

frequently at the higher bench heights or cadences, subjects have a tendency not to straighten the legs when completing the ascension on the bench.

e. Leg length affects the test results in extreme cases.

For estimating $\dot{V}O_2$max, various physiological responses to submaximal loads can be measured, but heart rate (HR) is most useful for this purpose. Heart rate may be measured more easily than other functions, and more data is available for maximum heart rate and its relation to age, sex, etc. than for maximum values of other physiological variables.

The treadmill test can be fixed at one work load, and the heart rate response to this load can be measured; the lower the heart rate the better the fitness (estimated $\dot{V}O_2$max). The disadvantage of this kind of test is that the work load is likely to be too low for some subjects and too high for others. The effects of factors other than work load on the heart rate are greater at the lower work loads than at higher work loads.

If $\dot{V}O_2$ and HR are measured at a number of increasing submaximal loads, one can predict the $\dot{V}O_2$max using a generalized HR/$\dot{V}O_2$ curve. This has been done in a number of studies, and the correlation coefficients between predicted and measured $\dot{V}O_2$ range from 0.47 to .90 (20). The most popular methods, all of which use HR and $\dot{V}O_2$ at one or more submaximal loads, are those by Åstrand and Rhyming (21), Margaria and others (22), and Maritz and colleagues (23). In the Tecumseh study, we had measured $\dot{V}O_2$max on 474 males, ages 10–39 years, in whom we also had $\dot{V}O_2$ and HR data on 5 submaximal loads. The numbers of subjects in each age group are shown in table 2. We estimated the $\dot{V}O_2$ using the method of Åstrand and Rhyming (21), Margaria et al. (22), and Maritz et al. (23), which basically involves fitting regression lines to the individual submaximal HR–$\dot{V}O_2$ points and extrapolating to an estimated maximum HR based on age. The results of this analysis are shown in figure 1 and table 3. It should be emphasized that the estimated maximum HR was used, not the measured maximum HR.

All methods overpredict for ages 10–14 years. In each age group, the method of Maritz consistently provides the smallest standard error of prediction, although these are still quite large, ranging from .28 to .39 $l \cdot min^{-1}$.

Correlation coefficients between measured $\dot{V}O_2$max and $\dot{V}O_2$max predicted by these three methods range from .50 to .84, with the method of Maritz consistently providing the highest values. It was concluded that, in an unselected male population, the predicting of $\dot{V}O_2$max from submaximal measures of HR and $\dot{V}O_2$ by

Table 2. Descriptive data for Tecumseh male subjects

Age	*n*	$\dot{V}O_2$max (l/min)	Mean heart rate max (beats/min)
10–14 years	118	2.02	184
15–19 years	129	3.04	190
20–24 years	59	3.04	190
25–29 years	43	3.14	186
30–34 years	42	2.91	182
35–39 years	83	2.80	178

Table 3. Correlation coefficients: Measured $\dot{V}O_2$max versus predicted $\dot{V}O_2$max for three prediction methods

Age	Åstrand (21) (1 point)	Margaria (22) (2 points)	Maritz (23) (4 points)
10–39 years	.80	.76	.83
10–14 years	.82	.80	.84
15–19 years	.66	.66	.71
20–24 years	.58	.50	.70
25–29 years	.54	.58	.65
30–34 years	.64	.55	.75
35–39 years	.61	.62	.68

Source: Reproduced from Washburn and Montoye (20) with permission.

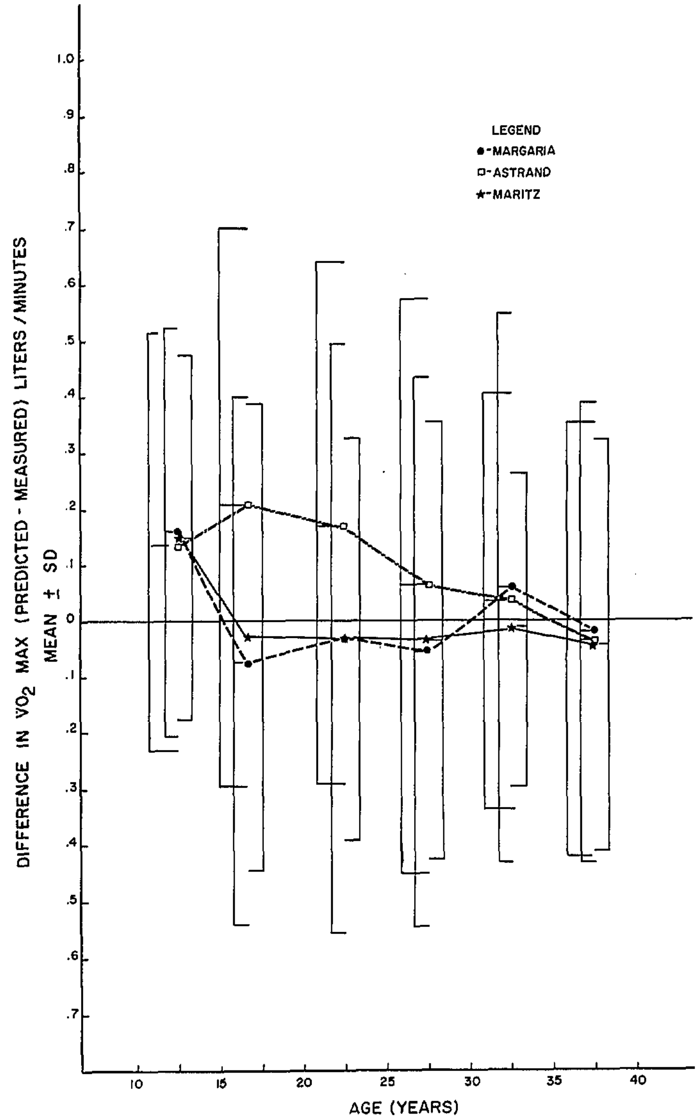


Figure 1. Difference in $\dot{V}O_2$max (predicted – measured) in l/min when predicting $\dot{V}O_2$max by methods of Åstrand-Rhyming (21), Margaria (22) and Maritz (23). Reproduced from Washburn and Montoye (20) with permission.

the methods described provide a reasonably accurate estimate of the mean $\dot{V}O_2$max for a group; however, estimates for individuals are subject to considerable error.

The fact that predicted $\dot{V}O_2$max and measured $\dot{V}O_2$max are not perfectly correlated may be the result of error in the methods of predicting $\dot{V}O_2$max from submaximal heart rate and oxygen consumption, error in the measurement of $\dot{V}O_2$max, or both. The methods of estimating $\dot{V}O_2$max are dependent on several assumptions—heart rate/oxygen uptake relationships are linear up to and including maximum values, maximum heart rate among subjects within an age group is similar, and oxygen uptake at a given work load is similar when an individual is tested several times.

Several investigations into the validity of these assumptions have been conducted (23, 24, 25). It has been shown that heart rate/oxygen uptake relationships are linear across moderate levels of work, but at higher levels the heart rate becomes asymptotic, at least in some cases. Therefore, when heart rate/oxygen uptake regressions are extrapolated to an estimated maximal heart rate, predicted $\dot{V}O_2$max will tend to be underestimated. In our data, except in age group 10–16 years, $\dot{V}O_2$max was slightly underpredicted. In 51 cases (11%), the heart rate plateaued or tended to, which might explain the slight underprediction in the means. This underprediction is consistent with other studies reported (26, 27).

Variation in maximum heart rate for a given age group has been reported to be quite small with a coefficient of variation of approximately 5% (24). We did not find this to be the case. The coefficient of variation ranged from 7% to 11%, with a value of about 10% when all subjects are combined into one group. Hence, assuming the same maximum heart rate for all subjects of the same age may be a significant source of error in predicting $\dot{V}O_2$max.

When oxygen uptake is measured, the day-to-day variability in heart rate/oxygen uptake relationships may be largely because of changes in heart rate (28). Davies (24) made a number of heart rate/oxygen uptake measurements on two experienced subjects working on a bicycle ergometer over a 6-month period. The contribution of random error due to either day-to-day physiological variation or error of measurement was assessed from the 95% confidence limit for any new observation and was ±500 ml O_2 for the two subjects. This deviation alone would result in approximately a 7% error in the estimation of $\dot{V}O_2$max from submaximal heart rate/oxygen uptake relationships.

The overprediction of $\dot{V}O_2$max shown by all regressions in age group 10–14 years may be the result of the failure to measure a true $\dot{V}O_2$max in these subjects, although each boy was encouraged to exert a maximum effort. The $\dot{V}O_2$max and maximum heart rate (2.04 ± .57 l/min, 184 ± 14 beats/min respectively) are lower than others reported in the literature (29). This points out the difficulty in measuring a true $\dot{V}O_2$max in an unselected population of youngsters ages 10–14 years. It may be that, for this group of subjects, $\dot{V}O_2$max predicted from linear regression is more indicative of their true capacity for aerobic work than the measured $\dot{V}O_2$max. This difficulty may also be true to a lesser degree in older subjects. Most other studies have utilized a highly selected population who often are used to exerting maximum effort. One might assume that this difficulty in the present study could be circumvented by testing each subject a number of times. However, when this is done, the participation rate likely would be inversely related to the number of trials, so one would end up with a highly select, nonrepresentative sample.

Table 4. Correlation coefficients: Measured $\dot{V}O_2$max versus predicted $\dot{V}O_2$max for various numbers of points in regression equation

	Number of points in regression equation			
Age	2	3	4	5
10–39 years	.83	.82	.83	.82
10–14 years	.84	.85	.84	.84
15–19 years	.71	.66	.71	.68
20–24 years	.70	.68	.70	.69
25–29 years	.62	.60	.65	.60
30–34 years	.71	.78	.75	.78
35–39 years	.69	.68	.68	.67

Source: Reproduced from Washburn and Montoye (20) with permission.

The submaximal HR/$\dot{V}O_2$ points deviate very little from a straight line. Therefore, it was of interest to determine if there is any advantage to using more than two or three points. The results of this analysis are shown in table 4, and the particular submaximal work loads corresponding to the HR/$\dot{V}O_2$ points are shown in table 5. Clearly, if the work loads are fairly strenuous and two points are selected that are reasonably widely spaced, very little is gained from using more than two points (20). In the Åstrand-Rhyming nomogram (21), only one $\dot{V}O_2$/HR point is used.

It would appear feasible to measure $\dot{V}O_2$ at submaximal loads in the NHANES. Nevertheless, it is a great deal simpler to measure only HR at various submaximal loads. The question can then be raised: How good is the estimate of $\dot{V}O_2$max when only submaximal HR and work loads are used? Using only one submaximal HR/work load point, Åstrand (30) reported that the prediction of $\dot{V}O_2$max was almost as good as when one

Table 5. Treadmill grades selected for use in regression (Speed = 3 mph)

Number of points in regression equation	Treadmill grade (%)				
	3	6	9	12	15
2		*			*
3	*		*		*
4		*	*	*	*
5	*	*	*	*	*

Source: Reproduced from Washburn and Montoye (20) with permission.

submaximal HR/$\dot{V}O_2$ point is used (r = .65 and .72, respectively, for females and .72 and .78, respectively, for males). This assumes that $\dot{V}O_2$ at submaximal work loads can be estimated fairly accurately from that work load. Again the data from Tecumseh provides an opportunity to test this assumption. The diagonal line in figures 2–5 represents the estimated $\dot{V}O_2$ calculated from the equations presented in the Guidelines of the American College of Sports Medicine (31, 32), which are as follows:

$\dot{V}O_2$ for horizontal walking

$\dot{V}O_2$ in ml/kg/min = speed in m/min × 0.1 ml
O_2/kg min perm/min
+ 3.5 ml O_2kg/min

In this formula, the resting oxygen consumption (1 MET) is assumed to be 3.5 ml kg/min. For horizontal walking at 1 meter·min^{-1}, the oxygen consumption is assumed to be 0.1 ml/kg/min plus $\dot{V}O_2$ for vertical work in grade walking.

$\dot{V}O_2$ in ml/kg/min = % grade expressed as a fraction
× speed in m/min
× 1.8 ml/kg/min

Mean measured $\dot{V}O_2$ at various work loads on the treadmill is plotted together with two standard errors of the mean in figures 2–5. Except for young children, who are known to be inefficient in walking, the fit is remarkably good. However, what is important is the standard deviation around those means, because this reflects the accuracy of the estimates. These are given in tables 6 and 7. The average standard deviation is about 1.5 to 2.0 ml/kg/min, which represents the variation in estimated oxygen uptake, and this, of course, will influence how well $\dot{V}O_2$max may be estimated from submaximal work loads and heart rate. The subjects were in a steady state in $\dot{V}O_2$ by the third minute.

Measured $\dot{V}O_2$max was available for the subjects of figures 2–5, as well as heart rates at submaximal work loads. The results of two methods for predicting $\dot{V}O_2$max from heart rate measured during submaximal exercise are shown in table 8 (columns 3 and 4). These techniques could be used when it is not possible to

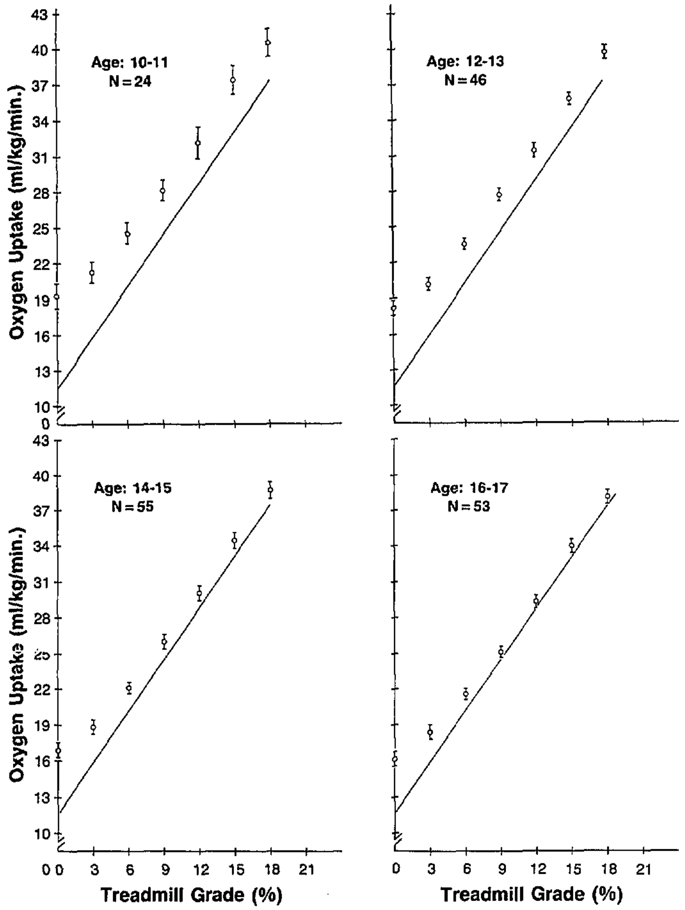


Figure 2. Measured $\dot{V}O_2$ for boys age 10–17 years, walking at 3 mph at various grades on a motor-driven treadmill. Mean value for each grade is shown (0) plus or minus 2 SE (bars). Sloping line represents $\dot{V}O_2$ as estimated by *Guidelines* formulas (32, pp. 138–140). Reproduced from Montoye et al. (31) with permission.

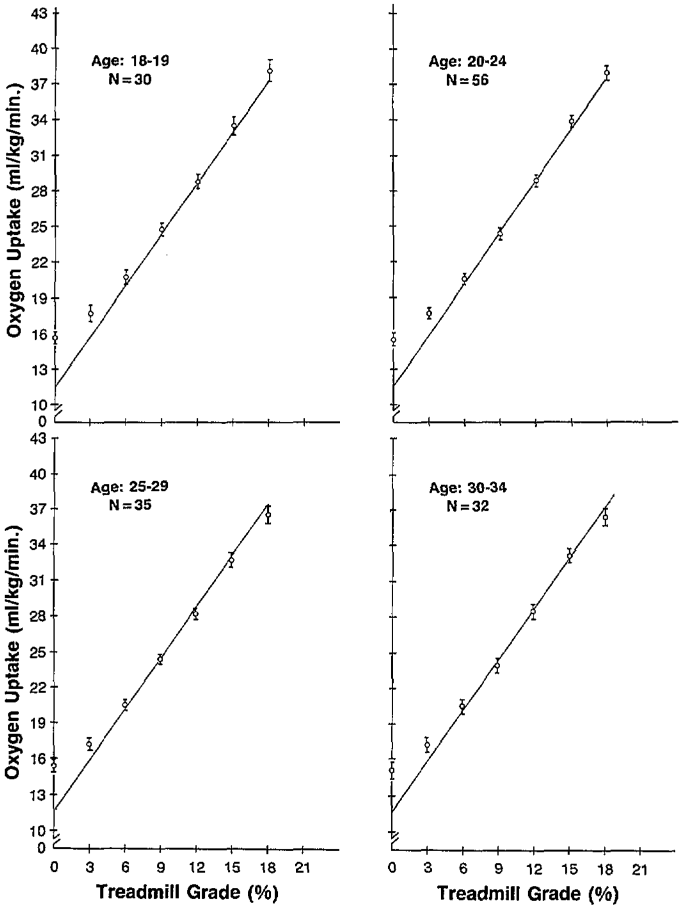


Figure 3. Measured $\dot{V}O_2$ for males age 18–34 years, walking at 3 mph at various grades on a motor-driven treadmill. Mean value for each grade is shown (0) plus or minus 2 SE (bars). Sloping line represents $\dot{V}O_2$ as estimated by *Guidelines* formulas (32, pp. 138–140). Reproduced from Montoye et al. (31) with permission.

measure $\dot{V}O_2$ or to exercise subjects to exhaustion. The first method utilized heart rate at a fixed workload. For this analysis, we recorded heart rate for each subject at 9% grade and correlated these heart rates with measured $\dot{V}O_2$max. The results are shown in table 8 (column 3). The estimates are very poor because we have little direct information that bears on $\dot{V}O_2$max. This test is comparable with tests like the Harvard Step Test, and the values of the coefficients in column 3 are comparable with what has been reported in the literature for the step test (17, 33, 34).

The second method for estimating $\dot{V}O_2$max using only submaximal heart rate involved plotting the heart rate recorded during the last 30 seconds previous to minutes 3, 4, 9, and 12 (corresponding to treadmill grades of 0%, 3%, 6%, 9%) against treadmill time. In this case, treadmill time served as an indication of work load. A straight regression line was then fitted to each individual's heart rate/time points and extrapolated to the estimated maximal heart rate for the age (220 − age). As in the previous analysis, heart rates above 170 (ages 10–19 years), 165 (ages 20–29 years), and 160 (ages 30–39 years) were not used in the analysis. These estimates of maximum treadmill times were correlated with measured $\dot{V}O_2$max, and the resulting coefficients are shown in table 8 (column 4). The values are not much different from those in column 3. This is not surprising because, again, only treadmill grade or time

Table 6. Standard deviation of measured $\dot{V}0_2$ in ml/kg/min at various grades on a treadmill-belt, males: Tecumseh Study

		Treadmill grade (%)						
Age	*n*	0	3	6	9	12	15	18
10–11 years	24	2.6	2.3	2.2	2.1	3.4	3.1	2.9
12–13 years	46	1.9	1.7	1.5	1.8	1.9	1.8	2.2
14–15 years	55	2.0	2.1	1.6	2.4	2.0	2.5	2.8
16–17 years	53	2.0	2.0	1.5	1.7	1.9	2.2	2.2
18–19 years	30	1.6	2.0	1.7	1.4	1.8	2.1	2.6
20–24 years	56	1.9	1.6	1.6	1.8	1.8	2.0	2.2
25–29 years	35	1.6	1.5	1.2	1.3	1.4	1.9	2.1
30–34 years	32	1.8	1.6	1.9	1.7	1.6	1.7	2.3
35–39 years	99	2.1	1.6	1.8	1.9	2.0	2.7	—
40–44 years	106	1.9	1.7	1.6	1.7	2.0	—	—
45–49 years	78	2.0	1.7	2.1	2.1	2.3	—	—
50–54 years	42	1.7	1.5	1.9	1.7	1.9	—	—

Note: Speed = 3 mph.

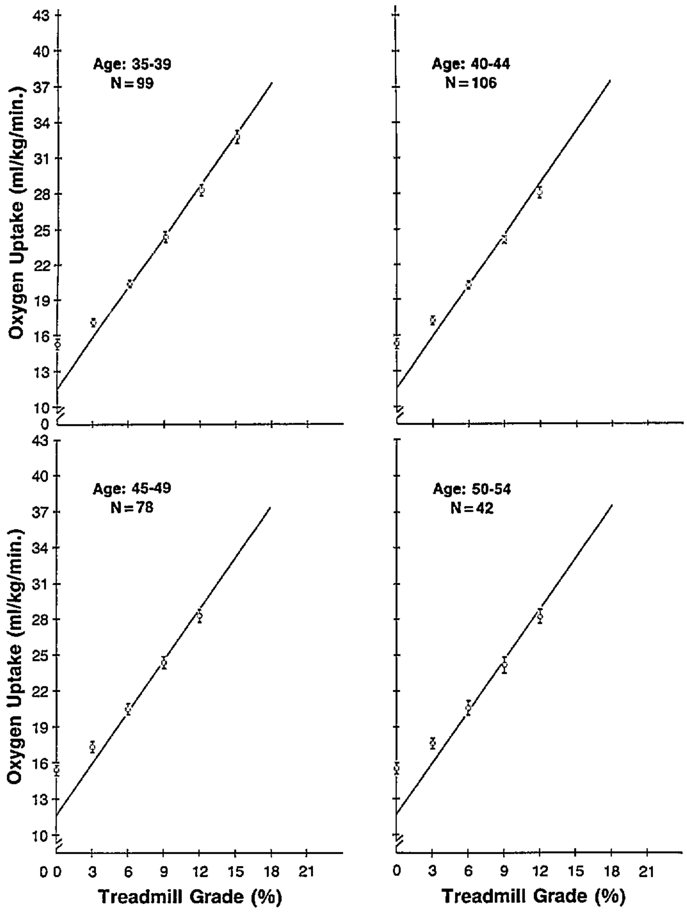


Figure 4. Measured $\dot{V}O_2$ for males age 35–54 years, walking at 3 mph at various grades on a motor-driven treadmill. Mean value for each grade is shown (0) plus or minus 2 SE (bars). Sloping line represents $\dot{V}O_2$ as estimated by *Guidelines* formulas (32, pp. 138–140). Reproduced from Montoye et al. (31) with permission.

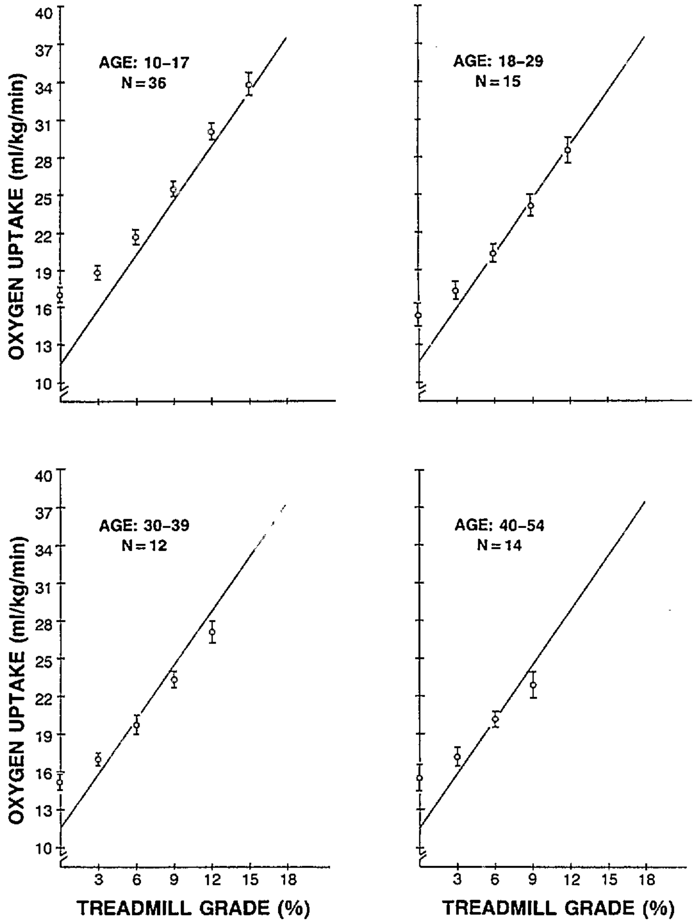


Figure 5. Measured $\dot{V}O_2$ for females age 10–54 years, walking at 3 mph at various grades on a motor-driven treadmill. Mean value for each grade is shown (0) plus or minus 2 SE (bars). Sloping line represents $\dot{V}O_2$ as estimated by *Guidelines* (32, pp. 138–140). Reproduced from Montoye et al. (31) with permission.

Table 7. Standard deviation of measured $\dot{V}O_2$ in ml/kg/min at various grades on a treadmill-belt, females: Tecumseh Study

		Treadmill grade (%)					
Age	n	0	3	6	9	12	15
10–17 years	36	2.0	1.7	1.7	2.0	2.0	2.6
18–29 years	15	1.8	1.4	1.3	1.7	2.0	—
30–39 years	12	1.1	0.8	1.3	1.1	1.5	—
40–54 years	14	1.9	1.3	1.1	2.1	—	—

Note: Speed = 3 mph.

Table 8. Correlation coefficients expressing the relationship between measured $\dot{V}O_2$max and an estimate from heart rates at submaximal work loads in 367 males

Age	n	HR at 9% treadmill grade	Estimated max treadmill time from submax HR and time [1]
10–14 years	82	−.62 (4.8)	.51 (5.2)
15–19 years	99	−.47 (6.1)	.43 (6.2)
20–24 years	48	−.16 (4.0)	.33 (3.8)
25–29 years	32	−.49 (4.1)	.72 (3.2)
30–34 years	38	−.34 (2.8)	.31 (2.8)
35–39 years	68	−.40 (4.6)	.54 (4.3)
Weighted Mean	367	−.44 (4.7)	.47 (4.7)

[1] For this analysis, submaximal heart rates and their corresponding time during the treadmill test were plotted, and individual least squares regression lines fitted to the points and extrapolated to estimate HRmax (220 − age). This provided an estimated maximum treadmill time that was correlated with measured $\dot{V}O_2$max.

Source: Reprinted by permission of the American Alliance for Health, Physical Education, Recreation and Dance, 1900 Association Drive, Reston, Va. 22091

and submaximal heart rates are available. Excluding exercise, heart rates below 120 did not improve the $\dot{V}O_2$ estimates.

This last measure is similar to estimating $\dot{V}O_2$max using the Åstrand-Rhyming nomogram with one work load and a corresponding heart rate. There have been a number of studies in which the nomogram estimate of $\dot{V}O_2$max and measured $\dot{V}O_2$max have been correlated (17). The correlation coefficients range from .33 to .87 with a mean of .60. Recently, Kasch (33) reported a similar coefficient (0.58) using 83 males, with a mean age of 48.

The heart rate at a fixed work load is also similar to the fitness measurement used in the second round of examinations in Tecumseh, which was a modification of the Harvard Step Test. This consisted of stepping onto an 8-inch bench for 3 minutes at the rate of 24 steps (4-count sequences) per minute. This is a rate of energy expenditure of roughly five times the basal metabolic rate. The electrocardiogram, from which heart rates were measured, was recorded while the subject was reclining, sitting, and standing before exercise, during exercise, and for 5 minutes after exercise while the subject was again sitting.

The test was administered to more than 4,000 males and females in the Tecumseh study population, aged 10–69 years, who were not excluded for medical reasons. Pregnant women were excluded. The subjects usually came to the clinic with their families. They were examined in the afternoon or early evening during the week or on Saturday morning. They were not required to be in a postabsorptive state. All tests were administered in an air-conditioned laboratory (72°–76°F). We did not measure $\dot{V}O_2$max in any respondents in this round of examination. Therefore, it is not possible to determine the relation between the HR response to the exercise and $\dot{V}O_2$max. However, there have been a number of reports of this relationship (17) between the Harvard Step Test and $\dot{V}O_2$max. The correlation coefficients range from .35 to .77, with a mean of .54.

This particular step test was selected for several reasons. A treadmill was not available. The test was administered in the basement of the local hospital, and a low ceiling would not have permitted the use of a treadmill anyway. This was also one of the reasons why a relatively low (8 inches) bench was employed. The other reason for this bench height was the fact that this is about the height of the average stair. It was felt that the incidents of tripping would thus be minimized. There were several significant problems with this test. The exercise stress was insufficient for many young subjects, and it overstressed some of the older and/or overweight subjects. Many subjects were unable to maintain the proper cadence. The technician frequently was faced with the subjective decision of whether the test should be terminated if she could not motivate the subject to maintain the pace. Unfortunately, a soft-hearted technician will often allow the test to be completed even though the pace is not maintained. Finally, a number of subjects, as they became tired, tripped on the bench.

A comparison of the heart rate response to this 8-inch step test was compared with serum total cholesterol (35), serum uric acid (36) and glucose tolerance (37). The same was done for measured and estimated $\dot{V}O_2$max (37, 38). A summary of the results is shown in table 9.

Strength

Strength tests are generally classified as follows.

Static strength

This may be defined as the maximum force that can be applied to a fixed object, such as a dynamometer, from a defined immobile position. This is also called isometric strength.

Dynamic strength

This is the maximum force that can be exerted once to move a load through a specified range of motion of a joint with the body in some defined position. Another term applied is isotonic strength. An example of this test is the maximum weight one can curl with the flexors of the arm.

Static muscular endurance

This is the ability to maintain a given amount of force in a fixed position for a prolonged period of time. An example is the ability to hold oneself in a chinning position on a horizontal bar.

Dynamic muscular endurance

This is the strength and endurance required to repeat a muscular effort over a period of time. Examples are the number of situps or chinups an individual can perform.

Generally, body weight is correlated with all of these strength tests, and the various kinds of strength tests are generally correlated. There is little in the way of evidence supporting a relationship between strength and health. There have been some studies suggesting abdominal strength may be related to the maintenance of a normal curvature of the spine and, hence, inversely related to incidence of low-back pain. Unfortunately, the measurement of abdominal strength in a field study is difficult unless some minimal, pass/fail test is used. Strength is obviously related to one kind of physical performance.

In Tecumseh, two static strength tests were employed. The one was a grip strength test in which a commercially available adjustable grip dynamometer was used. These instruments are easily calibrated by suspending known weights. A preliminary study was done to determine how hand size was related to grip strength performance. Hand width was found to be the most useful hand measurement. The width of the hand was measured with a simple sliding scale with three distances corresponding to three settings on the dynamometer. Finer settings were found to be unnecessary (39). Two trials were allowed with each hand, with suitable rest periods in between.

The other test was designed to measure the arm flexors, and it is illustrated in figure 6. Again, two trials were allowed. The instrument allowed for adjustments in the cable, such that the elbow joint was at 90° when force was applied.

These tests are practical in field studies, requiring no electrical power. Minimum medical screening is required to exclude subjects, for example, with an abdominal weakness (hernia).

At a minimum, the strength tests should be adjusted for body weight so subjects of various sizes can be compared. If other anthropometric measurements on the respondents are available, such as height, biacromial diameter, and biiliac (bicristal) diameter, each strength score may be adjusted for all these size measurements by multiple regression. The strength test scores may be combined and adjusted for multiple anthropometric measurements by canonical analysis as was done in the Tecumseh study (40). However, body weight is the most important single adjustment because it reflects the strength of an individual in terms of moving his/her own body weight.

The grip or arm dynamometer may utilize a strain gauge, but this requires electrical power. Spring dynamometers are commercially available, inexpensive, require no electrical power, and maintain their calibration for long periods of time and are therefore recommended. An alternative is a tensiometer, which is available commercially. This instrument was developed for measuring tension on aircraft cables and requires no electrical power. It can be used for grip or arm strength. However, the spring scales are simpler and more dependable.

Table 9. Relationship between measured $\dot{V}O_2$max and some submaximal exercise test scores to serum total cholesterol, serum uric acid, and glucose tolerance

	Serum total cholesterol		Serum uric acid		Blood glucose 1 hour after challenge	
Fitness measure	*r*	*n*	*r*	*n*	*r*	*n*
Heart rate after step test:						
Male, 10–64 years	[1].04	2385	[1].09	2298	[1].13	2347
Female, 10–64 years	[1].01	2318	[1].08	2237	[1].13	2248
$\dot{V}O_2$ measured, 10–35 years, estimated age 40–69 years: [2]						
Males	−.08	827	.07	733	−.05	733
Females	−.17	83	.22	71	−0.21	71
$\dot{V}O_2$ at heart rate of 150: [2]						
Males			.03	793	−.04	793
Females			.06	80	.00	80
Heart rate at 9% grade: [2]						
Males			−.04	793	.07	793
Females			−.07	80	.07	80
Treadmill time:						
Males			.10	516	−.04	516
Females			.15	80	−.31	80

[1] Average of 9 age group partial coefficients, effects of skinfolds removed.

[2] Effects of age, weight, and skinfolds removed.

Flexibility

A somewhat oversimplified definition of flexibility is the degree of movement (flexion and extension) possible by the articulating segments of the body around a joint. This ability is important in certain sports and, of course, defines the extent of normal movement of which a person is capable. Flexibility in the various joints is not closely correlated (41). Flexibility may be measured in the laboratory with mechanical and electrical goniometers or a flexometer, but these are more suitable for the laboratory than field studies.

A simple measurement of flexibility around the hip joint, the sit-and-reach test, is recommended for the NHANES. This test was not used in the Tecumseh study, but it probably should have been. It is safe, requires little equipment or little time to administer, and there is some evidence in the literature that poor flexibility in this joint may be related to low-back pain.

Body fatness

Skinfolds are recommended as a measure of body fatness. There are a number of calipers available commercially, and the better ones are designed to exert a pressure of 10g/mm^2. The most popular are the Harpenden and Lange calipers, but neither can be adjusted for pressure of the jaws. In that respect, the C-calipers developed at the University of Minnesota Laboratory of Physiological Hygiene are superior because the pressure can be easily adjusted. The Lange calipers were used in the second and third round of examinations in Tecumseh. In calibrating the pressure setting, it was determined that none of the three available exerted a pressure of precisely 10g/mm^2. Since that time, a Lange caliper has been modified in our laboratory, so that jaw pressure can be adjusted at will. Such a modification is recommended for the NHANES.

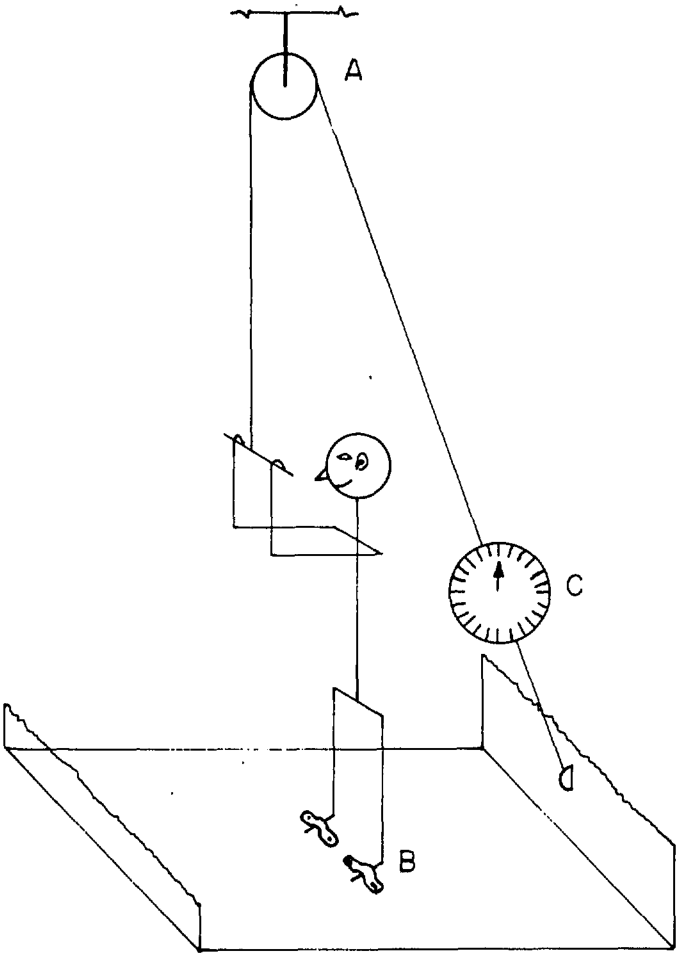


Figure 6. Illustration of apparatus for measuring arm strength in the Tecumseh Study. (*a*) Pulley; (*b*) toe straps to secure feet; (*c*) dynamometer. Reprinted by permission of the American Alliance for Health, Physical Education, Recreation and Dance, 1900 Association Drive, Reston, VA 22091.

Following is a description of the measurement of skinfolds at five sites; the last four were used in the Tecumseh study.

Biceps skinfold

This measurement is taken approximately over the center of the biceps muscle of the upper arm. The arm should be relaxed and hanging freely. The skinfold is lifted about a centimeter below the midpoint along the long axis of the muscle. The caliper, in a horizontal position, is allowed to compress the skinfold about a centimeter above the point where the thumb and finger grasp the skinfolds.

Triceps skinfold

This measurement is taken at a site halfway between the tip of acromial process and tip of the elbow. The arm should be hanging freely. The skinfold is lifted on the back of the right arm parallel to the long axis of the arm about 1 centimeter above the site. Calipers are allowed to compress skinfold about 1 centimeter below the point where the skinfold is lifted.

Subscapular skinfold

The skinfold is lifted at the tip of the right scapula on a diagonal plane about 45 degrees from the horizontal (laterally downward). The calipers are used about 1 centimeter laterally downward from this point.

Abdominal skinfold

The skinfold is lifted about 1 centimeter to the right of the umbilicus and parallel to the long axis of the body. The calipers are held so that the jaws are vertical, about 1 centimeter below the point at which the skinfold is lifted.

Waist skinfold

The skinfold is lifted just above the crest of the ilium. The fold is lifted to follow the natural diagonal line at this point (dorsally upward). The calipers are again used about 1 centimeter from the point at which the skinfold is lifted.

Reproducibility of skinfold measurement is generally good among the same subjects measured by the same investigators on more than one occasion. The coefficient of reliability is usually 0.90 or higher.

It has been observed that the skinfold measurements of various sites on the same subjects are highly correlated (r = 0.742 to 0.938) (38). We have observed much the same results in the Tecumseh Community Health Study in which the correlation coefficients between triceps and subscapular skinfolds range between 0.6 and 0.8 for age groups spanning 10 years among males and females. When the sum of four skinfolds was correlated with the sum of two, the coefficients were generally near 0.9. This means that the law of diminishing returns applies very soon, and a few skinfolds are about as good as taking measurements at many sites. Therefore, to measure skinfolds at more than three or four sites is not worth the additional time.

Table 10 shows a summary of a number of comparisons of skinfolds with hydrostatic weighing and is reproduced from Montoye (42), where the references to these studies may be found. A number of regression equations have been developed using several skinfolds to predict percent body fat. However, these have not been found to apply to all ages, both sexes, and various populations, so the sum of four skinfolds was used in Tecumseh.

Table 10. Relationship between pinch caliper skinfolds and total body density as measured by hydrostatic weighing

Correlation coefficient[1]	Number of skinfold sites	Subjects
.87	3	116 young men
.88	6 plus rel. wt.	116 young men
.74	6 plus rel. wt.	214 middle-aged men
.85	3	88 young men
.86	5 plus rel. wt.	88 young men
.69	12	94 young women
.69	12	88 middle-age women
.74	8	100 young men
.91	13	50 Chinese men
.70	13 (wt. partialled out)	29 Chinese women
.90	10	66 boys, age 9–12
.81	10	56 girls, age 9–12
.92	10	57 boys, age 13–17
.83	10	62 girls, age 13–17
.80	2	133 young men
.78	7	133 young men
.81	4 plus ht.	133 young men
.86	7	50 young men
.74	2–4	50 young women
.81–.88	2	28 young men
.74–.84	2–3	23 girls, age 14–18
.76–.85	2–3	55 males, age 17–59
.84[2]	4	60 males, age 18–34
.78[2]	4	45 women, age 22–29
.76[2]	4	48 boys, age 13–16
.78[2]	4	38 girls, age 13–16
.92	1 plus 1 girth, and 1 diameter	30 males, age 25–50
.78–.83	2 plus 2 girths	23 girls, age 14–18
.83	2 plus 1 girth, and 1 diameter	83 women, age 18–22
.89	2 plus 1 girth, and cup size	60 women, age 33–55
.7–.9	2–4	481 men and women, age 16–72 years

[1] Minus signs have been omitted. Higher density (that is, less fat) is inversely correlated with skinfold thickness.

[2] The inclusion of girths and other anthropometric measurements did not improve the relationships.

Source: Reproduced from Montoye (42) with permission.

Lessons from Tecumseh: Summary

1. It appears feasible to measure $\dot{V}O_2$ in NHANES, but the expense for equipment would be significant, and one would need test gases, chemical methods of analysis for analyzing test gases, and expertise available to maintain equipment (replacing O_2 sensors, for example).
2. Maximal tests require careful medical screening, and the response rate in females is likely to be poor.
3. With reasonable medical screening, it is safe and feasible to exercise males and females to exhaustion (ages 10–39 years), to a maximum HR of 160 (ages 40–59 years) and a HR of 150 (ages 60–69 years).
4. The recording of a readable ECG during exercise is practical and useful. A sufficient number of normal subjects showed significant ECG changes to warrant the recording of the ECG.
5. Males and females can be taught to walk on a motor-driven treadmill within 5–10 minutes, even though they had never before walked on a treadmill.
6. A steady state in $\dot{V}O_2$ is reached within 2 minutes of walking on a treadmill as long as the grade is not increased more than 3% at a stage.
7. Readable ECG during the exercise may be secured from almost 100% of the subjects.
8. The amount of repair to be expected with treadmill is minimal. In 3½ years of testing, 8 hours per day, 5 days per week, only one major breakdown occurred—a broken axle. This was repaired within 4 days. Any other malfunctions were corrected in a matter of minutes.
9. If proper precautions are taken after the exercise test (slow walking, recumbent position), no serious problems with hypotension should be expected.
10. Systolic blood pressure can be measured manually, with little difficulty during the treadmill exercise test, but measuring diastolic pressure is much more difficult. Good technicians with considerable experience can obtain reasonable results (5th phase), but in a national survey, the exercise diastolic blood pressures, if obtained manually, may be questionable. Of course, in some young subjects, if the 5th phase diastolic blood pressure is used as the criterion, sound can be heard at 0 pressure.
11. In the 8-inch bench step test used in Tecumseh, at least as much time is required to teach the skill as in walking on the treadmill. Some subjects have considerable trouble with the coordination needed.
12. Subjectivity in deciding when to terminate the step test is a problem with many subjects who seem not able to maintain the cadence.

13. No serious injuries occurred with the step test, but a number of minor ones were experienced, for example, shin abrasions.
14. Some subjects, despite urging, didn't complete the leg extension when stepping onto the bench.
15. Efficiency of walking ($\dot{V}O_2$) is not correlated with age from about age 18 years to age 65 years. Below age 18 years, the $\dot{V}O_2$ requirement is higher (that is, subjects are less efficient). In any particular age group, the variation in efficiency ($\dot{V}O_2$) at the same speed and grade is not great. The standard duration is about 1.5–2.0 ml O_2/kg/min.
16. If $\dot{V}O_2$ and heart rate are measured at a minimum of two submaximal work loads (grades) on the treadmill, $\dot{V}O_2$max can be predicted with reasonable accuracy (r = about .80). More than two points improve the prediction very little, if at all.
17. If only work load and heart rate are available at submaximal work loads, the prediction is poorer (r = about 0.45).
18. Because of the effects of ambient temperature on heart rate, exercise tests should be administered in an air-conditioned space.
19. Risk factors for coronary heart disease were significantly correlated with $\dot{V}O_2$max and submaximal treadmill measurements and with heart rate response to a step test. However, when the influence of body fatness was removed, little if any correlation remained.
20. Skinfolds can be measured quickly and with an acceptable degree of accuracy provided the technician is well trained.
21. The skinfolds caliper should be capable of being adjusted to exert a pressure of 10 g/mm^2. The calibration from the manufacturer is not dependable.
22. With minimal medical clearance, static grip and arm strength may be easily and quickly measured in males and females of all ages. Several trials should be given with suitable rest between trials.

Recommendations

Assessment of physical activity (males and females not in school)

1. The questionnaire-interview method should be employed to assess both occupational and leisure-time physical activity.
2. The most efficient technique would be to ask each respondent to complete a brief (10–15 minutes) self-administered questionnaire followed by an interview based on the questionnaire response.
3. The occupational physical activity form should provide for both employed and unemployed workers. The forms should be similar to, or a modification of, those employed in the Tecumseh study. The forms and interview should provide for multiple jobs if applicable.
4. A questionnaire and interview supplement should be developed for assessing homemaking and child care activities.
5. Leisure-time physical activities should be assessed by a questionnaire-interview similar to, or a modification of, that published by Taylor and colleagues (6). Supplements to aid in the interview should be developed from those employed in the Tecumseh study and the suggestions in the Taylor et al. article.
6. The forms should provide the opportunity to respond in such a way as to facilitate immediate coding for computer entry.
7. The computer software should provide for the following indexes:
    a. Total hours per week of occupational work, on the average.
    b. Average occupational work intensity, expressed in METS to avoid necessity of adjusting for body weight.
    c. The average number of hours of active leisure per week.
    d. List of leisure-time activities involving significant physical activity in which the respondent participated during the previous year together with estimates of frequency and duration.
    e. Average intensity expressed in METS for all physically active leisure-time activities, prorated for the past year.
    f. Estimate of total energy expended in active leisure activities (that is, c. times e.).
    g. Estimate of total energy expenditure in occupational and all leisure-time activities.

Assessment of physical fitness

(To be preceded by medical evaluation and clearance)

1. With recently developed electronic gas analyzer/computer equipment, it appears feasible to measure $\dot{V}O_2$max, and every effort should be made to do so.
2. If medical supervision is inadequate, a submaximal treadmill exercise test should be employed in which $\dot{V}O_2$, heart rate, ECG, and blood pressure are measured. The submaximal $\dot{V}O_2$ and heart rates can be used to estimate $\dot{V}O_2$max. Details are as follows:
    a. This would require only a small treadmill, 6 mph, 30-percent grade capacity, but with the capacity to handle body weights of 350 lb.
    b. The test should include a brief adaptation period (usually 5–10 minutes is sufficient), 0-percent grade, 2 mph gradually increasing to 3 mph (for subjects age 60 and over, keep at 2 mph).
    c. The belt speed should be 3 mph (2 mph for participants 60 years of age and over) throughout the test. For the first 2 minutes, the grade should be 0 percent, and the grade should increase by 2½ percent every 2 minutes.
    d. An ECG should be attached from which heart rate would be displayed and recorded during the last 30 seconds of each load. Test continues until an

established heart rate cutoff is reached or until exhaustion occurs, whichever comes first. The heart rate cutoff should be 170 for subjects ages 10–29 years; 160 for those ages 30–49 years; 150 for those ages 50 years and over. This would require about 14 minutes of exercise for ages 10–29 years; 12 minutes for ages 30 and over.

   e. Recording of ECG on tape would be advisable during the last 30 seconds at each load.

3. If it is not feasible to measure $\dot{V}O_2$, even at submaximal loads, heart rates and ECG should be measured at the various loads during the treadmill test described above. This would permit several indexes to be calculated. However, it should be recognized that these indexes are only rough estimates of $\dot{V}O_2$max.
   a. Work load at heart rate 170, 160, or 150.
   b. Extrapolation to estimate maximum work load.
   c. Estimation of $\dot{V}O_2$ at heart rates of 170, 160, or 150.
   d. Estimated max $\dot{V}O_2$ from load and estimated HR max.
4. Variable steps may be substituted for the treadmill, but this would be less desirable. If this is done, the heart rate should be displayed continuously and recorded during the last 30 seconds of each of two loads such that the heart rates are about:

   130 and 170 (ages 10–29 years)
   130 and 160 (ages 30–49 years)
   120 and 150 (ages 50 years and over)

   The step heights should be 8, 12, 16, 20, 24, 28, and 32 cm.

   Duration at each height should be 2 minutes

   Rate of stepping: 30/min (ages 10–49 years)
   24/min (ages 50 years and over)
5. Only two simple static (isometric) strength tests are recommended, both of which were used successfully in the Tecumseh study. These are tests of grip and arm strength. The design of the arm strength test should be modified slightly so the pull is against a floor mounting instead of around a pulley overhead. Grip strength should be measured by an adjustable dynamometer adjusted for hand size.
6. Flexibility in the various joints in the body are poorly correlated. Furthermore, the evidence relating poor flexibility to important health problems is sketchy. There is perhaps more scientific justification for being concerned about flexibility around the hip joint than flexibility of other joints. Because there is a simple, safe test for this, the sit-and-reach test, it is recommended.
7. Fatness should be assessed by skinfold calipers. Formulas for estimating percent of body fat have not been validated for a wide range of populations, but the sum of four or five skinfolds is useful.

References

1. Epstein, F.H., *et al.:* The Tecumseh Study: Design, progress, and perspectives. *Arch. Environ. Health,* 21:402–407, 1970.
2. Montoye, H.J.: *Physical Activity and Health: An Epidemiologic Study of an Entire Community.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975.
3. Montoye, H.J., and Taylor, H.L.: Measurement of physical activity in population studies: A review. *Human Biology,* 56:195–216, 1984.
4. Wessel, J.A., Montoye, H.J., and Mitchell, H.: Physical activity assessment: By recall record. *J. Public Health,* 55:1430–1436, 1965.
5. Reiff, G.G., Montoye, H.J., Remington, R.D., Napier, J.A., Metzner, H.L. and Epstein, F.H.: Assessment of physical activity by questionnaire and interview. Chapter 31 in *Physical Activity and the Heart,* Karvonen, M.J. and Barry, A.J., eds. Springfield, Ill.: Charles C. Thomas Co., 1967, pp. 336–371.
6. Taylor, H.L., Jacobs, D.R. Jr., Schucker, B., Knudsen, J., Leon, A.S., and Debacker, G.: A questionnaire for the assessment of leisure time physical activities. *J. Chron. Dis.,* 31:741–755, 1978.
7. Montoye, H.J., Block, W.D., Metzner, H.L., and Keller, J.B.: Habitual physical activity and serum lipids. Males, age 16–64 in a total community. *J. Chronic Dis.,* 29:697–709, 1976.
8. Montoye, H.J., Metzner, H.L., Keller, J.B., Johnson, B.C., and Epstein, F.H.: Habitual physical activity and blood pressure. *Med. and Sci. in Sports,* 4:175–181, 1972.
9. Montoye, H.J., Mikkelsen, W.M., Metzner, H.L., and Keller, J.B.: Physical activity, fatness, and serum uric acid. *J. Sports Med. and Physical Fitness,* 16:253–260, 1976.
10. Montoye, H.J., Block, W.D., Metzner, H.L., and Keller, J.B.: Habitual physical activity and glucose tolerance: Males 16–64 in total community. *Diabetes,* 16:172–176, 1977.
11. Cunningham, D.A., Montoye, H.J., Metzner, H.L., and Keller, J.B.: Active leisure time activities as related to age among males in a total population. *J. Gerontology,* 23:551–556, 1968.
12. Cunningham, D.A., Montoye, H.J., Metzner, H.L., and Keller, J.B.: Physical activity at work and active leisure as related to occupation. *Med. and Sci. in Sports,* 1:165–170, 1969.
13. Cunningham, D.A., Montoye, H.J., Metzner, H.L., and Keller, J.B.: Active leisure activities as related to occupation. *J. Leisure Research,* 2:104–111, 1970.
14. Balke, B.: *A Simple Field Test for the Assessment of Physical Fitness.* CARI Report 63–6. Oklahoma City, Civil Aeronautical Research Institute, Federal Aviation Agency, Apr. 1963.
15. Cooper, K.H.: A means of assessing maximal oxygen intake. *J. Amer. Med. Assoc.,* 203:201–204, 1968.
16. Welch, H. Endurance. In H. J. Montoye, Ed.: *An Introduction to Measurement in Physical Education,* pp. 80–90. Boston and Bacon, Inc., 1978.
17. Montoye, H.J.: Circulatory-respiratory fitness. Chapter 6 in Montoye, H.J. ed. *An Introduction to Measurement in Physical Education.* Boston: Allyn and Bacon, Inc., 1978, pp. 91–121.
18. Montoye, H.J.: Age and oxygen utilization during submaximal treadmill exercise. *J. Gerontol,* 37:396–402, 1982.
19. Guber, S., Montoye, H.J., Cunningham, D.A., and Dinka, S.: Age and physiological adjustment to continuous, graded, treadmill exercise. *Res. Quart.,* 43:175–186, 1972.
20. Washburn, R.A., and Montoye, H.J.: The validity of predicting $\dot{V}O_2$max in males age 10–39. *J. Sports Med. and Physical Fitness,* 24:41–48, 1984.
21. Åstrand, P.O., and Rhyming, I.: A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *J. Appl. Physiol.,* 7:218–221, 1954.
22. Margaria, R., Aghemo, P., and Rovelli, E.: Indirect determination of maximal O_2 consumption in man. *J. Appl. Physiol.,* 20:1070–1073, 1965.
23. Maritz, J.S., Morrison, J.F., Peter, J., Strydom, N.B., and Wyndham, C.H.: A practical method of estimating an individual's maximum oxygen intake. *Ergonomics,* 4:97–122, 1961.
24. Davies, C.T.M.: Limitations to the prediction of maximum oxygen intake from cardiac frequency measurements. *J. Appl. Physiol.,* 24:700–6, 1968.
25. Wyndham, C.H., Strydom, N.B., Maritz, J.S., Morrison, J.F., Peter, J. and Potgieter, Z.U.: Maximum oxygen intake and maximum heart rate during strenuous work. *J. Appl. Physiol.,* 14:927–36, 1959.

26. Colemen, A.E.: Validation of a submaximal test of maximal oxygen intake. *J. Sports Med.,* 16:106–11, 1976.
27. Rowell, L.B., Taylor, H.I. and Wang, Y.: Limitations to prediction of maximal oxygen intake. *J. Appl. Physiol.*, 19:919–27, 1964.
28. Montoye, H.J., Cunningham, D.A., Welch, H.G., and Epstein, F.H.: Laboratory Methods of Assessing Metabolic Capacity in a Large Epidemiological Study. *Am. J. Epidemiol.,* 91:39–47, 1970.
29. Woynarowska, B.: The validity of indirect estimations of maximal oxygen uptake in children 11–12 years of age. *Eur. J. Appl. Physiol.,* 43:19–23, 1980.
30. Åstrand, I.: Aerobic work capacity in men and women with special reference to age. *Acta Physiologica Scand.*, Vol. 49, Suppl. 169, p. 55.
31. Montoye, H.J., Ayen, T., Nagle, F., and Howley, E.T.: The oxygen requirement for horizontal and grade walking on a motor-driven treadmill. *Med. and Sci. in Sports and Ex.,* 17:640–645, 1985.
32. *Guidelines for Graded Exercise Testing and Exercise Prescription.* Second edition, Philadelphia: Lea and Febiger, 1980.
33. Kasch, F.W.: The validity of the Åstrand and Sjostrand submaximal tests. *Physician and Sportsmedicine,* 12:47–51, 1984.
34. Saris, W.H.M., de Koning, F., Elvers, J.W.H., de Boo, T., and Binkhorst, R.A.: Estimation of W170 and maximal oxygen consumption in young children by different treadmill test. In J. Ilmarinen & I. Valimake, Eds. *Children and Sports,* pp. 86–92. Berlin: Springer-Verlag, 1984.
35. Montoye, H.J., Block, W., Keller, J.B., and Willis, P.W. III.: Fitness, fatness, and serum cholesterol: An epidemiological study of an entire community. *Res. Quart.,* 47:400–408, 1976.
36. Montoye, H.J., Mikkelsen, W.H., Willis, P.W. III., and Keller, J.B.: Serum uric acid, body fatness, and heart rate response to exercise. *Med. and Sci. in Sports,* 7:233–236, 1975.
37. Montoye, H.J., Block, W., Keller, J.B., and Willis, P.W. III.: Glucose tolerance and physical fitness: An epidemiologic study in an entire community. *Europ. J. Appl. Physiol.,* 37:237–242, 1977.
38. Montoye, H.J., Mikkelsen, W.M., Block, W.D., and Gayle, R.: Relationship of oxygen uptake capacity, serum uric acid and glucose tolerance in males and females, age 10–69. *Ameri. J. Epid.,* 108:274–282, 1978.
39. Montoye, H.J., and Faulkner, J.A.: Determination of the optimum setting of an adjustable rip dynamometer. *Res. Quart.,* 35:29–36, 1964.
40. Lamphier, D.E., and Montoye, H.J.: Muscular strength and body size. *Human Biology,* 48:147–160, 1976.
41. Sigerseth, P.O.: Flexibility. Chapter 7, pp. 122–149 in Montoye, H.J., ed. *An Introduction to Measurement in Physical Education.* Boston: Allyn and Bacon, Inc., 1978.
42. Montoye, H.J.: Measurement of body fatness. Chapter 8, pp. 150–177, in Montoye, H.J., ed. *An Introduction to Measurement in Physical Education.* Boston: Allyn and Bacon, Inc., 1978.

NAME ______________________________
(Last) (Initial) (First)
I.D. Number ______________
DATE ______________

OCCUPATIONAL ACTIVITY QUESTIONNAIRE

SECTION A. WORK HISTORY

Please fill in the following information about jobs you have held in the past 12 months. Jobs include both full-time and part-time. You may have a single job or hold two or more jobs at once. If you change responsibilities with the same employer, consider it to be a change of jobs if a substantial change occurs in physical effort or tasks demanding physical activity. For example, changing from bookkeeping to construction work within the same company would be a "change of job". If you have held more than 3 jobs in the past 12 months, please answer for the 3 jobs which required the most physical effort.

1. Have you been employed at a job for pay in the past 12 months? No ___ ---> YOU ARE DONE WITH THIS QUESTIONNAIRE.
Yes ___ ---> GO TO QUESTION 2

2. For how many months were you employed in the past 12 months? ________ months

3. How many weeks were you on vacation (not working in any job)? ________ weeks.

Work History for the Past 12 Months

	Job 1	Job 2	Job 3
4. Occupation:			
5. Name of Business:			
6. How many hours per week do/did you usually work?			
7. List the months in the past year in which you did this work: (i.e, Feb-June)			

EPI PACT Version 2-5-87

SECTION B. TRANSPORTATION TO AND FROM WORK.

1. Check the months you used particular kinds of transportation to and from each job each of your jobs? In a particular month if you used more than one mode of transportation, indicate the approximate number of days you used each.

Mode of Transportation	Job 1												Job 2												Job 3											
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Auto, bus or subway																																				
Bike																																				
Walk																																				

What is the approximate distance (round trip) that you walked or biked on the way from home to work and back home again for each of your jobs?

	JOB 1	JOB 2	JOB 3
2. Distance walked or biked:	in miles ______	in miles ______	in miles ______
or	in blocks ______	in blocks ______	in blocks ______

SECTION C. PHYSICAL ACTIVITY AT WORK

We are interested in the level of activity that people do in different occupations. Think about the activities you do in a typical work week during the time you have worked in the past 12 months.

First, read each of the following categories and mark whether or not you do them in a typical week.

Then go back through the list and record, for each activity marked 'Yes', the number of hours per week you spend on each activity. The number of hours in each activity should add to the number of hours you gave for each job in part A, question 6.

	Job 1 ____ hrs/wk	Job 2 ____ hrs/wk	Job 3 ____ hrs/wk
1. Sitting - Light Work . desk work . driving a car or pick-up truck . using handtools . light assembly or repair	__ yes--> ____ __ no	__ yes--> ____ __ no	__ yes--> ____ __ no
2. Sitting - Moderate Work . working heavy levers . riding mower or forklift . crane operation . operating a heavy truck, frequently getting on and off rig and using arms	__ yes--> ____ __ no	__ yes--> ____ __ no	__ yes--> ____ __ no
3. Standing - Light Work . bartending . store clerk behind a counter . assembling light machine parts at own pace . using hand tools . car wash attendant	__ yes--> ____ __ no	__ yes--> ____ __ no	__ yes--> ____ __ no
4. Standing - Light-Moderate Work . assembling or repairing heavy machine parts . light welding . stocking shelves . packing or unpacking small objects . sanding, scrubbing, polishing floors with power equipment.	__ yes--> ____ __ no	__ yes--> ____ __ no	__ yes--> ____ __ no

	Job 1 ____ hrs/wk	Job 2 ____ hrs/wk	Job 3 ____ hrs/wk
5. Standing - Moderate Work, includes: . assembling or work with light machine parts at fast rate on assembly line . lifting up to 45 pounds every 5 min. or so for a few seconds at a time. . cranking up dollies . hitching trailers . operating large levers, jacks . pulling on wires . twisting cables . jerking on ropes, cables, etc., such as rewiring houses.	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
6. Standing - Moderate-Heavy Work includes: . lifting more than 45 pounds every 5 min. or so for a few seconds . masonary . painting . paper-hanging	__yes--> __no	__yes-->____ __no	__yes-->____ __no
7. Walking (don't include from home to work) This includes walking without carrying anything except a brief case or something similar) . between buildings . in hallways . roving store clerk	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
8. Walking up and down stairs (give number of flights of stairs per day instead of hours; about 10 stairs per flight; up and down counts as 1 flight). Each flight generally takes about 30 seconds. If you climb up and down more than 20 flights in the usual day, include the time under #7 above. If you usually climb fewer flights than 20 in a day, record the number of flights, but igonore the time.	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
9. Walking - carrying something: . carrying trays, dishes . gas station mechanic work --- changing tires, wrecker work, etc.	__yes--> __no	__yes--> __no	__yes-->____ __no

		Job 1 hrs/wk	Job 2 hrs/wk	Job 3 hrs/wk
10.	Lifting and carrying objects less than 45 lbs. for more than a few seconds at a time	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
11.	General Heavy Industrial Labor: . handyman	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
12.	Carpentry . chopping wood, using hand axe or saw	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
13.	Standing or walking, lifting and carrying objects 45-64 lbs for more than a few more than a few seconds at a time	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
14.	Standing or walking, lifting and carrying objects 65-84 lbs for more than a few seconds at a time	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
15.	Standing or walking, lifting and carrying objects 85-100 pounds or more for more than a few seconds at a time	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
16.	Using heavy power tools such as pneumatic tools (jackhammers, drills, tampers)	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
17.	Using heavy tools other than power tools such as shovel, pick, tunnel bar, spade	__yes-->____	__yes-->____	__yes>____

	____Job 1____ hrs/wk	____Job 2____ hrs/wk	____Job 3____ hrs/wk
18. Moving, pushing heavy objects, 75 lbs or more . desks, file cabinets, heavy stock, furniture . moving van work	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
19. Other Activities (specify)			
________________________________	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
________________________________	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
________________________________	__yes-->____ __no	__yes-->____ __no	__yes-->____ __no
Total the number of hours for each job. This should equal the totals on Section A.	____________	____________	____________

/MB
2-5-87

OCCUPATIONAL ACTIVITY QUESTIONNAIRE

Notes for Interviewers

This questionnaire is designed to be self-administered. However, when the questionnaire is picked up, the interviewer should check it over very carefully to see that there were no omissions, that the occupation(s) are clearly understood, that the answers seem reasonable and that the additional tasks (No. 19) are clearly understood by the interviewer because metabolic rates will have to be designated for these activities.

Section A

A question may be raised about vacation time. If the total vacation time was two weeks or less, ignore this information. If there were more than two weeks of vacation, this should be rounded off to the closest month and subtracted from the pro rate for the year. The same should be done for transportation to and from work. Lay offs should be reflected in question 2 in this section.

Section B

For some persons, transportation to and from work will be the only sources of physical activity. Make certain the distances are recorded as round trip distances. The respondent may indicate he/she travels from one job to the other without going home. In these instances, if the transportation is by auto, bus, or subway, ignore the information. If the transportation is by walking or biking, ask sufficient questions to be able to calculate and record the actual distances travelled.

Section C

#1 and #2:	Frequently truck drivers load and unload cargo. Make certain this is reflected in the responses. If there is essentially no loading or unloading, only #1 should be completed insofar as driving a truck is concerned. If there is loading and unloading, this should be reflected in #10, #13, #14, #15 or #18.
#7:	If there is difficulty estimating time spent walking, it might be helpful to estimate distance per week.
#8:	There is more difficulty estimating time spent ascending or descending stairs, so respondent is asked to indicate the approximate number of flights per day. Make certain he/she indicates this rather than hours. One flight is equivalent to about 10 steps and includes ascending and descending as one flight. Include only work days.
#16 and #17:	Number 16 refers to use of heavy power tools of various sorts. Number 17 involves no power equipment.

Appendix B

WORK SHEET: PHYSICAL ACTIVITY CODE

(1) OCCUPATION

Occupation: Farm Machinery Dealer

Activity	$\frac{WMR}{BMR}$		Hrs./Week		$\frac{WMR}{BMR}$ × Hrs./Week	
Sitting	1.5	×	19	=	28.5	
Standing (light work)	3.0	×	19	=	57.0	
Walking	4.0	×	19	=	76.0	
			57		161.5	2.83
			(Total A)		(Total B)	Ave. WMR/BMR for Occ. (B/A)

(2) ACTIVE LEISURE

Activity	$\frac{WMR}{BMR}$		Hrs./Week		$\frac{WMR}{BMR}$ × Hrs./Week	
Rowing (pleasure)	3.5	×	0.2	=	0.7	
Boat Fishing	3.0	×	0.9	=	2.7	
Hunting	6.0	×	0.4	=	2.4	
			1.5		5.8	3.87
			(Total C)		(Total D)	Ave. WMR/BMR for Active Leisure (D/C)

(3) QUIET LEISURE*

Activity (constant for all)	$\frac{WMR}{BMR}$		Hrs./Week		$\frac{WMR}{BMR}$ × Hrs./Week
Reading, TV, etc.	1.8	×	32.5	=	58.5
			(Total E) 91 hrs.-A-C		(Total F)

(4) TOTAL LEISURE

34.0	=	64.3	1.89
(Total G) (C+E)		(Total H) (D+F)	Ave. WMR/BMR Total Leisure (H/G)

(5) MEALS - SLEEPING

Activity (constant for all)	$\frac{WMR}{BMR}$		Hrs./Week		$\frac{WMR}{BMR}$ × Hrs./Week
Meals	1.8	×	21.0	=	37.8
Sleep	1.0	×	56.0	=	56.0
			77.0		93.8
			(Total I)		(Total J)

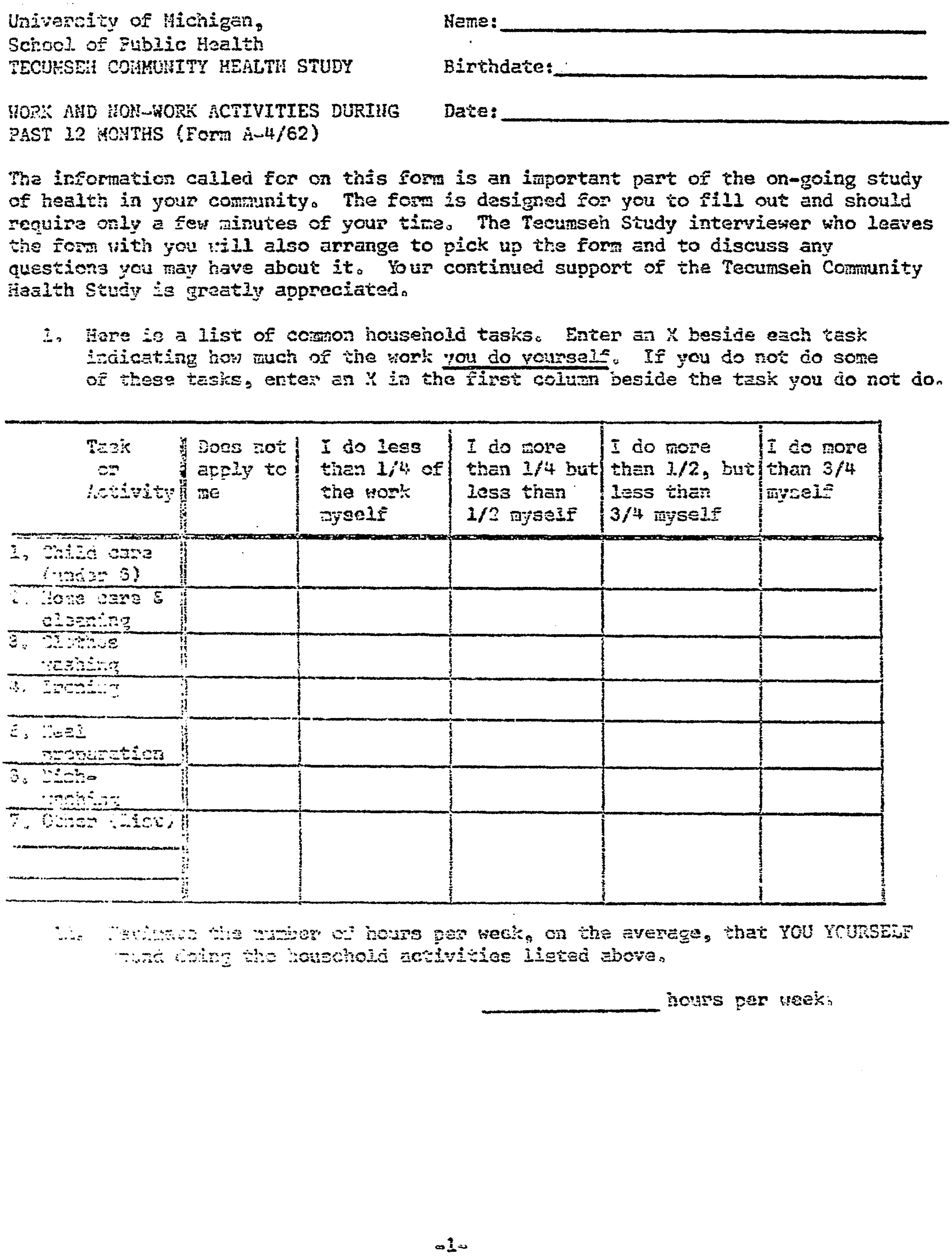
(6) Total (Occ.+Leisures+Means and Sleep Constant) = 319.6 (B+H+J)

(7) Average WMR/BMR for Occupation and Leisure = 1.90 B+H+J/168

---

*Total hours per week equals 7 days × 24 hours = 168 hours. It is assumed each subject averages 8 hours per day sleeping and 3 hours per day eating, or, 77 hours per week. The remaining 91 hours (168-77) are utilized to obtain Total E "quiet leisure hours", calculated by subtracting occupation and active leisure hours from 91. The figure in (7) may be converted to kcal./24 hrs. by multiplying the figure in (7) by 24 and by the respondents weight in kg.

Appendix C



University of Michigan,
School of Public Health
TECUMSEH COMMUNITY HEALTH STUDY

WORK AND NON-WORK ACTIVITIES DURING
PAST 12 MONTHS (Form A-4/62)

Name: ______________________________

Birthdate: ______________________________

Date: ______________________________

The information called for on this form is an important part of the on-going study of health in your community. The form is designed for you to fill out and should require only a few minutes of your time. The Tecumseh Study interviewer who leaves the form with you will also arrange to pick up the form and to discuss any questions you may have about it. Your continued support of the Tecumseh Community Health Study is greatly appreciated.

I. Here is a list of common household tasks. Enter an X beside each task indicating how much of the work <u>you do yourself</u>. If you do not do some of these tasks, enter an X in the first column beside the task you do not do.

Task or Activity	Does not apply to me	I do less than 1/4 of the work myself	I do more than 1/4 but less than 1/2 myself	I do more than 1/2, but less than 3/4 myself	I do more than 3/4 myself
1. Child care (under 6)					
2. Home care & cleaning					
3. Clothes washing					
4. Ironing					
5. Meal preparation					
6. Dish-washing					
7. Other (List) ______________ ______________					

II. Estimate the number of hours per week, on the average, that YOU YOURSELF spend doing the household activities listed above.

______________ hours per week.

2. If you are responsible for the care of any children or adults in your household, check in the appropriate column the age and number of children or adults that you care for.

Children

NONE ☐

Number of Children	Age of children under 2	2-5	6-11	12-15	16-18
1					
2					
3					
4					
5					
If more than 5 how many? ______					

Adults (Do not count yourself)

☐ None
☐ One
☐ Two
☐ Three
☐ Four
☐ More than four
(How many?) ______

3. Of all your household tasks, which one requires the most physical effort on your part? Please describe this effort in a short sentence or two.

____

____

4. Please check all of the household appliances listed that you use in your household tasks. (Do not include stove, iron, etc.).

☐ 1. Clothes washer
☐ 2. Clothes dryer
☐ 3. Dishwasher
☐ 4. Vacuum cleaner
☐ 5. Vacuum cleaner attachments
☐ 6. Electric floor waxer
☐ 7. Electric floor scrubber
☐ 8. Clothes mangle
☐ 9. Other (list if any)
____

5. Type of dwelling

☐ 1. House
☐ 2. Apartment
☐ 3. Trailer
☐ 4. Other (describe)

5A Number of bedrooms

☐ None
☐ One
☐ Two
☐ Three
☐ More than three
(How Many?) ______

6. Does your home have one or more flights of stairs?

☐ Yes ☐ No

6A If yes, approximately how many stairs? ______

6B. How many times a day, on the average, do you climb the stairs?
______ times per day.

Chapter **16**

Fitness and Activity Assessments Among U.S. Army Populations: Implications for NCHS General Population Surveys

James A. Vogel, Ph.D.
U.S. Army Research Institute of Environmental Medicine
Natick, Massachusetts

Introduction

Background

The military forces of this country represent the largest population for which physical fitness is routinely assessed. Field measures of aerobic power, muscle strength, and muscular endurance, along with body weight (and in some cases body fat), are measured twice yearly in the U.S. Army through age 60. Field measures are defined as those conducted by Army units without the aid of equipment or indoor facilities. The purpose of these periodic fitness evaluations is both as an indicator of the adequacy of training to meet performance goals and as a motivator to individuals to train and improve their fitness levels.

In addition to these periodic field measures, extensive population surveys of laboratory-measured fitness and activity assessment have been made in a wide variety of Army units over the past 10 years by the Exercise Physiology Division, U.S. Army Research Institute of Environmental Medicine. These assessments have been part of an ongoing research program to study factors influencing fitness in the Army. This chapter presents a description of the survey methods and sample data from both approaches.

Fitness components

Fitness components of concern to the Army include aerobic power, muscle strength, and strength endurance (anaerobic power). These components were selected to reflect the three categories of muscular contraction based on their respective sources and pathways of energy generation as illustrated in figure 1. Thus aerobic power represents the capacity for exercise that derives its energy primarily through the citric acid cycle and respiratory chain. Muscular power (or muscular endurance), on the other hand, represents activities that derive energy primarily, although not exclusively, from

---

Note: Human subjects participated in these studies after giving their free and informed voluntary consent. Investigators adhered to AR 70-25 and USAMRDC Regulation 70-25 on Use of Volunteers in Research.

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

the anaerobic process of glycolyses, and muscle strength represents activities that derive energy primarily from stored phosphagens. Motor fitness aspects, such as agility, flexibility, and coordination are not typically assessed and will not be included in this discussion.

As an adjunct to fitness concerns, body weight and body composition are also included as fitness components in the military. Body weight and fat standards were originally part of the fitness program and fitness regulations. Because of a considerable increase in emphasis in this area, they are now considered under separate regulations.

Objective

The purpose of this chapter is to document the methodologies used by the Army to evaluate fitness both in the field and in the laboratory and to present a compilation of available data from both sources. Body composition procedures and example data are also presented. Finally, a number of physical activity assessment questionnaires that have been employed in Army studies are included.

Field assessment of fitness

Background

The Army has conducted periodic assessments of fitness of its soldiers since World War II although the specific test events, standards, applicable population, and frequency have varied over the years. Prior to 1980, a five-event fitness test was administered for men that included an inverted crawl, run-dodge-jump, horizontal ladder, bent-leg situp, and 2-mile run. For women soldiers, events consisted of an 80-meter shuttle run, modified pushups, run-dodge-jump, modified situps and a 1-mile run.

In an attempt to improve and streamline fitness testing, a new testing program was implemented in 1980 that dropped events that did not evaluate physical fitness capacity per se (as opposed to motor fitness) and applied events uniformly to both men and women. With emphasis on eliminating equipment and enhancing the objectivity of scoring, three events were chosen: 2-mile run for time, maximal number of extended leg pushups, and maximal number of bent-knee situps that can be performed in a 2-minute period.

Although the 2-mile run for time can be considered a good estimation of aerobic power (1), pushups and situps leave much to be desired in covering the remaining components of strength and strength endurance. In fact, both of these tests must be considered strength endurance events that are limited only to the shoulder and abdominal muscles. Neither of these events correlate well with common soldiering tasks (2) but nevertheless serve the puporse of stimulating participation in physical training programs.

Methods

The three-event fitness test, originally called the Army Physical Readiness Test, is now referred to as the

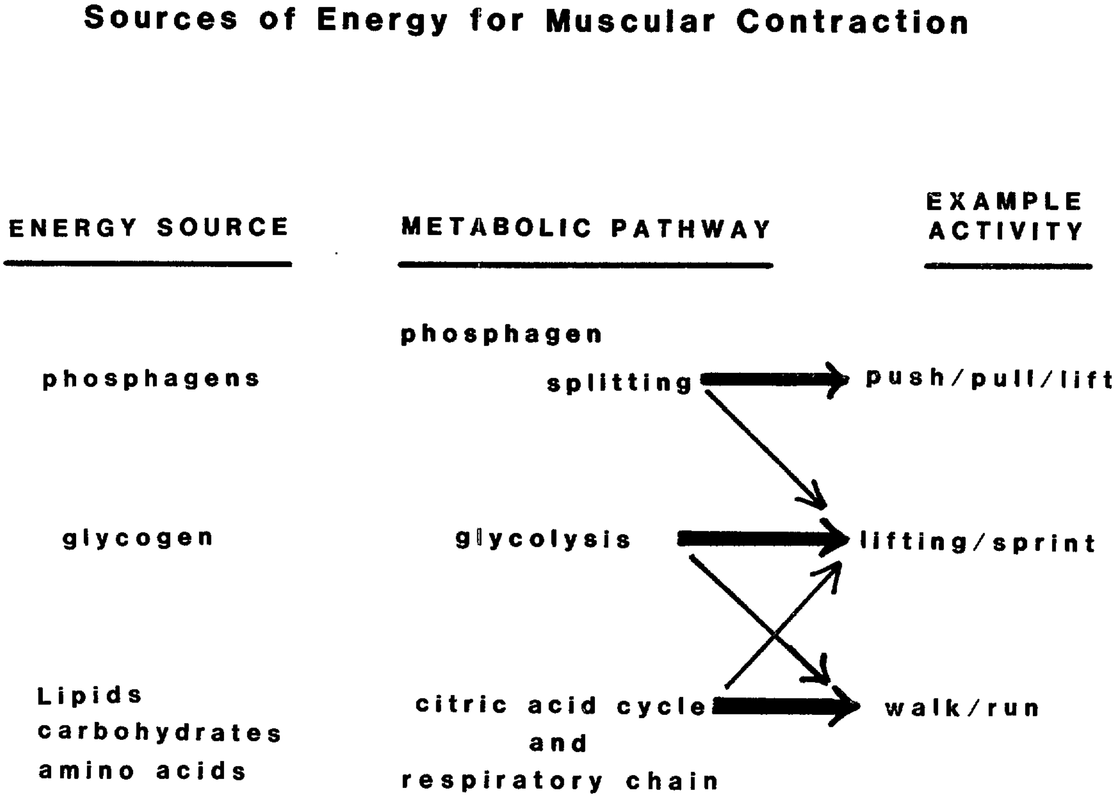
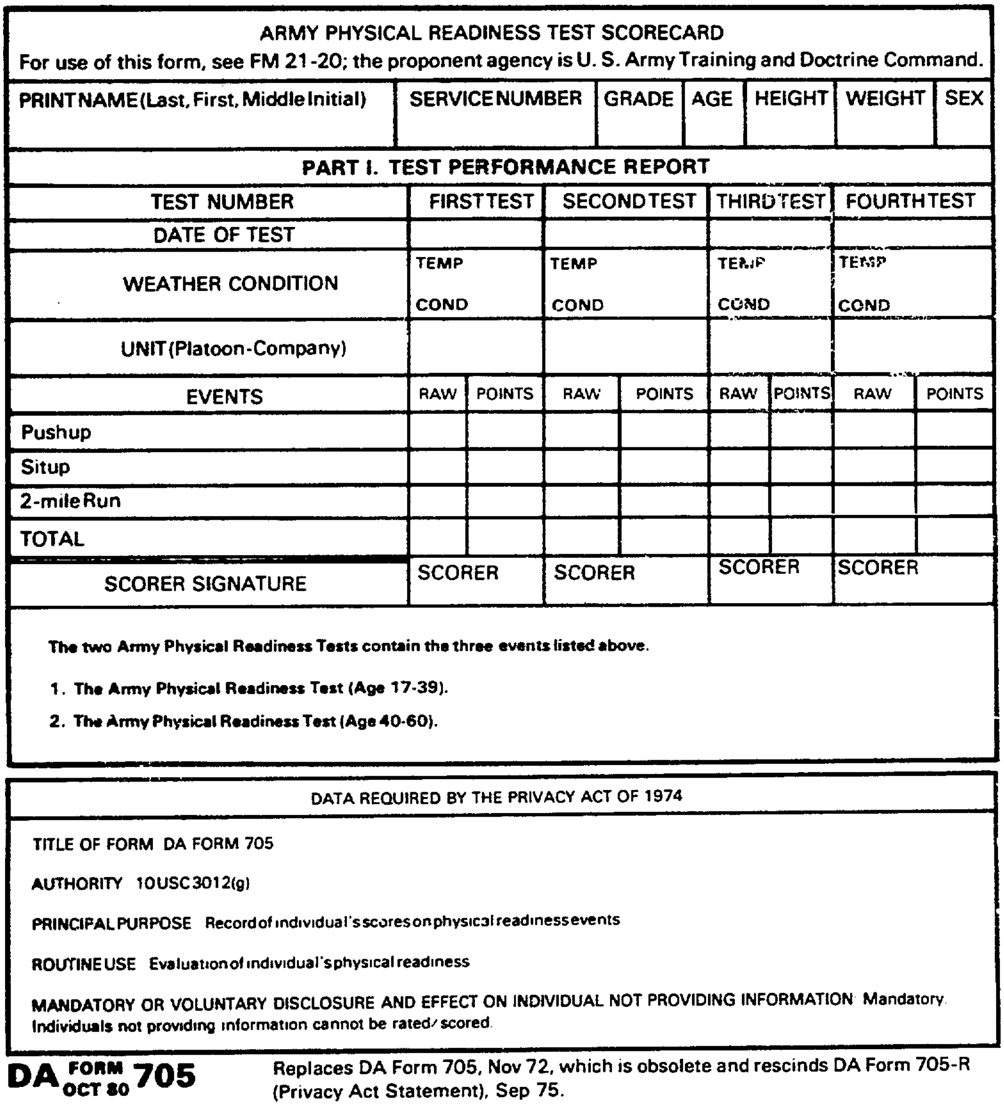


Figure 1. Energy sources and pathways for muscular contraction.



ARMY PHYSICAL READINESS TEST SCORECARD

For use of this form, see FM 21-20; the proponent agency is U. S. Army Training and Doctrine Command.

PRINT NAME (Last, First, Middle Initial)	SERVICE NUMBER	GRADE	AGE	HEIGHT	WEIGHT	SEX

PART I. TEST PERFORMANCE REPORT

TEST NUMBER	FIRST TEST		SECOND TEST		THIRD TEST		FOURTH TEST	
DATE OF TEST								
WEATHER CONDITION	TEMP COND		TEMP COND		TEMP COND		TEMP COND	
UNIT (Platoon-Company)								
EVENTS	RAW	POINTS	RAW	POINTS	RAW	POINTS	RAW	POINTS
Pushup								
Situp								
2-mile Run								
TOTAL								
SCORER SIGNATURE	SCORER		SCORER		SCORER		SCORER	

The two Army Physical Readiness Tests contain the three events listed above.

1. **The Army Physical Readiness Test (Age 17-39).**
2. **The Army Physical Readiness Test (Age 40-60).**

DATA REQUIRED BY THE PRIVACY ACT OF 1974

TITLE OF FORM DA FORM 705

AUTHORITY 10USC3012(g)

PRINCIPAL PURPOSE Record of individual's scores on physical readiness events

ROUTINE USE Evaluation of individual's physical readiness

MANDATORY OR VOLUNTARY DISCLOSURE AND EFFECT ON INDIVIDUAL NOT PROVIDING INFORMATION Mandatory. Individuals not providing information cannot be rated/scored.

DA FORM OCT 80 705 Replaces DA Form 705, Nov 72, which is obsolete and rescinds DA Form 705-R (Privacy Act Statement), Sep 75.

Figure 2. Data card for recording Army's physical fitness test scores.

Army Physical Fitness Test (APFT). It is required to be taken twice yearly through age 60. Personnel 40 years of age and older must receive a medical clearance to participate in training and testing that consist of a physical examination and coronary disease risk assessment (3).

The test is administered by the soldier's unit or organization and is recorded on a score card (figure 2) that is retained in the unit's administrative files. Raw scores (time for run and number of pushups and situps) are converted into a relative score. The soldier must achieve the minimum standard in each event that represents 60 points and must also achieve a total of 180 points overall. Minimum passing and maximum score standards have been established (table 1) and are currently being upgraded (table 2). These standards are adjusted for gender and age, regardless of occupation or

Table 1. Army Physical Readiness Test minimum standards
[Revised after 1 Oct 1986, see table 3]

	2-mile run time		Pushups		Situps	
Age	Males	Females	Males	Females	Males	Females
17–25 years	17:55	22:14	40	16	40	27
26–30 years	18:30	22:29	38	15	38	25
31–35 years	19:10	24:04	33	14	36	23
36–39 years	19:35	25:34	32	13	34	21
40–45 years	20:00	26:00	20	10	25	15
46–50 years	21:00	27:00	20	10	25	15
51–55 years	22:00	28:00	15	8	20	10
56–60 years	23:00	29:00	15	8	20	10

Table 2. New Army Physical Fitness Test minimum standards
[As of 1 Oct 1986]

Age	2-mile run time		Pushups		Situps	
	Males	Females	Males	Females	Males	Females
17–21 years	15:45	18:45	42	18	52	50
22–26 years	16:36	19:36	40	16	47	45
27–31 years	17:18	21:00	38	15	42	40
32–36 years	18:00	22:36	33	14	38	35
37–41 years	18:42	23:36	32	13	33	30
42–46 years	19:12	24:00	26	12	29	27
47–51 years	19:36	24:30	22	10	27	24
52 years and over	20:00	25:00	16	9	26	22

Table 3. Army physical fitness test scores before and after basic initial entry training

Test event	Time	Male			Female		
		n	Mean	SD	*n*	Mean	SD
Pushup	Prebasic	791	27	8	529	8	8
(number/2 min)	Postbasic	814	44	11	765	22	9
Situp	Prebasic	791	42	13	529	38	13
(number/2 min)	Postbasic	815	60	9	765	56	10
Run	Prebasic [1]	751	7:25	1:58	450	9:38	1:49
(min)	Postbasic [2]	812	14:06	1:12	757	17:47	:57

[1] 1 mile.
[2] 2 miles.
Source: Myers et al. (2).

assignment. Special units and schools may impose higher standards. Failure to meet the minimum standard for each event and the total score require repeat testing after a suitable period of remedial training. APFT scores become a part of the individual's annual performance rating.

The procedure for each of the three events is as follows:

Pushup: Start from a front-leaning rest position with hands and feet comfortably apart, arms extended, body in a straight line. Body is then lowered by bending the elbows to a point where the upper arms are parallel to the ground and then return to the starting position.

Situp: Start by lying on back with knees bent at 90-degree angle, ankles held by another individual, hands interlocked behind head. Upper body is raised forward to and past the vertical position and then lowered back to the ground to the starting position.

2-mile run: Time is measured that is required to run a measured 2-mile course.

Findings

Even though the APFT is administered twice yearly to some 700,000 soldiers, these data are not gathered centrally, and therefore it is not possible to summarize them. Large population studies have been conducted that have included the collection of APFT results, and some samples of these are given here. Table 3 presents APFT scores from a sample of basic initial entry trainees collected in 1982 at the Fort Jackson Training Center. Values are given for both men and women before and after the 7 weeks of recruit training. Marked improvement was evident in all events.

Tables 4–6 illustrate APFT data from a cross-section of soldiers assigned to a variety of units at a large Army

Table 4. Army physical fitness test pushup scores for soldiers assigned to a variety of units at a large Army post

Category	Male				Female			
	n	Mean	SD	Range	*n*	Mean	SD	Range
Total	1,014	50	15	13–99	255	32	12	10–78
17–20 years	154	56	11	20–79	60	34	11	17–64
21–27 years	368	56	13	13–99	146	33	12	15–78
28–39 years	286	50	12	15–99	48	29	11	12–69
40 years and over	206	36	16	15–80	—	—	—	—
Black	238	53	14	13–99	89	32	11	12–69
Hispanic	120	55	13	20–85	18	35	8	20–46
White	620	48	16	15–99	141	32	12	10–78

Source: Fitzgerald (4).

Table 5. Army physical fitness test situp scores for soldiers assigned to a variety of units at a large Army post

	Male				Female			
Category	*n*	Mean	SD	Range	*n*	Mean	SD	Range
Total	1,014	51	14	12–99	255	51	13	16–86
17–20 years	154	59	10	35–79	60	55	12	30–81
21–27 years	366	57	11	28–99	146	52	12	27–86
28–39 years	287	50	12	12–84	48	43	11	25–74
40 years and over	207	43	17	20–99	—	—	—	—
Black	239	57	12	25–99	89	51	12	26–75
Hispanic	118	55	12	33–84	18	55	10	40–71
White	621	50	15	12–99	141	50	13	16–86

Source: Fitzgerald (4).

Table 6. Army physical fitness test 2-mile run scores for soldiers assigned to a variety of units at an Army post [min:sec]

	Male				Female			
Category	*n*	Mean	SD	Range	*n*	Mean	SD	Range
Total	1,006	14:55	2:05	10:06–24:00	254	17:45	2:21	12:30–27:24
17–20 years	152	13:50	1:38	10:06–17:30	59	17:01	2:23	12:30–23:12
21–27 years	363	14:12	1:48	10:06–19:18	146	17:53	2:01	13:30–23:12
28–39 years	287	15:40	1:59	10:12–23:00	49	18:16	3:01	13:00–27:24
40 years and over	206	15:53	2:06	11:00–24:00	—	—	—	—
Black	236	14:27	2:06	10:12–23:00	90	17:38	1:58	13:00–23:12
Hispanic	116	14:25	1:52	10:06–20:00	18	16:45	1:54	12:30–21:06
White	618	15:09	2:04	10:12–24:00	139	17:58	2:33	13:00–27:24

base in 1984 (4). Values are tabulated according to gender, age, and ethnicity. Percentile values for the entire sample are presented in table 7.

Laboratory surveys of fitness

Background

Since the inception of the fitness research program at the U.S. Army Research Institute of Environmental Medicine in 1974, a number of Army populations have been sampled for physical fitness using standardized laboratory procedures. Most of these studies have been documented in individual reports (5–9) and have been summarized in two recent publications (10, 11). Some of the larger and more representative surveys are presented in this chapter.

Methods

Aerobic power

Our laboratory typically surveys aerobic fitness by directly measuring maximal O_2 uptake. The interrupted-load, uphill treadmill running procedure is used as originally described by Taylor et al. (12) and Mitchell et al. (13) and illustrated in figure 3. Although only 12–15 tests can be performed each day per treadmill, we believe that the much greater reliability, reproducibility, and consistency justifies this more elaborate procedure, compared with $\dot{V}O_2$max prediction procedures, especially when equipment, personnel, and space are not overriding constraints. The limitations of prediction techniques are well known (14, 15). They tend to give poor estimates at the extremes, and they are affected by

Table 7. Percentile values for Army physical fitness test scores for soldiers assigned to a variety of units at a large Army post

	Male			Female		
Percentile	Pushups	Situps	2-mile run	Pushups	Situps	2-mile run
5	72	72	11:30	50	70	14:05
20	65	66	13:06	40	62	16:00
35	55	60	14:00	38	56	17:00
50	50	51	15:00	30	50	17:30
65	45	46	15:30	26	45	18:24
80	40	40	17:00	22	40	19:34
95	20	25	18:18	17	30	21:26

Source: Fitzgerald (4).

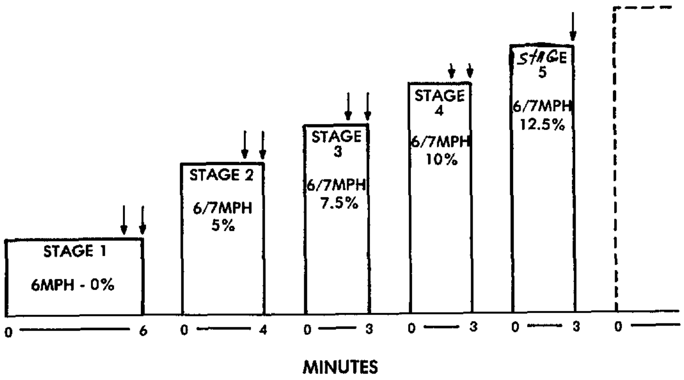


Figure 3. Procedural protocol employed for the actual measurement of maximal oxygen uptake on the treadmill. Arrows indicate points of gas collection. Speeds are those used for men and are reduced by 1 mph for women.

other factors that influence heart rate. Direct measures of $\dot{V}O_2$max permit the following of individuals through training programs or interventions and the ability to compare groups between studies, all of which are subject to considerable error when using predictive methods. We have also chosen the treadmill running mode of exercise rather than the cycle ergometer for the obvious reasons of application to soldiering tasks, because it avoids the possibility of being compounded by local muscle fatigue, and because it results in higher values.

In our procedure, female subjects perform an initial warm-up load at 5 mph (134 m/min) for 6 minutes. This is followed by 3 or 4 additional runs (all separated by 5 minutes of rest) of 4 or 3 minutes in duration at either 5 or 6 mph, with increasing increments in grades of 2.5% until a leveling off of $\dot{V}O_2$ is achieved. The leveling-off criterion is defined as an increase of less than 0.15 $l{\cdot}min^{-1}$ per 2.5% grade increase. Male subjects follow the same protocol, except that they begin at 6 mph (161 m/min) followed by speeds of 6 or 7 mph for subsequent incremental loads. Expired gas is collected for analysis during the final minute of each load. Subjects are not allowed to touch the treadmill side railings.

We have found it desirable to depart from this standard protocol when testing personnel over the age of 40, where we wish to combine electrocardiographic stress evaluation along with aerobic power determination. In this case, we employ a continuous uphill walking treadmill protocol to facilitate good-quality electrocardiographic traces. It consists of walking at a constant velocity of 3.3 mph (90 m/min) and elevating the treadmill incline 5% every 3 minutes without intervening rest. $\dot{V}O_2$ is measured as described previously. If $\dot{V}O_2$ does not plateau, the highest $\dot{V}O_2$ achieved is taken to represent $\dot{V}O_2$max.

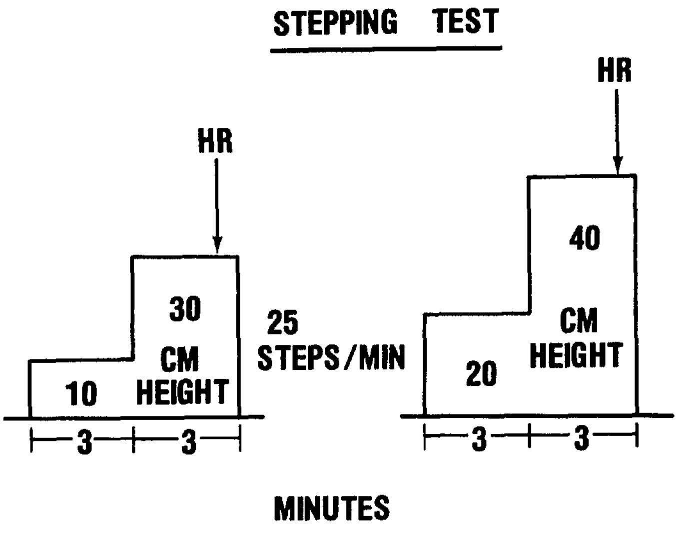
Although we typically do not utilize predictive techniques in the laboratory for assessing aerobic fitness in the Army, we have examined them for application to initial entry testing for occupational classification. Some of these data are presented here for comparative purposes. The two predictive procedures used included the familiar Astrand-Ryhming cycle ergometer single load test (16) and a step test.

The procedure followed in the cycle ergometer test is as originally described by Astrand and Ryhming in which the subject pedals at a resistance that will result in a heart rate response between 120 and 170 at the end of 6 minutes. The resistance setting (watts or kilopound meters) and heart rate are applied to a nomogram for the estimation of $\dot{V}O_2$max.

The step-test procedure is illustrated in figure 4. Subjects began by stepping at a cadence of 25 complete steps per minute at a step height of either 10, 20, or 30 cm, depending on an estimate of their fitness level. The heart rate observed at the end of 3 minutes is used to adjust the load to a higher step (30 or 40 cm) for an additional 3 minutes for the final heart rate reading. A stepstool with folddown steps is utilized (figure 5). $\dot{V}O_2$max is again estimated from the observed heart rate and exercise intensity (step height and frequency) (16). In both procedures, heart rate is measured electrocardiographically with disposable chest electrodes. We have observed correlation coefficients of 0.63 and 0.64 with actual treadmill $\dot{V}O_2$max for the Astrand-Ryhming cycle and step-test procedures, respectively.

Anaerobic power/muscular endurance

Our laboratory has employed two procedures to assess anaerobic power capacity for Army fitness evaluations: the Wingate power test (17, 18) and the



MALE: $\dot{V}O_{2\,MAX} = \dfrac{195-61}{HR_S-61} \cdot \dot{V}O_{2\,STEP}$

FEMALE: $\dot{V}O_{2\,MAX} = \dfrac{198-72}{HR_S-72} \cdot \dot{V}O_{2\,STEP}$

$\dot{V}O_{2\,STEP}$ 30 CM = 26 ML/KG . MIN.
40 CM = 32 ML/KG . MIN.

Figure 4. Step test procedure for estimation of maximal oxygen uptake.

Figure 5. Folddown stepping stool for step test determination of maximal oxygen uptake.

Thorstensson isokinetic endurance test (19). Both are designed to evaluate the capacity to generate muscular power from the anaerobic glycolytic energy pathways.

The Wingate protocol involves pedalling at maximal velocity for 30 seconds against a resistance based on body weight selected to elicit maximal power output over a 30-second period. The exercise is performed on a modified mechanical braked ergometer (20). The weighted pendulum is replaced by a counterbalanced lever arm to which a weight is attached so that resistance can be applied instantaneously. Resistance applied is 4.41 joules/pedal revolution/kg body weight. After achieving a near maximal pedal rate with no resistance applied, the lever arm is dropped applying the resistance, and the subject continues pedaling all out for 30 seconds. Mean power output over the 30-second period is computed.

For the isokinetic endurance test, the subject performs repeated knee extensions against a lever arm connected to a dynamometer and speed control device that maintains angular velocity at 180 degrees per second (Cybex II apparatus). Fifty knee extension contractions are performed requiring 60 seconds. The mean peak torque over the 50 contractions is computed.

Muscular strength

Muscular strength is a measure of the maximal force that can be generated in a single contractile effort. It may also represent the peak power that can be generated in a dynamic exercise of no longer than 5 seconds, thus measuring only energy that is immediately available within the muscle. We have developed a variety of strength measures for our assessment batteries in order to include various modes of activity and several different muscle groups.

Dynamic-isokinetic. If possible, we prefer to utilize dynamic measures of strength as opposed to static or isometric because most real-life tasks are dynamic. Our typical strength assessment battery includes the use of the Cybex II dynamometer and isokinetic apparatus to measure elbow flexion and extension and knee extension and flexion (21). At least two velocities are employed, 30 and 180 degrees per second. The average of three single contractions is recorded.

Dynamic-lift. Also included in our muscle strength battery is a measure of the maximal lift capacity to a height of 5 or 6 feet. This represents a total or composite strength of several muscle groups. The procedure (22, 23) involves lifting a weighted carriage that rides on a vertical track in incremental steps until the maximum lift weight is achieved (figure 6). This procedure is presently employed to screen all new Army and Air Force applicants.

Isometric. Isometric or static maximal contractions are used in our assessment batteries when time, safety, or other constraints may apply. Handgrip maximal force is included because it has been found to correlate well with general body strength (24). We utilize a noncommercial handgrip dynamometer (figure 7) (21) that includes an adjustable grip surface to account for the 15 degrees ulnar deviation of the hand. It is coupled to a load cell transducer.

Because one of the primary strength tasks in the Army is lifting, we also include an isometric 38-cm upright pull-force measure (figure 8) (25). This involves pulling vertically on a bar from a squatting position centered over the force transducer. It measures the strength of many of the muscle groups involved in lifting. Three other isometric measures have been employed: knee extension, trunk extension, and upper torso pull-down force (26) (figures 9–11).

Findings

Aerobic power

Most of our Army population surveys of aerobic fitness have been reported recently (10, 11). Extracts of these are presented here.

Table 8 presents data for new male and female recruits as they enter the Army. These data are representative of the civilian population entering the military service. Absolute $\dot{V}O_2max$ is 40% less in women but

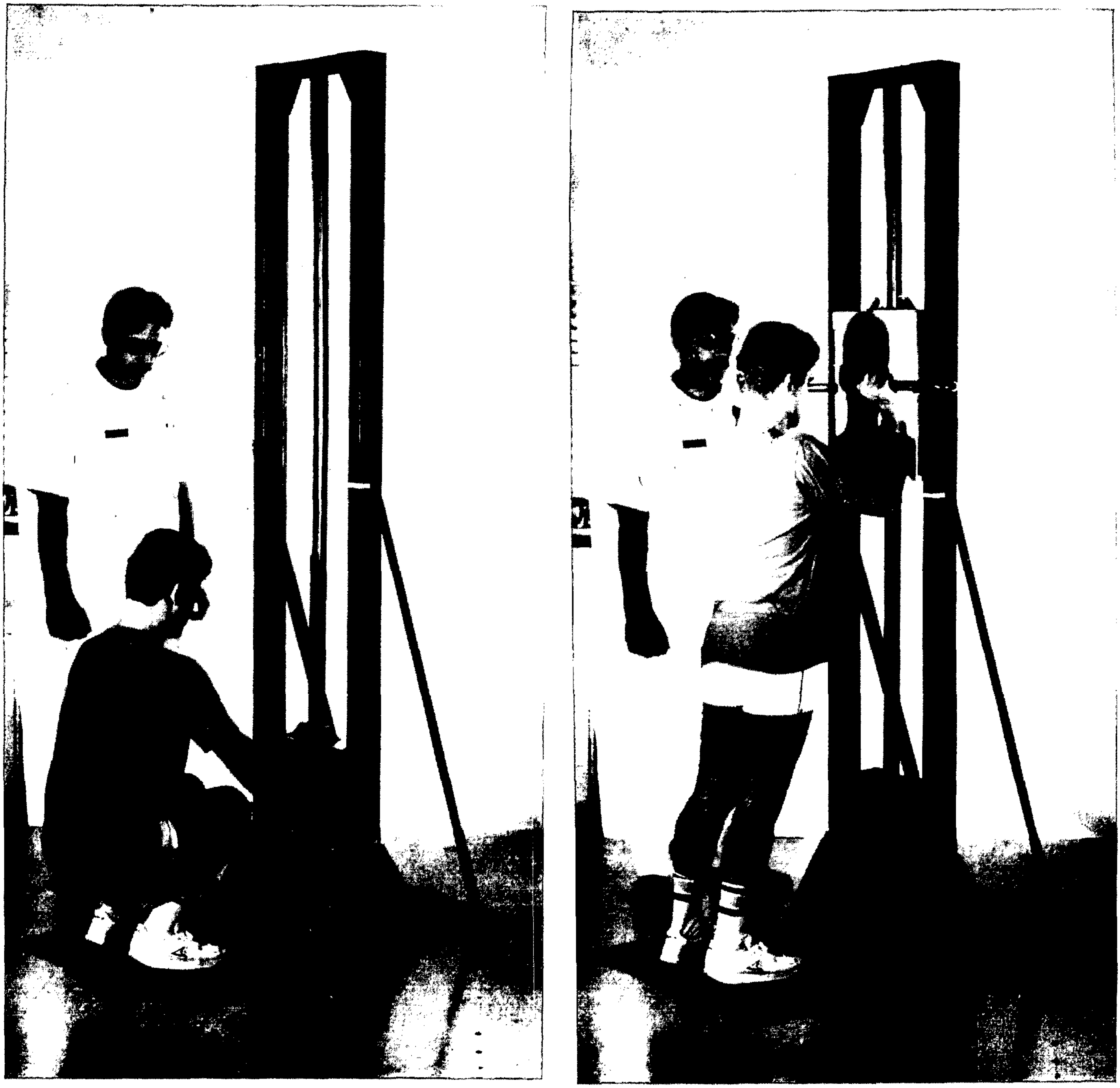


Figure 6. Incremental dynamic lift device used to determine maximal lift capacity.

only 15% less when adjusted for difference in fat-free weight. The relatively small overlap between genders is illustrated in figure 12. Figure 13 illustrates the typical decrements in $\dot{V}O_2max$ with age in high-intensity and low-intensity training Army units. Table 9 illustrates the influence of occupational intensity on $\dot{V}O_2max$, body weight, and body composition. Table 10 summarizes our most recent Army survey study, where $\dot{V}O_2max$ was measured directly on the treadmill. This latter study represents a cross-section of a large Army post with a wide variety of units and occupations. Table 11 presents data obtained with the predictive $\dot{V}O_2max$ step test and cycle ergometer procedures in Army recruits.

Anaerobic power/muscular endurance

Considerably less data are available regarding anaerobic power values in Army populations because this has only recently been added to our test batteries. Table 12 summarizes the data from two recent studies on military populations.

Muscle strength

A recent report from our laboratory has summarized muscle strength data from military population samples (11). Tables 13–17 give a compilation of data from U.S.

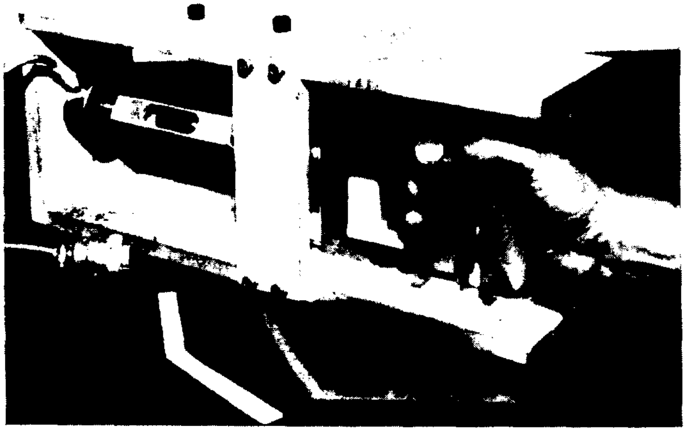


Figure 7. Handgrip device with adjustable grips and load cell transducer.

military samples. Table 13 presents mean values of isometric strength of various muscle groups, and table 14 presents mean values for peak isokinetic torque of two muscle groups at two different velocities. Tables 15 and 16 give mean male-female comparisons of strength and lifting capacity. The variance in values in male and female samples is presented in table 17 and illustrated in figure 14 for lifting capacity to 152 cm.

Body composition

Background

Body composition in terms of its two main components, fat and fat-free mass, is included in this presenta-

Figure 8. 38-centimeter isometric upright pull force device.

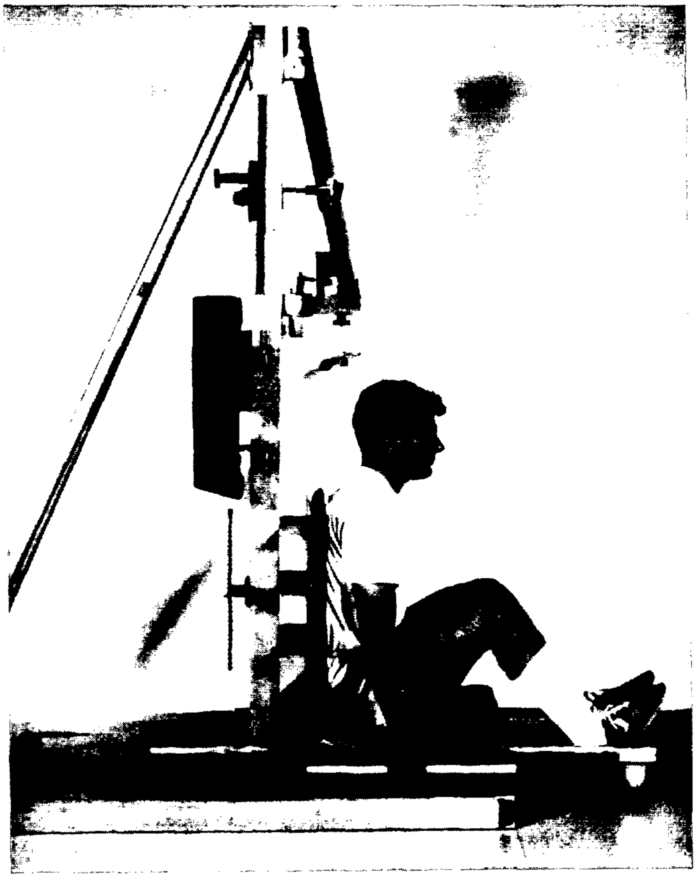


Figure 9. Isometric measurement of knee extension strength.

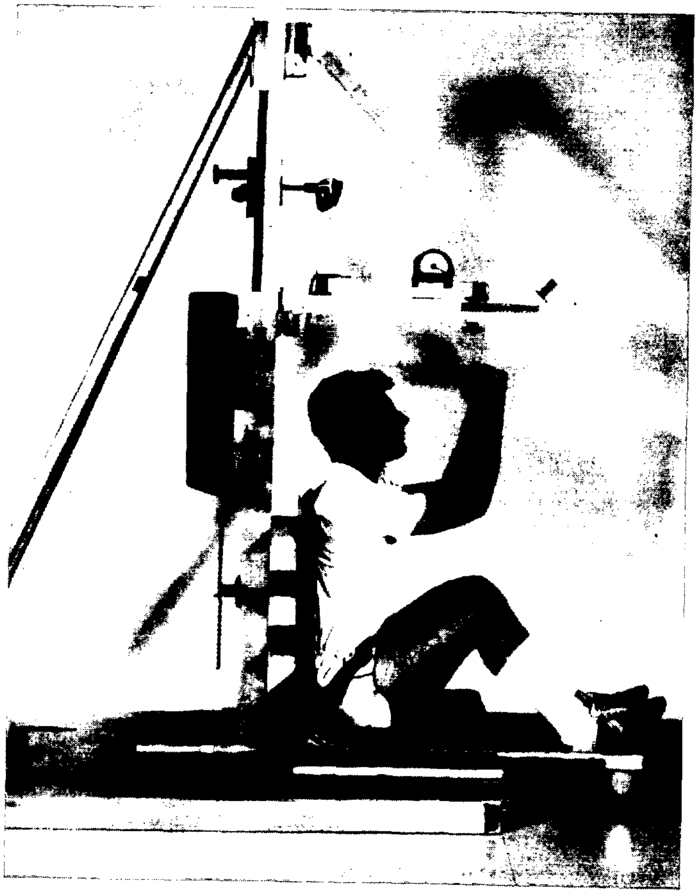


Figure 10. Isometric measurement of shoulder-arm strength.

Table 8. Maximal O_2 uptake, anthropometric and related variables of men and women entering the Army from civilian life, preinitial entry training

Variable	Male (*n* = 210)		Female (*n* = 212)		F/M
	Mean + SD	Range	Mean + SD	Range	
Age (years)	19.7 ± 2.2	17–25	19.7 ± 1.9	17–25	—
Height (cm)*	174.7 ± 6.9	153.7–195.1	162.0 ± 6.4	146.7–183.2	—
Body weight (kg)*	70.5 ± 10.7	45.8–105.5	58.6 ± 7.0	42.2–77.5	.83
Body fat (% of BW)*	15.6 ± 5.6	6.0–32.7	28.4 ± 4.5	12.4–38.8	—
Lean body mass (kg)*	59.1 ± 7.0	40.7–80.6	41.8 ± 4.4	32.7–53.1	.71
$\dot{V}O_2$ ($l\cdot min^{-1}$)*	3.60 ± 0.50	2.31–5.35	2.18 ± 0.32	1.24–3.14	.61
$\dot{V}O_2max$ ($ml\cdot kgBW^{-1}\cdot min^{-1}$)*	51.1 ± 5.1	32.4–63.7	37.5 ± 3.7	24.1–47.1	.73
$\dot{V}O_2max$ ($ml\cdot kgLBM^{-1}$)*	60.9 ± 5.6	44.4–79.5	52.4 ± 5.4	32.0–70.1	.86
HRmax ($beats\cdot min^{-1}$)	190.7 ± 6.8	172–210	189.8 ± 7.4	164–210	—
$\dot{V}_Imax$ [$l\cdot min^{-1}$ (BTPS)]*	139 ± 21.3	83.9–194.0	88.6 ± 15.7	46.1–131.7	—

* Mean differences significant at 1% confidence level.
Source: Vogel et al. (10).

tion because of its obvious relationship to fitness capacity and exercise performance. Although the relative proportions of muscle and fat have a direct influence on fitness appearance, they also are related to functional fitness capacity and health-related fitness. Even though some degree of body fat stores are necessary as energy sources and mechanical cushioning, excess stores are a burden to the body in that they represent excess weight that must be transported by the active muscle mass. In

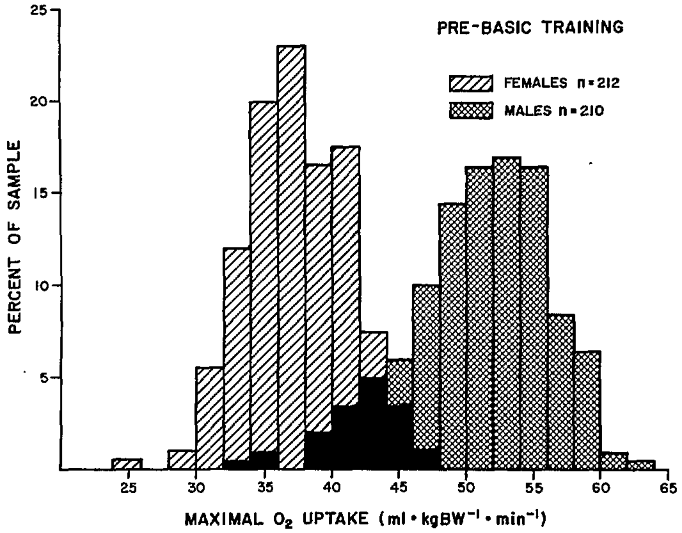


Figure 12. Distribution of maximal oxygen uptake in male and female recruits at the beginning of basic initial entry training. From Vogel et al. (10).

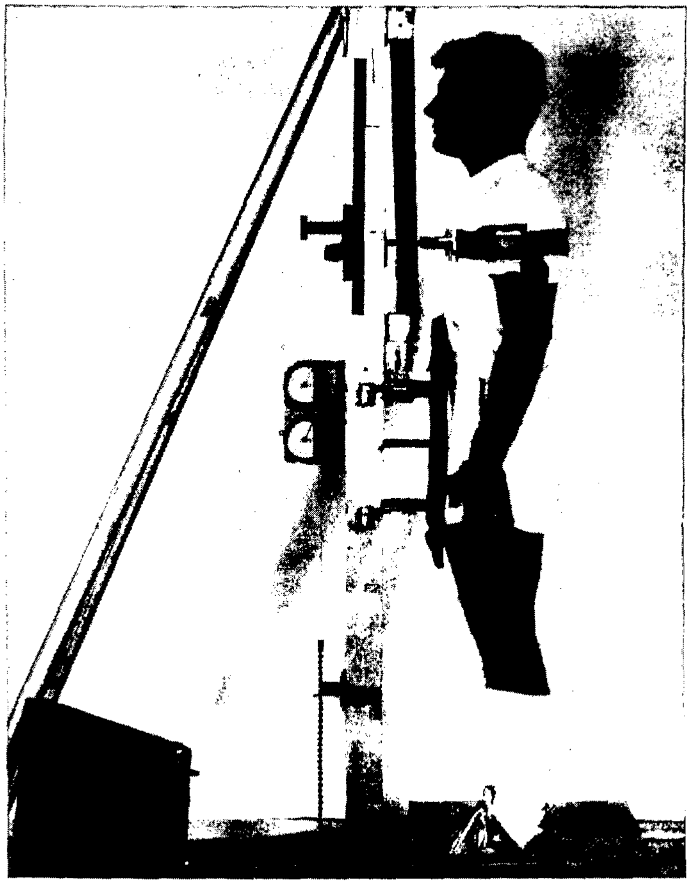


Figure 11. Isometric measurement of trunk extension strength.

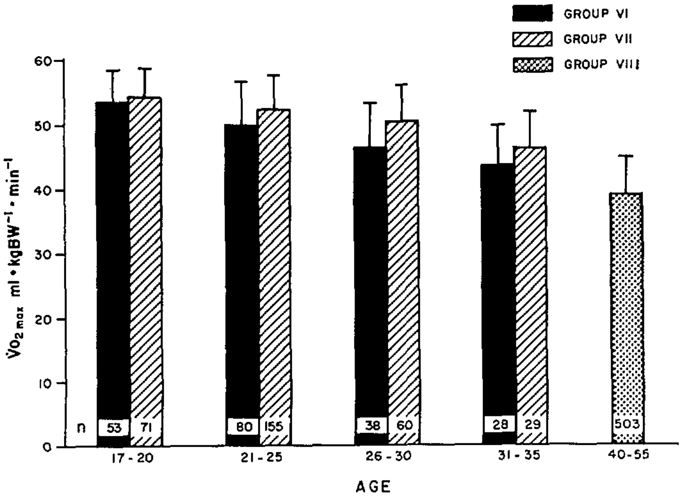


Figure 13. Influence of age on maximal oxygen uptake in three groups of male soldiers. Group VI: untrained; VII: trained; VIII: over age 40. From Vogel et al. (10).

Table 9. Effect of occupational physical intensity level on maximal oxygen uptake, body weight, and body fat

Category	Occupational rating Heavy	 Moderate	 Light
Group I (variable training intensity)			
n	82	20	40
$\dot{V}O_2$max (ml·kgBW^{-1}·min^{-1})**	50.6 ± 6.4	46.7 ± 7.7	47.1 ± 7.5
Body weight (kg)*	72.2 ± 10.0	70.6 ± 9.4	74.5 ± 12.0
Body fat (% of BW)*	17.2 ± 5.0	19.6 ± 6.7	19.99 ± 6.3
Group II (high training intensity)			
n	122	62	81
$\dot{V}O_2$max (ml·kgBW^{-1}·min^{-1})	53.0 ± 5.0	52.4 ± 5.9	50.5 ± 5.7
Body weight (kg)	72.2 ± 10.0	70.6 ± 9.4	74.5 ± 12.0
Body fat (% of BW)**	18.7 ± 5.4	18.4 ± 5.5	20.9 ± 6.0

* ANOVA F (<.05)
** ANOVA F (<.01)
Source: Vogel et al. (10)

this respect, they detract from the aerobic capacity of the individual when transporting his or her own body weight such as in walking or running. Thus, there is a general relationship between aerobic power adjusted for body weight and percent body fat content as illustrated in figure 15 (10). Muscle mass, or measurable fat-free mass, is related to strength or anaerobic power capacity because the force that can be generated by a muscle is related to its cross-sectional area. Figure 16 illustrates the relationship between lifting capacity and fat-free mass.

Methods

Suitable methods for assessing body composition in a field (outside the laboratory) setting are of great

Table 10. $\dot{V}O_2$max of soldiers assigned to a variety of units and occupations at one Army post
[ml·kg^{-1} body weight·min^{-1}]

Category	Male n	 Mean	 SD	Female n	 Mean	 SD
Combined	956	48.0	6.3	240	39.7	4.6
Age:						
17–20 years	126	52.0	4.3	50	41.1	5.3
21–27 years	332	50.2	5.7	143	39.7	4.0
28–39 years	275	45.1	5.7	46	38.3	5.3
40 years and over	223	46.0	6.5	—	—	—
Black	213	48.5	6.3	84	38.4	4.1
Hispanic	103	48.0	6.8	17	41.3	4.2
White	603	47.9	6.2	131	40.3	4.7

Source: Fitzgerald (4).

Table 11. $\dot{V}O_2$max in Army recruits predicted from the step test and Astrand-Ryhming procedures
[ml·kg^{-1} body weight ·min^{-1}]

Category	Male n	 Mean	 SD	Female n	 Mean	 SD
Astrand-Rhyming Cycle	273	44.3	8.0	274	38.8	7.8
Step test	444	48.3	6.3	387	34.9	5.5

Source: Teves, Wright, and Vogel (22).

Table 12. Anaerobic power values in male military populations

Study reference	Mean	SD	Range	Mean	SD	Range
Wingate test (mean power, watts)	Upper body			Lower body		
7	424	73	301–567	440	101	238–683
27	383	42	312–481	611	57	520–699
Isokinetic test (mean peak torque, Nm)	Elbow extensors			Knee extensors		
7	23	7	12–52	78	17	52–121
27	19	3	14–25	77	13	58–105

Table 13. Mean isometric strength values in male military populations
[Values in kg]

Reference	Subjects	Hand-grip	Knee extension	Trunk extension	38-cm pull	Horizontal arm pull
28	Army recruits (n = 102)				103.9	
6	Army recruits (n = 769)		158.2	79.0		
2	Army recruits (n = 462)	52.6			148.8	
29	Infantry (n = 50)		161	77		
30	Navy recruits (n = 350)	46.1				71.1
11	Infantry (n = 32)	56.2	186.0	89.0	130.6	
31	Navy trainees (n = 69)	52.2				70.0
32	Infantry	54.0	167.6	80.0	138.0	

Table 14. Mean isokinetic muscle strength values in male infantry soldiers

Reference	Elbow flexion 30°/sec	 180°/sec	Knee extension 30°/sec	 180°/sec
29 (n = 50)	56.5	40.5	215.0	178.0
11 (n = 32)	53.6	41.1	222.5	128.6

Table 15. Male-female mean comparisons of muscle strength in military populations

Reference	Isometric handgrip (kg)			Isometric knee extension (kg)			Isometric trunk extension (kg)			Isometric 38-cm upright pull (kg)			Isokinetic trunk extension at 36°/sec peak torque,N,m		
	M	F	F/M	M	F	F/M	M	F	F/M	M	F	F/M	M	F	F/M
30	46.1	28.5	.62												
32	54.0	34.1	.63	167.6	99.3	.59	80.0	51.3	.64	138.0	83.7	.60	286.9	163.2	.57
6				158.2	106.6	.67	79.0	56.6	.72						
2										148.8	95.2	.63			
28										103.9	58.3	.56			

Table 16. Male-female mean comparisons of one-repetition maximum lift capacity in military populations

Reference	Max lift to 132 cm (kg)			Max lift to 152 cm (kg)			Max lift to 183 cm (kg)			Max lift to shoulder height (kg)			Max lift to elbow height (kg)		
	M	F	F/M	M	F	F/M	M	F	F/M	M	F	F/M	M	F	F/M
11	77.7	35.5	.66												
32	57.6	32.5	.56												
2				65.5	34.4	.53	62.1	30.4	.49	50.8	30.2	.59			
23							51.8	25.8	.50				58.6	30.7	.52

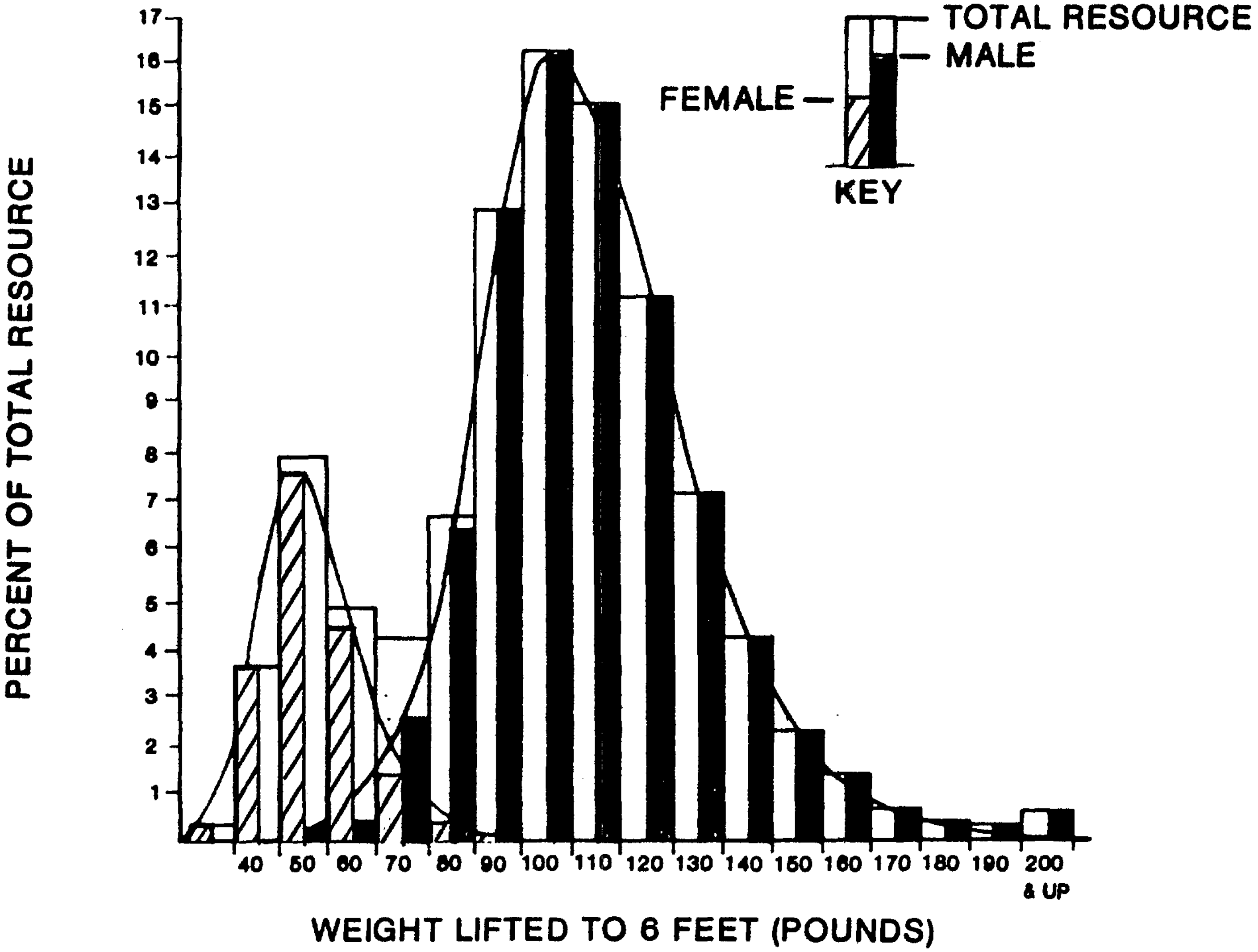


Figure 14. Distribution of lifting capacity to 6 feet in male and female Air Force recruits. From McDaniel, Skandis, and Madole (23).

interest to the military and the topic of considerable recent research. This stems from the importance of the military services' weight control programs to maintain adequate appearance and enhance physical performance. Although the Services have traditionally enforced weight control through weight-for-height standards, their limitations were recognized. In 1981 the Services were instructed to supplement these tables with a secondary body fat standard to handle the overmuscular individual who does not meet weight tables standards but is not obese. The Army chose to meet this requirement by establishing an age and gender-adjusted standard for percent body fat (table 18). Body fat assessment was performed in medical facilities with the skinfold caliper technique. Of the many skinfold body fat equations available, the Army chose the Durnin-Womersley equations because it is age adjusted and commonly used by other NATO military services.

The Durnin-Womersley procedure employs four skinfolds: bicep, tricep, subscapular, and suprailiac and is represented by the following equations:

$$\%\ \text{fat (males)} = [4.95 - (1.1739 - 0.06227 \times \log 10\ \text{sum of 4 SF} - 0.000555 \times \text{age}) - 4.5] \times 100$$

$$\%\ \text{fat (females)} = [4.95 - (1.1572 - 0.0647 \times \log 10\ \text{sum of 4 SF}\ 0.00038 \times \text{age}) - 4.5] \times 100$$

The practical limitation to the skinfold procedure is the potential for large intermeasurer error stemming from variation in site location, extent of pinch, and application of calipers. This problem was particularly severe in the Army experience, where skinfolds were being performed at about 100 locations with at least that number of measurers or more. Even though considerable effort was made to train and credential the measurers, variability was evident and acceptance of the procedure suffered.

Because of these difficulties with the skinfold procedure, it was decided to explore other anthropometric variables that would give adequate predictability of body

Table 17. Values of isometric strength and lifting capacity of Army recruits

Test	Male (n = 980)		Female (n = 1,004)	
	Mean	SD	Mean	SD
Handgrip (kg)	47.5	7.4	30.2	5.5
38-cm pull (kg)	124.8	21.2	77.1	13.5
IDL 152 (kg)	60.6	10.7	29.8	5.4
IDL 183 (kg)	56.7	10.5	25.6	4.7

Source: Teves, Wright, and Vogel (22).

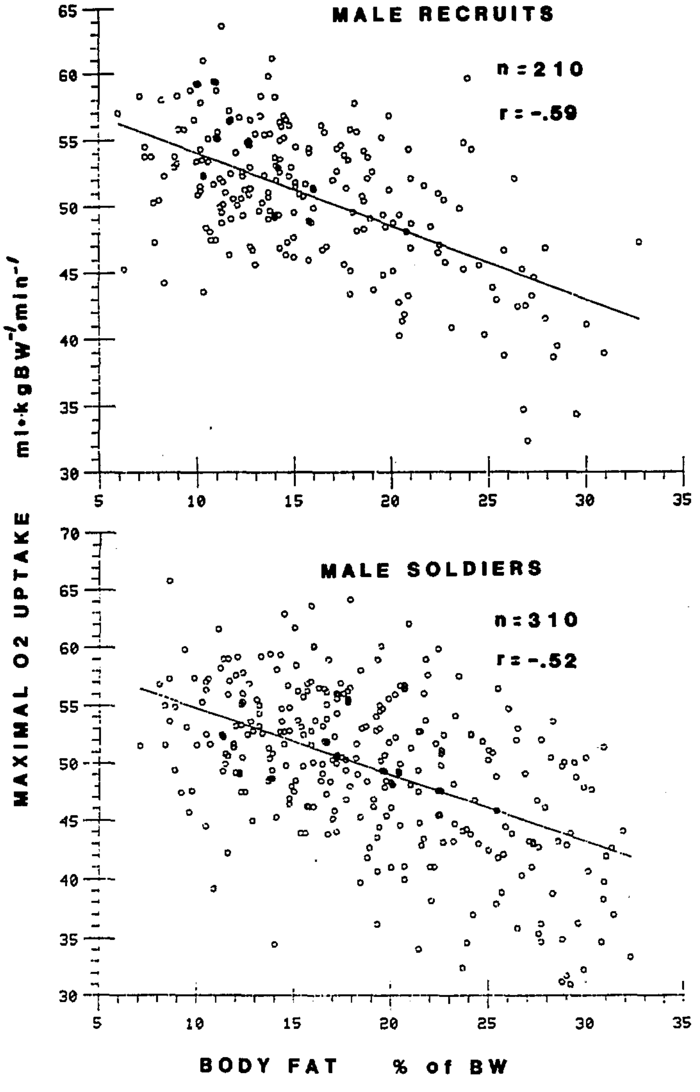


Figure 15. Relation between maximal oxygen uptake (weight adjusted) and percent body fat. From Vogel et al. (10).

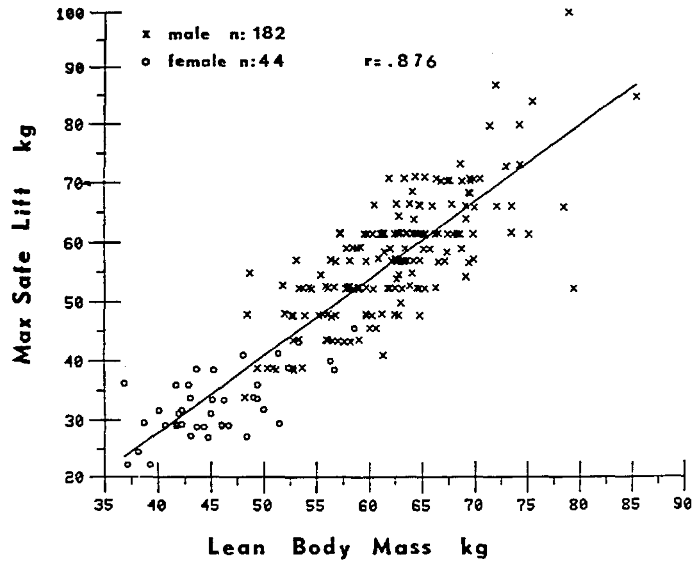


Figure 16. Relation between maximal lift capacity and fat-free mass.

Table 18. US Army maximal limits for percent body fat

Sex	17–20 years	21–27 years	28–39 years	40 years and over
Male	20	22	24	26
Female	28	30	32	34

Table 19. Percent body fat values of U.S. Army populations taken by the Durnin-Womersley skinfold procedure

			% body fat					
			Male			Female		
Reference	Subjects	Age	*n*	Mean	SD	*n*	Mean	SD
8	Recruits:							
	pre		87	16.3	5.0	57	28.2	4.6
	post		87	14.5	3.8	57	26.2	3.8
33	Recruits:							
	pre	17–20 years	353	15.3	4.7	146	27.7	4.2
		21–25 years	507	16.1	5.2	277	28.8	4.5
		26–30 years	77	18.1	5.2	60	28.3	4.3
		31–35 years	12	22.4	4.6	16	31.0	4.8
34	Infantry	40–53 years	260	26.0	4.6			
11	Infantry	17–20 years	73	15.8	4.1			
		21–25 years	95	17.9	6.1			
		26–30 years	38	19.3	5.9			
		31–35 years	28	20.0	5.8			

fat and could be applied at the unit/organizational level and thereby relieve the medical facilities of this responsibility. Our laboratory carried out a major research project in 1984 to seek such a field procedure. This consisted of hydrostatic weighing and anthropometric measures on nearly 1,500 Army personnel. The outcome was the derivation of a circumference technique employing the following equations:

$$\%\text{ body fat (male)} = 46.892 - 68.678 \times \log_{10}\text{height} + 76.462 \times \log_{10}(\text{abdominal-neck circumference})$$

$$\begin{aligned}\%\text{ body fat (female)} = & -35.601 - 0.515 \times \text{height} \\ & + 0.173 \times \text{hip circumference} \\ & - 1.574 \times \text{forearm circumference} \\ & - 0.533 \times \text{neck circumference} \\ & - 0.200 \times \text{wrist circumference} \\ & + 105.328 \times \log_{10}\text{body weight}\end{aligned}$$

Abdominal circumference is measured at the umbilicus; hip circumference, at the largest protrusion of the buttocks; and forearm, at the largest point (extended). Correlation coefficients and standard error of estimate for the male and female equations are: $r = .817$, SEE = 4.020, and $r = .820$, SEE = 3.598, respectively (4). These circumference procedures and equations will be implemented by the Army in April 1986.

Findings

Table 19 summarizes percent body fat values derived by the Durnin-Womersley skinfold procedure for a sample of U.S. Army units. Table 20 presents a summary of our recent hydrostatic weighing data on a large Army population along with similar data from the U.S. Navy. The significant age effect on percent body fat is readily observed in these data.

Physical activity assessment

Assessments of physical activity levels are not routinely made of military populations as is physical fitness and body composition. They are, however, an integral part of research studies concerned with fitness and physical training. Five samples of activity questionnaires that have been employed in our studies are shown in figure 17. They are presented as examples. No data are available concerning their validity.

Conclusions and recommendations

This chapter reviews the selection of and the methods employed by the U.S. Army to assess physical fitness both in the field and within the laboratory. Representative data from various Army population surveys are presented on new recruits as well as soldiers across a wide age span. The recruit population represents a selected civilian sample, typical of healthy, active young Americans. The data presented here offer a substantial data base that is suitable for comparative purposes with the NCHS general population surveys.

Physical fitness assessment, or the measure of functional exercise capacity, is a valuable supplement to health surveys for two reasons. First, it can be considered as an additional level of detection of disease or

Table 20. Percent body fat values in U.S. Army populations by the hydrostatic weighing procedure
[mean ± SD]

		% body fat	
Age	*n*	Male	Female
17–20 years	160	16.5 ± 5.6	28.1 ± 5.3
21–27 years	383	17.8 ± 6.4	27.2 ± 5.9
28–39 years	318	22.8 ± 7.1	30.9 ± 5.8
40+ years	258	24.2 + 5.2	—

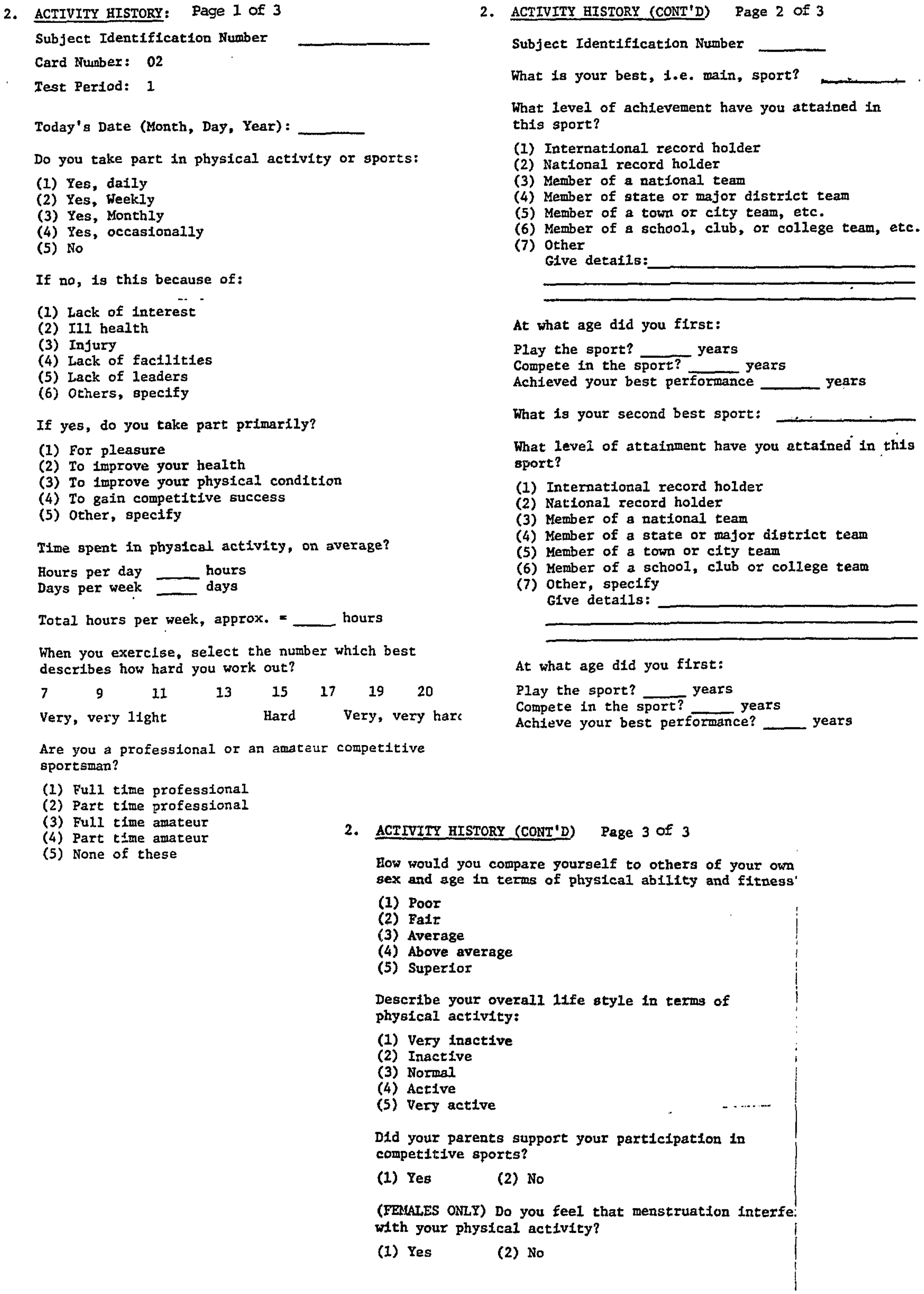
Source: Fitzgerald (4).

incapacity where the body is subjected to a load or demand. Thus it increases the chances for detection, as compared with examinations carried out only in the resting, nonchallenged state. Secondly, it also provides demographic data on the functional capacity of our population that would be useful for a wide variety of employment and industrial related matters.

Surveys of physical fitness should include all three components of exercise capacity: aerobic power, muscle strength, and muscular endurance because they represent three distinct energy-generating systems and therefore three separate capacities for muscular function. Body composition should also be added because most aspects of fitness must be interpreted in terms of the portions of fat or fat-free mass.

Decisions regarding the selection of fitness measurements are based on a number of considerations including time and space available, safety, motivation and cooperation required of the participant, and the degree of reliability and sensitivity desired. In our experience in the laboratory setting, indirect or predictive methods of exercise capacity have considerable limitations and insufficient accuracy and reliability. We have concluded that the dependability and accuracy of direct measurements, even in small numbers, outweigh the larger numbers that can be obtained with indirect, predictive procedures. In the case of aerobic fitness, we suggest that actual measurement of oxygen uptake is preferable to estimation from heart rate and that measurements at maximal effort are preferable to those at submaximal effort. In respect to strength and strength endurance, we suggest that actual maximal capacity measurements of lifting, pushing, pulling or cranking, for example, are more meaningful than such measurements as isometric handgrip or elbow flexion force.

A major limitation to the use of direct measurements of exercise capacity in population survey settings is the safety concerns. Although the safety of maximal aerobic testing can be greatly ensured by careful screening and monitoring, the safety of direct measurements of maximal lifting may be more difficult. However, in our experience, the use of equipment such as a weight machine where the body motion can be limited to safe positions and carefully monitored results in a very high safety level.



2\. ACTIVITY HISTORY: Page 1 of 3

Subject Identification Number ____________

Card Number: 02

Test Period: 1

Today's Date (Month, Day, Year): ______

Do you take part in physical activity or sports:

(1) Yes, daily
(2) Yes, Weekly
(3) Yes, Monthly
(4) Yes, occasionally
(5) No

If no, is this because of:

(1) Lack of interest
(2) Ill health
(3) Injury
(4) Lack of facilities
(5) Lack of leaders
(6) Others, specify

If yes, do you take part primarily?

(1) For pleasure
(2) To improve your health
(3) To improve your physical condition
(4) To gain competitive success
(5) Other, specify

Time spent in physical activity, on average?

Hours per day ____ hours
Days per week ____ days

Total hours per week, approx. = ____ hours

When you exercise, select the number which best describes how hard you work out?

7 9 11 13 15 17 19 20

Very, very light Hard Very, very har

Are you a professional or an amateur competitive sportsman?

(1) Full time professional
(2) Part time professional
(3) Full time amateur
(4) Part time amateur
(5) None of these

2\. ACTIVITY HISTORY (CONT'D) Page 2 of 3

Subject Identification Number ______

What is your best, i.e. main, sport? ______

What level of achievement have you attained in this sport?

(1) International record holder
(2) National record holder
(3) Member of a national team
(4) Member of state or major district team
(5) Member of a town or city team, etc.
(6) Member of a school, club, or college team, etc.
(7) Other
Give details: ____________________

At what age did you first:

Play the sport? _____ years
Compete in the sport? _____ years
Achieved your best performance _____ years

What is your second best sport: ______

What level of attainment have you attained in this sport?

(1) International record holder
(2) National record holder
(3) Member of a national team
(4) Member of a state or major district team
(5) Member of a town or city team
(6) Member of a school, club or college team
(7) Other, specify
Give details: ____________________

At what age did you first:

Play the sport? _____ years
Compete in the sport? _____ years
Achieve your best performance? _____ years

2\. ACTIVITY HISTORY (CONT'D) Page 3 of 3

How would you compare yourself to others of your own sex and age in terms of physical ability and fitness'

(1) Poor
(2) Fair
(3) Average
(4) Above average
(5) Superior

Describe your overall life style in terms of physical activity:

(1) Very inactive
(2) Inactive
(3) Normal
(4) Active
(5) Very active

Did your parents support your participation in competitive sports?

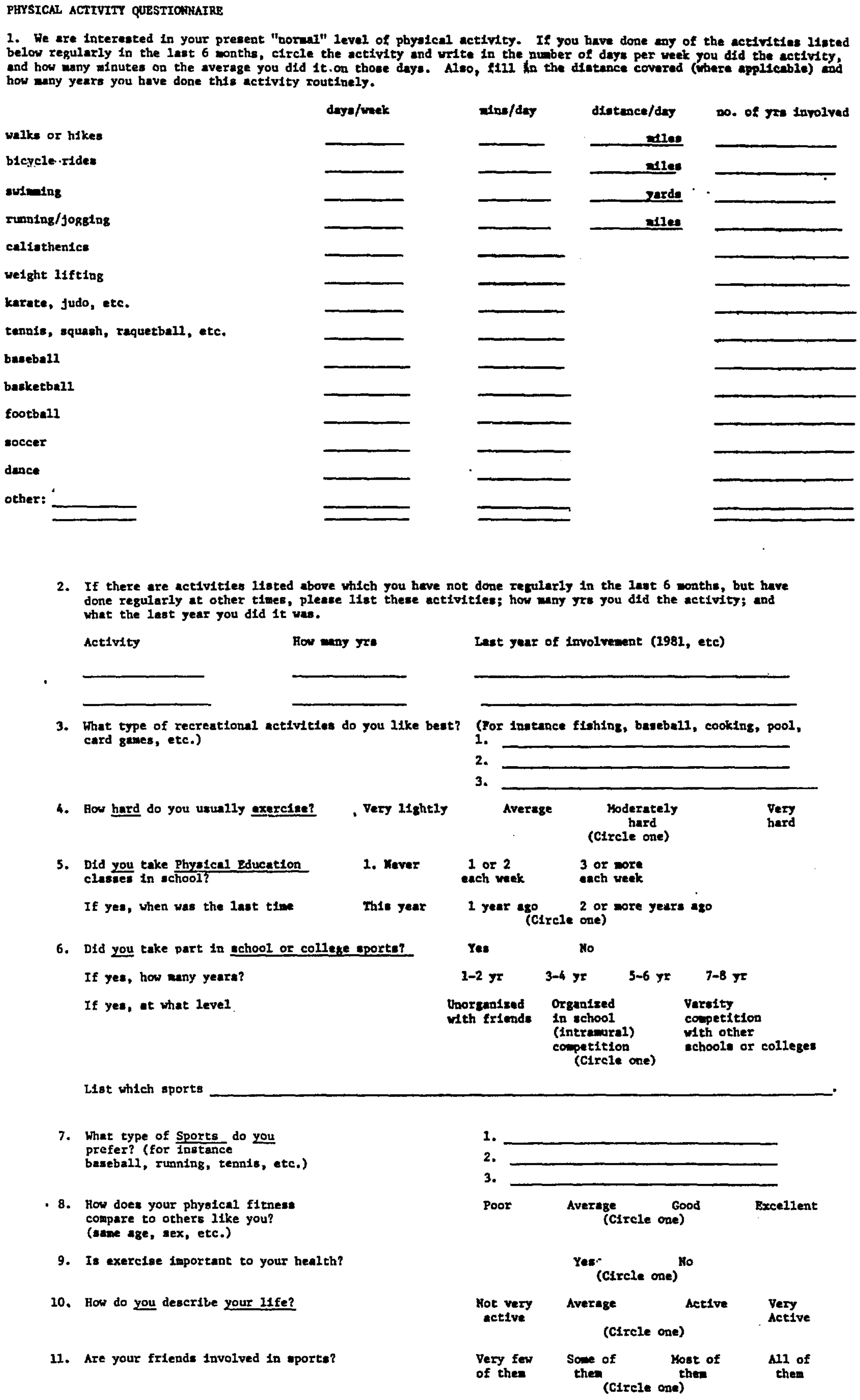
(1) Yes (2) No

(FEMALES ONLY) Do you feel that menstruation interfe with your physical activity?

(1) Yes (2) No

Sample A

Figure 17. Samples of physical activity questionnaire used in Army fitness research studies.



PHYSICAL ACTIVITY QUESTIONNAIRE

1. We are interested in your present "normal" level of physical activity. If you have done any of the activities listed below regularly in the last 6 months, circle the activity and write in the number of days per week you did the activity, and how many minutes on the average you did it on those days. Also, fill in the distance covered (where applicable) and how many years you have done this activity routinely.

	days/week	mins/day	distance/day	no. of yrs involved
walks or hikes	______	______	______ miles	______
bicycle rides	______	______	______ miles	______
swimming	______	______	______ yards	______
running/jogging	______	______	______ miles	______
calisthenics	______	______		______
weight lifting	______	______		______
karate, judo, etc.	______	______		______
tennis, squash, raquetball, etc.	______	______		______
baseball	______	______		______
basketball	______	______		______
football	______	______		______
soccer	______	______		______
dance	______	______		______
other: ______	______	______		______
______	______	______		______

2. If there are activities listed above which you have not done regularly in the last 6 months, but have done regularly at other times, please list these activities; how many yrs you did the activity; and what the last year you did it was.

Activity	How many yrs	Last year of involvement (1981, etc)
______	______	______
______	______	______

3. What type of recreational activities do you like best? (For instance fishing, baseball, cooking, pool, card games, etc.)
1. ______
2. ______
3. ______

4. How hard do you usually exercise? Very lightly / Average / Moderately hard / Very hard (Circle one)

5. Did you take Physical Education classes in school? 1. Never / 1 or 2 each week / 3 or more each week

If yes, when was the last time: This year / 1 year ago / 2 or more years ago (Circle one)

6. Did you take part in school or college sports? Yes / No

If yes, how many years? 1-2 yr / 3-4 yr / 5-6 yr / 7-8 yr

If yes, at what level: Unorganized with friends / Organized in school (intramural) competition / Varsity competition with other schools or colleges (Circle one)

List which sports ______

7. What type of Sports do you prefer? (for instance baseball, running, tennis, etc.)
1. ______
2. ______
3. ______

8. How does your physical fitness compare to others like you? (same age, sex, etc.) Poor / Average / Good / Excellent (Circle one)

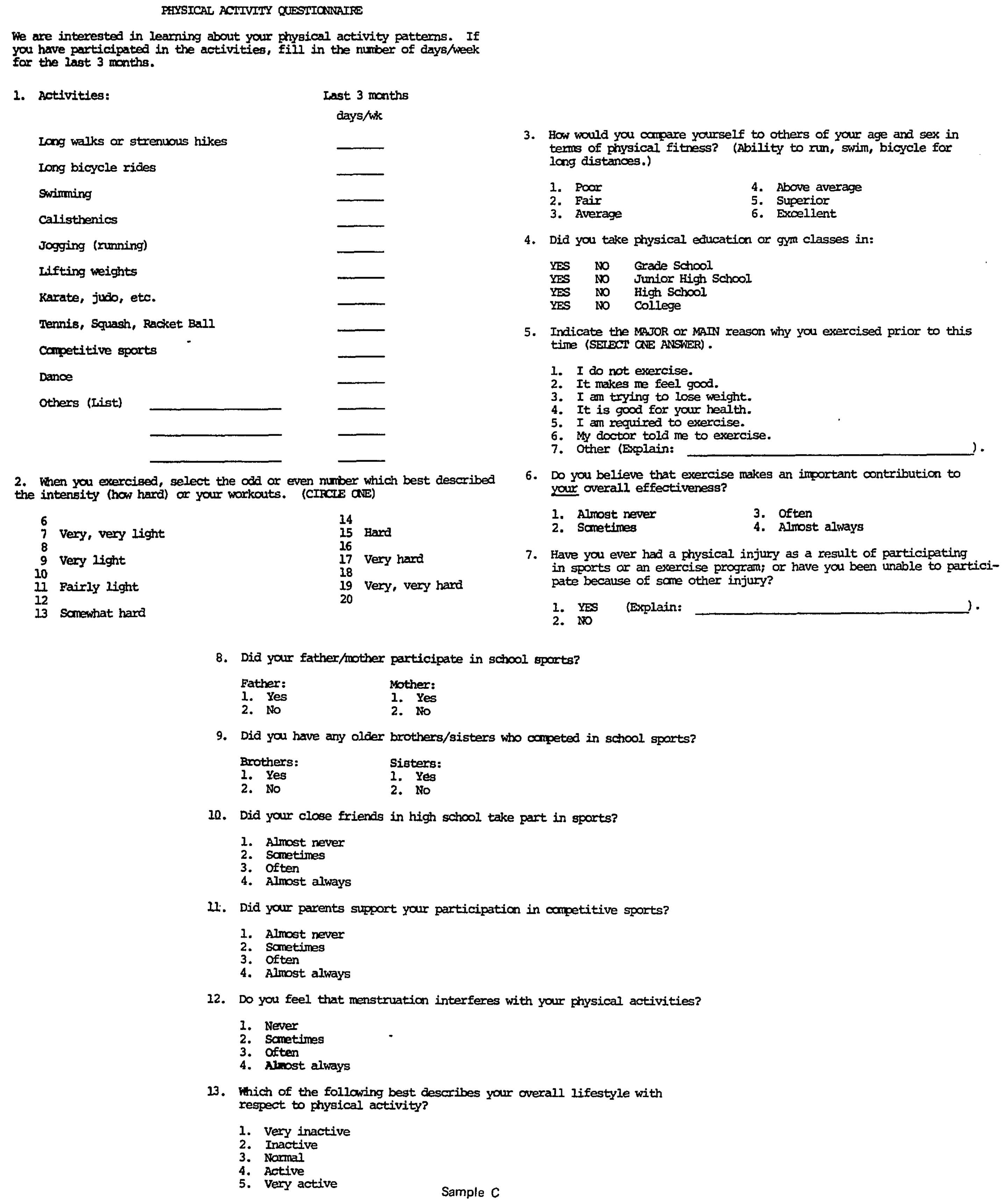
9. Is exercise important to your health? Yes / No (Circle one)

10. How do you describe your life? Not very active / Average / Active / Very Active (Circle one)

11. Are your friends involved in sports? Very few of them / Some of them / Most of them / All of them (Circle one)

Sample B

Figure 17. Samples of physical activity questionnaire used in Army fitness research studies—continued.



PHYSICAL ACTIVITY QUESTIONNAIRE

We are interested in learning about your physical activity patterns. If you have participated in the activities, fill in the number of days/week for the last 3 months.

1. Activities:

	Last 3 months days/wk
Long walks or strenuous hikes	______
Long bicycle rides	______
Swimming	______
Calisthenics	______
Jogging (running)	______
Lifting weights	______
Karate, judo, etc.	______
Tennis, Squash, Racket Ball	______
Competitive sports	______
Dance	______
Others (List) ________________	______
________________	______
________________	______

2. When you exercised, select the odd or even number which best described the intensity (how hard) or your workouts. (CIRCLE ONE)

6
7 Very, very light
8
9 Very light
10
11 Fairly light
12
13 Somewhat hard
14
15 Hard
16
17 Very hard
18
19 Very, very hard
20

3. How would you compare yourself to others of your age and sex in terms of physical fitness? (Ability to run, swim, bicycle for long distances.)

1. Poor
2. Fair
3. Average
4. Above average
5. Superior
6. Excellent

4. Did you take physical education or gym classes in:

YES NO Grade School
YES NO Junior High School
YES NO High School
YES NO College

5. Indicate the MAJOR or MAIN reason why you exercised prior to this time (SELECT ONE ANSWER).

1. I do not exercise.
2. It makes me feel good.
3. I am trying to lose weight.
4. It is good for your health.
5. I am required to exercise.
6. My doctor told me to exercise.
7. Other (Explain: ______________________________).

6. Do you believe that exercise makes an important contribution to <u>your</u> overall effectiveness?

1. Almost never
2. Sometimes
3. Often
4. Almost always

7. Have you ever had a physical injury as a result of participating in sports or an exercise program; or have you been unable to participate because of some other injury?

1. YES (Explain: ______________________________).
2. NO

8. Did your father/mother participate in school sports?

Father:
1. Yes
2. No

Mother:
1. Yes
2. No

9. Did you have any older brothers/sisters who competed in school sports?

Brothers:
1. Yes
2. No

Sisters:
1. Yes
2. No

10. Did your close friends in high school take part in sports?

1. Almost never
2. Sometimes
3. Often
4. Almost always

11. Did your parents support your participation in competitive sports?

1. Almost never
2. Sometimes
3. Often
4. Almost always

12. Do you feel that menstruation interferes with your physical activities?

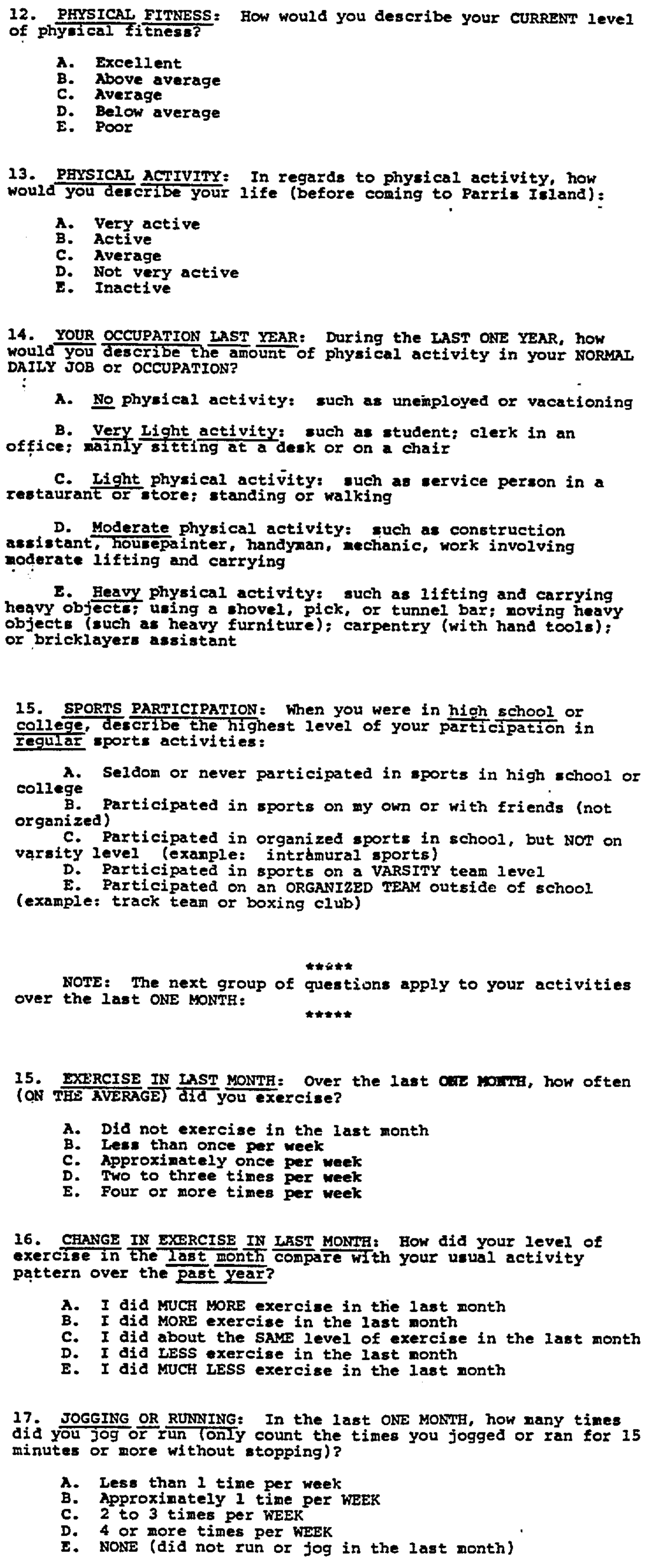
1. Never
2. Sometimes
3. Often
4. Almost always

13. Which of the following best describes your overall lifestyle with respect to physical activity?

1. Very inactive
2. Inactive
3. Normal
4. Active
5. Very active

Sample C

Figure 17. Samples of physical activity questionnaire used in Army fitness research studies—continued.



12. PHYSICAL FITNESS: How would you describe your CURRENT level of physical fitness?

A. Excellent
B. Above average
C. Average
D. Below average
E. Poor

13. PHYSICAL ACTIVITY: In regards to physical activity, how would you describe your life (before coming to Parris Island):

A. Very active
B. Active
C. Average
D. Not very active
E. Inactive

14. YOUR OCCUPATION LAST YEAR: During the LAST ONE YEAR, how would you describe the amount of physical activity in your NORMAL DAILY JOB or OCCUPATION?

A. No physical activity: such as unemployed or vacationing

B. Very Light activity: such as student; clerk in an office; mainly sitting at a desk or on a chair

C. Light physical activity: such as service person in a restaurant or store; standing or walking

D. Moderate physical activity: such as construction assistant, housepainter, handyman, mechanic, work involving moderate lifting and carrying

E. Heavy physical activity: such as lifting and carrying heavy objects; using a shovel, pick, or tunnel bar; moving heavy objects (such as heavy furniture); carpentry (with hand tools); or bricklayers assistant

15. SPORTS PARTICIPATION: When you were in high school or college, describe the highest level of your participation in regular sports activities:

A. Seldom or never participated in sports in high school or college
B. Participated in sports on my own or with friends (not organized)
C. Participated in organized sports in school, but NOT on varsity level (example: intramural sports)
D. Participated in sports on a VARSITY team level
E. Participated on an ORGANIZED TEAM outside of school (example: track team or boxing club)

*****

NOTE: The next group of questions apply to your activities over the last ONE MONTH:

*****

15. EXERCISE IN LAST MONTH: Over the last ONE MONTH, how often (ON THE AVERAGE) did you exercise?

A. Did not exercise in the last month
B. Less than once per week
C. Approximately once per week
D. Two to three times per week
E. Four or more times per week

16. CHANGE IN EXERCISE IN LAST MONTH: How did your level of exercise in the last month compare with your usual activity pattern over the past year?

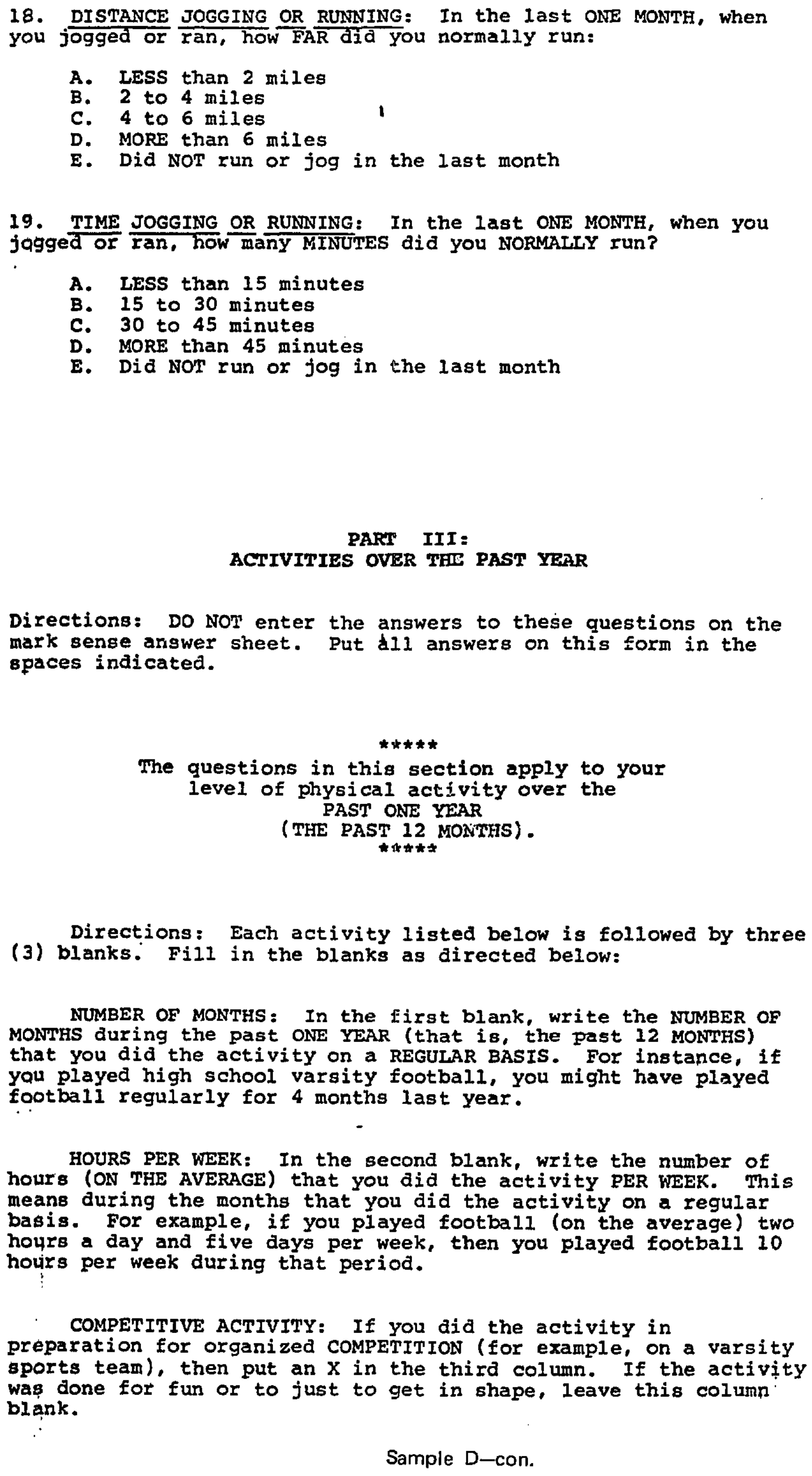
A. I did MUCH MORE exercise in the last month
B. I did MORE exercise in the last month
C. I did about the SAME level of exercise in the last month
D. I did LESS exercise in the last month
E. I did MUCH LESS exercise in the last month

17. JOGGING OR RUNNING: In the last ONE MONTH, how many times did you jog or run (only count the times you jogged or ran for 15 minutes or more without stopping)?

A. Less than 1 time per week
B. Approximately 1 time per WEEK
C. 2 to 3 times per WEEK
D. 4 or more times per WEEK
E. NONE (did not run or jog in the last month)

Sample D

Figure 17. Samples of physical activity questionnaire used in Army fitness research studies—continued.



18. DISTANCE JOGGING OR RUNNING: In the last ONE MONTH, when you jogged or ran, how FAR did you normally run:

A. LESS than 2 miles
B. 2 to 4 miles
C. 4 to 6 miles
D. MORE than 6 miles
E. Did NOT run or jog in the last month

19. TIME JOGGING OR RUNNING: In the last ONE MONTH, when you jogged or ran, how many MINUTES did you NORMALLY run?

A. LESS than 15 minutes
B. 15 to 30 minutes
C. 30 to 45 minutes
D. MORE than 45 minutes
E. Did NOT run or jog in the last month

PART III:
ACTIVITIES OVER THE PAST YEAR

Directions: DO NOT enter the answers to these questions on the mark sense answer sheet. Put all answers on this form in the spaces indicated.

*****

The questions in this section apply to your
level of physical activity over the
PAST ONE YEAR
(THE PAST 12 MONTHS).

*****

Directions: Each activity listed below is followed by three (3) blanks. Fill in the blanks as directed below:

NUMBER OF MONTHS: In the first blank, write the NUMBER OF MONTHS during the past ONE YEAR (that is, the past 12 MONTHS) that you did the activity on a REGULAR BASIS. For instance, if you played high school varsity football, you might have played football regularly for 4 months last year.

HOURS PER WEEK: In the second blank, write the number of hours (ON THE AVERAGE) that you did the activity PER WEEK. This means during the months that you did the activity on a regular basis. For example, if you played football (on the average) two hours a day and five days per week, then you played football 10 hours per week during that period.

COMPETITIVE ACTIVITY: If you did the activity in preparation for organized COMPETITION (for example, on a varsity sports team), then put an X in the third column. If the activity was done for fun or to just to get in shape, leave this column blank.

Sample D—con.

Figure 17. Samples of physical activity questionnaire used in Army fitness research studies—continued.

LEAVE ALL COLUMNS BLANK
for activities that you did NOT do REGULARLY
in the past year.

	How many MONTHS out of the PAST YEAR did you do this as a REGULAR ACTIVITY?	How many HOURS per WEEK did you do this when it was a REGULAR ACTIVITY?	Place an X in this column if the ACTIVITY was done COMPETITIVELY
1. Basketball (non-game)			
2. Stream Fishing			
3. Baseball/Softball			
4. Golf			
5. Volleyball			
6. Calisthenics			
7. Soccer/Lacrosse			
8. Basketball (Game Play)			
9. Racquetball/Squash/Handball			
10. Snow / Water Skiing			
11. Touch Football			
12. Tennis			
13. Ice Skating/Roller Skating			
14. Hunting/Hiking			
15. Swimming (non-competitive)			
16. Bicycling			
17. Aerobic Dancing			
18. Wrestling/Boxing/Martial Arts			
19. Hockey			
20. Competitive Football/Rugby			
21. Gymnastics			
22. Swimming (competitive)			
23. Running			
24. Cross-country skiing			
25. Other: ____________			
26. Other: ____________			

Sample D—con.

Figure 17. Samples of physical activity questionnaire used in Army fitness research studies—continued.

We are interested in your present "normal" level of physical activity. If you have done any of the activities listed below regularly in the last 3 months, circle the activity and write in the number of days per week you did the activity, and how many minutes on the average, you did it on those days. Also, fill in the distance covered where applicable.

ACTIVITY	DAYS/WEEK	MINUTES/DAY	DISTANCE/DAY
Walks or Hikes			
Bicycle Rides			
Swimming Laps			
Running/Jogging			
Calisthenics			
Weight Lifting			
Karate, Judo, etc.			
Tennis, Squash, Raquetball			
Baseball			
Basketball			
Aerobic Dance			
Other			

How hard do you usually exercise? (circle one)

Very Light Average Moderately Hard Very Hard

Sample E

Figure 17. Samples of physical activity questionnaire used in Army fitness research studies—continued.

References

1. Mello, R.P., M.M. Murphy and J.A. Vogel. Relationship between the Army two mile run test and maximal oxygen uptake. US Army Rsch Inst. Envrn. Med. Technical Report T3/85, December 1984.
2. Myers, D.C., D.L. Gebhardt, C.E. Crump and E.A. Fleishman. Validation of the military entrance physical strength capacity test. US Army Rsch. Inst. for the Behav. and Social Sciences, Technical Report 610, January 1984.
3. Zoltick, J.M., H.A. McAllister, and J.L. Bedynek. The United States Army cardiovascular screening program. *J. Cardiac Rehabil.* 4:530–535, 1984.
4. Fitzgerald, P.I. Unpublished data from the US Army Research Institute of Environmental Medicine, 1985.
5. Daniels, W.L., D.M. Kowal, J.A. Vogel and R.M. Stauffer. Physiological effects of a military training program on male and female cadets. *Aviat. Space Env. Med.* 50:562–566, 1979.
6. Knapik, J.J., J.E. Wright, D.M. Kowal and J.A. Vogel. The influence of and women. *Aviat. Space Env. Med.* 51:1086–1090, 1980.
7. Murphy, M.M., J.J. Knapik, J.A. Vogel and F.R. Drews. Relationship of anaerobic power capacity to performance during a 5-day sustained combat scenario. U.S. Army Rsch. Inst. Env. Med Technical Report T5/84, 1984.
8. Patton, J.F., W.L. Daniels and J.A. Vogel. Aerobic power and body fat of men and women during Army basic training. *Aviat. Space Env. Med.* 51:492–496, 1980.
9. Patton, J.F. and J.A. Vogel. An evaluation of physical fitness in the "Pro-Life" program, 2d Infantry Division, Korea. Proc. U.S. Army Science Conf., June 1976, Vol III, West Point, N.Y.
10. Vogel, J.A., J.F. Patton, R.P. Mello and W.L. Daniels. An analysis of aerobic capacity in a large United States population. *J. Appl. Physiol.* 60:494–500, 1986.
11. Vogel, J.A. A review of physical fitness as it pertains to the military services. U.S. Army Rsch. Inst. Env. Med. Technical Report 14/85, July 1985.
12. Taylor, H.L., E. Buskirk and A. Henschel. Maximal oxygen uptake as an objective measure of cardiorespiratory performance. *J. Appl. Physiol.* 8:73–80, 1955.
13. Mitchell, J.H., J. Sproule and C.B. Chapman. The physiological meaning of maximal oxygen uptake test. *J. Clin. Invest.* 37:538–547, 1957.
14. Rowell, L.B., H.L. Taylor and Y. Wang. Limitations to prediction of maximal oxygen intake. *J. Appl. Physiol.* 19:919–927, 1964.
15. Davies, C.T.M. Limitations to the prediction of maximum oxygen intake from cardiac frequency measurements. *J. Appl. Physiol.* 24:700–706, 1968.
16. Astrand, P.-O. and I. Ryhming. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *J. Appl. Physiol.* 7:218–221, 1954.
17. Shephard, R.J. Computer programs for solution of the Astrand nomogram and the calculation of body surface area. *J. Sports Med.* 10:206–210, 1970.
18. Bar-Or, O., R. Dotan and O. Inbar. A 30s all-out ergometric test—its reliability and validity for anaerobic capacity. *Israeli J. Medical Sci.* 13:326–327, 1977.
19. Thorstensson, A. Muscle strength, fibre types and enzyme activities in man. Acta Physiologica Scandinavia, Supplementum 443, 1976.
20. Frederick, F.A., R.C. Langevin, J. Miletti, M. Sacco, M.M. Murphy and J.F. Patton. Development and assessment of the Monark cycle ergometer for anaerobic muscular exercise. U.S. Army Rsch. Inst. Envr. Med. Technical Report No. T6/83, 1983.
21. Ramos, M.U. and J. Knapik. Instrumentation and techniques for the measurement of muscular strength in the human body. U.S. Army Rsch. Inst. Env. Med. Technical Report No. T2/80, 1980.
22. Teves, M.A., J.E. Wright and J.A. Vogel. Performance on selected candidate screening test procedures before and after Army basic and advanced individual training. U.S. Army Rsch Inst. Env. Med. Technical Report No. T13/85, June 1985.
23. McDaniel, J.W., R.J. Skandis and S.W. Madole. Weight lift capabilities of Air Force basic trainees. US Air Force Aerospace Med. Rsch. Lab. Technical Report No. 83–0001, 1983.
24. Borchart, J.W. A cluster analysis of strength tests. *Rsch Quart.* 39:258–264, 1968.
25. Knapik, J.J., J.A. Vogel and J.E. Wright. Measurement of isometric strength in an upright pull at 38 cm. U.S. Army Rsch. Inst. Env. Med. Technical Report No. T3/81, 1981.
26. Knapik, J., D. Kowal, P. Riley, J. Wright and M. Sacco. Development and description of a device for static strength measurement in the Armed Forces Examination and Entrance Station. U.S. Army Rsch. Inst. Env. Med. Technical Report No. T2/79, 1979.
27. Patton, J.F. and A. Duggan. The evaluation of tests of anaerobic power (British) Army Personnel Rsch. Estab. Memorandum Report No. 85M 503, March 1985.
28. McConville, J.T., E. Churchill, T. Churchill and R. White. Anthropometry of women of the U.S. Army—1977. Comparable data for the U.S. Army men. U.S. Army Natick R&D Command Technical Report No. TR–77/029, 1977.
29. Wright, J.E., J.A. Vogel, J.B. Sampson, J.J. Knapik, J.F. Patton and W.L. Daniels. Effects of travel across time zones (jet-lag) on exercise capacity and performance. *Aviat. Space Env. Med.* 54:132–137, 1983.
30. Robertson, D.W. Development of an occupational strength test battery. U.S. Navy Personnel R&D Center Technical Report 82–42, 1982.
31. Marcinik, E.J., J.A. Hodgdon, K. Mittleman and J.J. O'Brien. Aerobic/calisthenic and aerobic/circuit weight training program for Navy men: a comparative study. U.S. Naval Health R&D Center Technical Report No. 84–6, 1984.
32. Sharp, D.S., J.E. Wright, J.A. Vogel, J.F. Patton, W.L. Daniels, J. Knapik and D.M. Kowal. Screening for physical capacity in the U.S. Army: an analysis of measures a predictive of strength and stamina. U.S. Army Rsch Inst. Env. Med. Technical Report No. T8/80, 1980.
33. Knapik, J.J., R.L. Burse and J.A. Vogel. Height, weight, percent body fat and indices of adiposity for young men and women entering the U.S. Army. *Aviat. Space Envir. Med.* 54:223–231, 1983.
34. Patton, J.F., J.A. Vogel, J. Bedynek, D. Alexander and R. Albright. Response of 40 and over aged military personnel to an unsupervised, self-administered aerobic training program. *Aviat. Space Envir. Med.* 54:138–143, 1983.

Chapter **17**

Fitness and Activity Measurement in the 1981 Canada Fitness Survey

Thomas Stephens, Ph.D.
National Center for Health Statistics

Cora Lynn Craig
Canada Fitness and Lifestyle Research Institute
Ottawa, Ontario

Introduction

Survey objectives

The broad objective of the 1981 Canada Fitness Survey was to acquire descriptive population statistics on the topics of physical activity and physical fitness. Four specific objectives were identified when planning began: (a) To acquire reliable data for making cross-sectional comparisons within the Canadian population to support the sponsor's policy and program development; (b) to establish population-based norms for the physical-fitness test battery; (c) to establish baseline levels of fitness against which future comparisons could be made through repetitions of the 1981 survey; and (d) to contribute data to a time series on physical activity dating from the 1976 Survey of Fitness, Physical Recreation and Sport (1).

Although these were the official objectives that motivated the sponsor and guided survey development throughout all its stages, there were other objectives, of a more analytic nature, that had their own impact upon the protocol. Foremost of these was a desire to exploit the unique opportunity presented by the Canada Fitness Survey (CFS) to analyze the association between physical activity and fitness (2) and between physical activity and health status.

Figure 1 is a schematic of the conceptual model adopted to guide development and, subsequently, analysis. The dashed line indicates the boundaries of survey subject matter, although techniques such as record linkage could conceivably be used to extend these boundaries to measures of morbidity and mortality as outcomes of activity and fitness (3).

Context for survey development

The proposal for a Canada-wide survey of population fitness dates from 1972, when the National Conference on Fitness and Health was convened in Ottawa (4). In that same year, the President's Council on Physical Fitness and Sport was conducting a survey of physical activity in the U.S. population (5), a development that must have given impetus to the Canadian proposal.

Note: The success of the Canada Fitness Survey is attributable in large measure to its dedicated field staff, advisors, and consultants. With regard to this paper, we would especially like to thank Claude Bouchard, Phil Campagna, Blake Ferris, Adele Furrie, Norm Gledhill, Gaston Godin, Jo Hauser, Fernand Landry, Francois Peronnet, Art Quinney, William D. Ross, Michael T. Sharratt, and Roy J. Shephard. Their comments on an earlier draft helped ensure that the descriptive sections are accurate; we are, of course, responsible for the interpretations and recommendations.

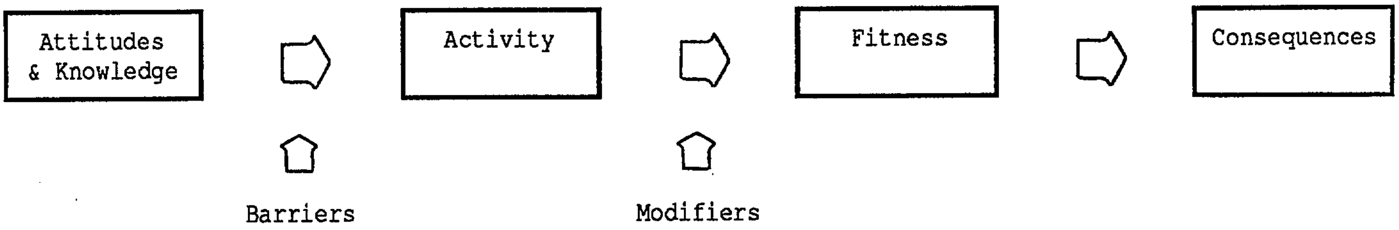


Figure 1. Model of fitness and its interrelationships.

The 1972 conference coincided with two other events that would help create a climate of support for finally implementing the recommended survey 9 years later: One was the achievement of a nationwide system of public insurance for hospitalization and physician care; the second was the conclusion of a Canada-wide survey of nutritional status. With the health insurance system finally in place after years of development, attention began to turn from facilitating access to health care services to promoting the health of the population. Meanwhile, personnel of the Nutrition Canada Survey were reaching the conclusion that "the problem of overweight in adults has to be viewed as a result of calories consumed being in excess of energy expended" (6, p. 121) and went on to decry the reduced opportunities for physical activity in a technologically advanced society. (No data on energy expenditure were collected in the nutrition survey, however.)

Events during the years between survey proposal in 1972 and survey implementation in 1981 had a considerable impact on the final shape of the CFS. In 1976, considerable detail on physical activity patterns was collected from a probability sample of approximately 70,000 individuals through a nationwide questionnaire survey (1). This survey thus provided a point of reference for data on participation levels 5 years later (7). In 1977, a field battery of fitness tests—the Standardized Test of Fitness in Occupational Health (8)—was used to evaluate the physical fitness of over 10,000 working adults, thereby producing both a protocol and provisional norms for use by fitness leaders throughout Canada and, eventually, in the CFS. The cardiorespiratory test in this battery was a version of the Canadian Home Fitness Test (9), a step test designed for self-administration and distributed to the public as part of a package of informational and motivational materials on fitness. This test was used to assess the cardiorespiratory fitness of 3,850 individuals in the Canada Health Survey in 1978–79 (10, chapter 3).

Thus, by the time that serious planning on the CFS was initiated in 1980, there existed a fitness-test protocol developed for field use, provisional norms, experience with in-home fitness testing of a probability sample of free-living adults, and data on population activity patterns. From the outset, it was understood that the CFS would use the Standardized Test of Fitness (as it was called by then) and would supply new norms for the test items. The possibilities of modifying the test battery were thus highly constrained. In contrast, there were few preconditions by the sponsor on the design of the questionnaire.

Financial support and much of the initiative for this project came from Fitness Canada, the agency of the federal government charged with increasing physical activity in the population. As the central agency in a fitness-delivery system sharing responsibility with 10 provinces, 2 territories, and hundreds of municipalities, Fitness Canada was "first among equals" in articulating data needs.

At the time of planning the CFS, it was generally understood that the 1981 survey would be the first of a series conducted at more or less regular intervals. As of this writing, it is proposed to conduct a repetition in 1991 and a longitudinal followup of 1981 participants in the interim. This would conform to an approximate 5-year cycle, a frequency judged appropriate (11) for a survey as comprehensive and costly as the CFS.

Methods

Fitness measurement

General issues

The Canada Fitness Survey was designed to assess the health-related components of fitness rather than athletic fitness (12). This reflected the sponsor's mandate, which extended to the entire population and which grew out of a public health perspective rather than an elite-sports orientation. The specific components of fitness to be assessed were cardiorespiratory endurance, flexibility, muscular strength, and muscular endurance, as well as body fatness.

The design of the CFS protocol was constrained from the outset by several factors, the most significant being the sponsor's requirement that the Standardized Test of Fitness (STF) be the basic protocol. Another factor was a preference for in-home testing in the belief that this would enhance participation; consequent on this was a need for portable equipment and a reasonable duration for the household visit. Fortunately, these requirements were consistent with use of the STF.

One drawback of the STF was the fact that it had previously been applied only to persons ages 17–64. Because the CFS was to assess the fitness of the entire population, procedures were reviewed and adopted and, ultimately, ages 7–69 were tested. These and other minor modifications to the original STF protocol are described in the sections that follow.

The sequence of tests was the same for all survey participants and was organized into three stations: (a) Anthropometry, (b) screening and step test, and (c) grip strength, pushups, trunk flexion, and situps (see appendix A). The average time for one individual to complete the protocol was approximately 1 hour. When two or more persons were being tested during a household visit, each would be taken in turn through Station 1 (anthropometry), as this stage required both members of the two-person testing team. Subsequently, Stations 2 and 3 were each operated by one tester, with Subject B starting at Station 2 after Subject A had moved on to Station 3. Any waiting time was occupied with completion of the self-administered questionnaire. The average household visit, including setup and explanations, screening, testing, questionnaire, and feedback session, consumed about 2½ hours. All testing was by appointment, and subjects were requested to abstain from food, caffeine, and smoking for 2 hours and from alcohol for 6 hours prior to the test appointment.

Anthropometry

The anthropometry procedures of the STF were based on the Durnin and Womersley (13) approach to estimating body fatness from skinfold thickness. These procedures originally consisted of biceps, triceps, subscapular, and suprailiac sites; for the CFS, the medial calf was added. Height and weight were also measured, or estimated if measurement was impossible—a rare event in the survey. Chest, abdomen, and hip girth formed part of the original anthropometry battery. Added to these were five other measurements required for somatotyping by the Heath-Carter protocol (14): Diameter of the humerus and femur, and girth of the upper arm, thigh, and calf.

Considerable care was exercised in site selection, instrument placement, and reading. All measurements were taken by one tester; site locations were verified by the second. Measurements were read aloud, repeated by the second tester, and recorded. The skinfold measurements were performed twice and both values recorded. If the difference exceeded 4 mm, a third measurement was taken, and the mean of the two closest measurements was used as the datum. Tolerances and sites are shown in table 1 and figure 2.

The instrumentation for these procedures was: An Accuweigh platform scale, set square, and modified carpenter's tape for height; Harpenden skinfold calipers; modified Mitutoyo vernier caliper for diameters; and a K + E steel anthropometric tape for girths. All equipment was checked prior to each household visit to ensure that it was in good working order.

With some modifications to permit a single tester to carry out these procedures, and with the exception of the Heath-Carter measurements, this part of the CFS protocol has been incorporated into the Second Edition of the Standardized Test of Fitness (15). The third edition will include the norms from the CFS.

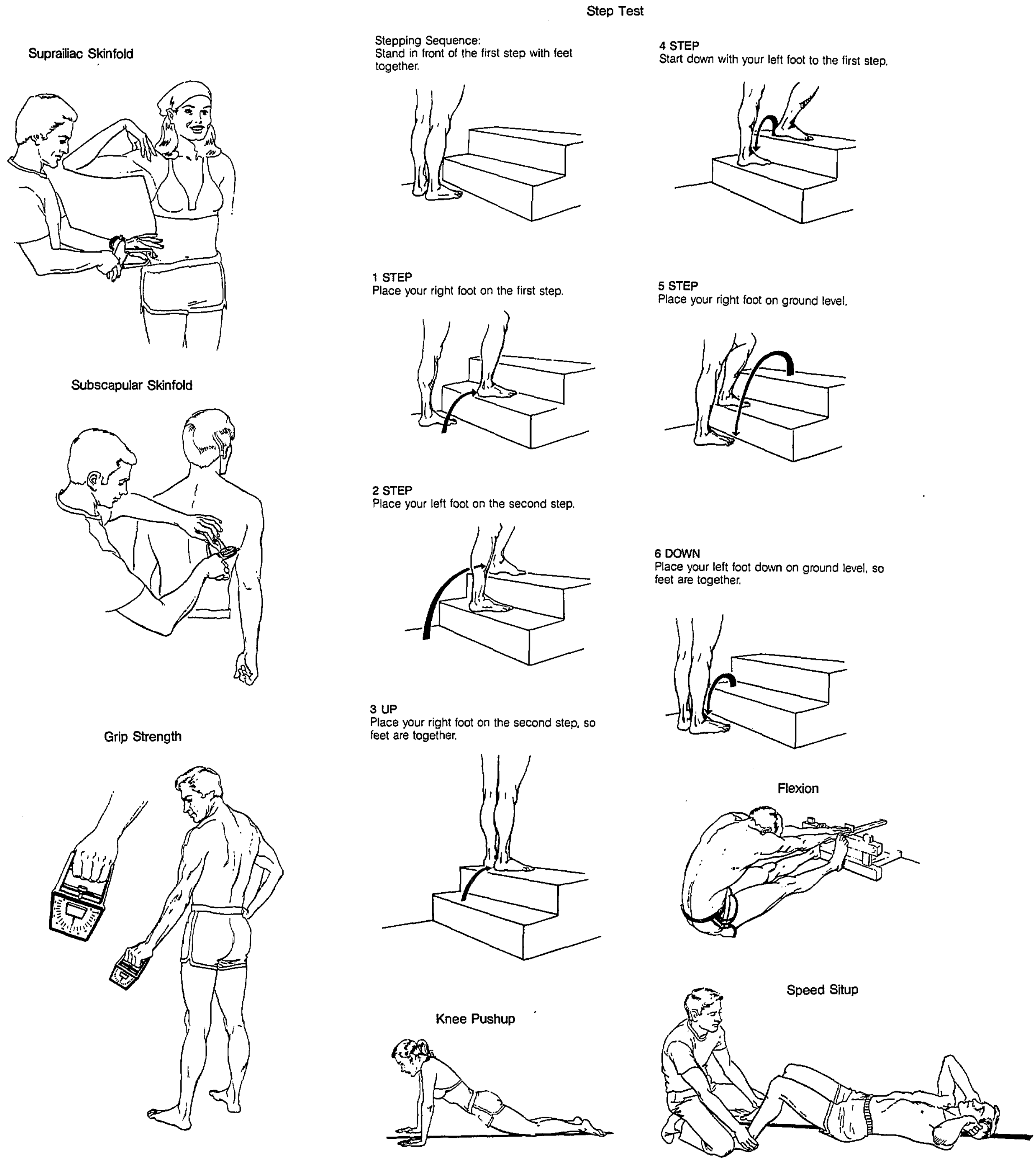
Cardiorespiratory fitness

The centerpiece of the STF is the Canadian Home Fitness Test (now called the Canadian Aerobic Fitness Test). Originally developed for self-administration in the home, the test provides an acceptably accurate prediction for adults of group means of maximal oxygen uptake as measured by both bicycle ergometer (16, 17) and progressive treadmill (9, 18). Its results are little affected by temperature or humidity (19). The principal drawback, noted by one of the original designers (20), is the low accuracy with which test subjects measure their own pulse. This was not a problem relevant to the CFS application and was amenable to training in any event.

The step test used in the CFS was the "Advanced Version" of the Canadian Home Fitness Test, with a maximum of three stages of climbing steps for 3 minutes per stage. The stepping rates at the initial and two

Table 1. Anthropometric procedures

Parameter	Tolerance	Remarks
Weight	0.1 kg	Light clothing, no shoes
Height (standing)	0.1 cm	Stretch with gentle traction
Triceps skinfold	0.2 mm	Midpoint, right side, arm relaxed after site located
Subscapular skinfold	0.2 mm	Lateral to inferior angle of right scapula (figure 2)
Biceps skinfold	0.2 mm	Same landmark as for triceps
Suprailiac skinfold	0.2 mm	Subject holds breath, site 3–5 mm above crest of right ilium at midline of body, fold angles slightly downward (figure 2)
Calf skinfold	0.2 mm	Right foot on step, vertical fold on inside of calf just above point of maximum girth
Elbow diameter	0.5 mm	Right arm at shoulder height, upper arm at 90 degrees
Knee diameter	0.5 mm	Right knee with foot on step, leg forming 90 degree angle
Upper arm girth	0.1 cm	Arm relaxed, same landmark as for biceps and triceps
Chest girth	0.1 cm	At mesosternale, arms relaxed, at end of expiration
Waist girth	0.1 cm	At level of noticeable narrowing or at lateral level of 12th or lower floating rib
Hip girth	0.1 cm	At level of greatest gluteal protuberance
Thigh girth	0.1 cm	1 cm below gluteal line, right side
Calf girth	0.1 cm	At maximum girth, right side



Suprailiac Skinfold

Subscapular Skinfold

Grip Strength

Step Test

Stepping Sequence:
Stand in front of the first step with feet together.

1 STEP
Place your right foot on the first step.

2 STEP
Place your left foot on the second step.

3 UP
Place your right foot on the second step, so feet are together.

4 STEP
Start down with your left foot to the first step.

5 STEP
Place your right foot on ground level.

6 DOWN
Place your left foot down on ground level, so feet are together.

Flexion

Knee Pushup

Speed Situp

Figure 2. Selected procedures. Reproduced from *The Standardized Test of Fitness Operations Manual,* Second Edition. Fitness Canada, Ottawa, 1981.

subsequent stages were adjusted for the subject's age and sex (table 2) and had been calibrated to produce a workload of 65–70 percent of the average maximum for an individual 10 years older, of the same age, and 10 years younger, respectively, for the three stages. For children ages 7–14, the third stage is equivalent to 85 percent of the maximum workload for the subject's age. Adult energy requirements for the slowest to the fastest stepping cadences shown in table 2 range from 4.4 METS (15.5 ml/kg/min) to 10.8 METS (37.7 ml/kg/min) for males and 4.7 METS (16.3 ml/kg/min) to 8.9 METS (21.1 ml/kg/min) for females (21). Stepping was performed on a pair of 20.3 cm steps (figure 2), and cadence was maintained with prerecorded music.

Assuming pretest screening did not show contraindications, the test proceeded with heart rate measurement immediately after each 3-minute stage was completed. For this purpose the tester placed a stethoscope over the heart for a 10-second count; only subjects falling within preestablished guidelines were allowed to continue to the next stage (table 2). The heart rate measured immediately after the final stage is the crucial one; entering this value in a regression equation with weight, height, age, and the energy cost of the last stepping stage produces an estimate of maximum oxygen uptake that correlates well ($r = 0.91$) with the directly measured value for adults obtained on a treadmill (9).

Table 2. Stepping tempo and heart rate limits for 3-stage step test of cardiorespiratory fitness

Age	Tempo [1] and heart rate (HR) [2]	Stage		
		1	2	3
7–10 years	Tempo	114	120	132
	HR	180	186	—
11–14 years	Tempo	120	132	144
	HR	180	186	—
15–19 years	Tempo:			
	Male	132	144	156
	Female	114	120	132
	HR	180	162	—
20–29 years	Tempo:			
	Male	132	144	156
	Female	102	114	120
	HR	174	156	—
30–39 years	Tempo:			
	Male	114	132	144
	Female	102	114	120
	HR	168	150	—
40–49 years	Tempo:			
	Male	102	114	132
	Female	84	102	114
	HR	156	144	—
50–59 years	Tempo:			
	Male	84	102	114
	Female	66	84	102
	HR	150	138	—
60–69 years	Tempo	66	84	—
	HR	144	138	—

[1] Steps/minute; no distinction between males and females is made for the 2 youngest or the oldest age group.

[2] Beats/minute; values at or above those indicated ruled out progression to the next stage of the test.

Children followed the same protocol as adults. After some early trials with different stepping heights and longer stepping sequences, the adult protocol was adopted and the measurement of the final heart rate found to be repeatable ($r = 0.71$ to 0.90). However, it turned out to be a poor predictor of maximum oxygen uptake independent of body mass (22).

Equipment for the step test consisted of two fiberglass stairs of 20.3 cm (8 inches) each with handrail, 110V cassette player and cassettes, stopwatch, and stethoscope. The cadence on the cassettes was checked before each household visit to ensure against changes resulting from stretching or other causes; stopwatches were checked for accuracy.

Muscular strength

A single test of muscular strength (the maximum force of a muscular contraction) was part of the STF and was originally added for motivational reasons as much as for the desire for data on this variable. The measure used, handgrip strength, cannot be claimed as a general indicator of overall muscular strength (23, 24). At best, it represents the strength of the forearm and hand.

Using a handgrip dynamometer, each subject exerted maximum force for two trials with each hand, the arm being held downward and away from the body (figure 2). The best performance for each hand was read to the nearest kilogram and recorded; the sum of these two values was used to represent grip strength. Testers encouraged subjects to do their best.

The equipment consisted of a TKK model 1201 dynamometer with a range of 0–100 kg. Dynamometers were calibrated prior to the survey and checked before each household visit to ensure that the needle returned to the zero position.

Muscular endurance

Muscular endurance is defined as the ability to contract muscles repeatedly without fatigue. The two tests of this attribute were speed situps and pushups, involving primarily the musculature of the abdomen, shoulders, and upper arms.

Two versions of pushups were used: Males pivoted from the toes and females from the knees (figure 2). Although developmental testing had confirmed that preadolescent girls had little trouble with pushups from the toes, it was apparent that mature women did, with too many scores of 0 and 1. Thus, the same protocol was adopted for all females to permit comparisons among age groups.

Both sexes were required to keep the upper body straight and to continue the pushups, without time limit, until they strained visibly or stopped of their own accord. The tester counted aloud in an encouraging manner each time the subject returned to the floor.

Situps were identical for both sexes. The subject began in a supine position, knees bent at 90° and ankles held by the tester (figure 2). With hands behind the head, the subject was required to sit up, touch the elbows to the knees, and return to the mat for each count. The tester provided moderate verbal encouragement in the same fashion as for the pushups. Although a 60-second time limit was imposed, there was no restriction on stopping and restarting during this period.

The only equipment required for the two tests of muscular endurance was a stopwatch and an exercise mat.

Flexibility

The ability to move joints through a wide range of motion is an attribute related to activities of daily life. In the STF, trunk flexion was assessed as an indicator of hip joint flexibility, which is associated with lower back pain (25).

After some slow stretching to loosen up, the subject was seated on the floor with feet 5 cm apart and flush against the crossbar of the measuring apparatus, with knees fully extended (figure 2). Without jerking, the subject bent forward to push a sliding marker along a scale and held the position for approximately 2 seconds. In addition to receiving moderate encouragement, the subject was advised that lowering the head enhances performance. Two attempts were made and recorded to the nearest 0.5 cm; the best was preserved as the datum.

A modified Wells and Dillon sit-and-reach flexometer was used for this test, with a scale value of 25 cm equivalent to touching one's toes. Detailed specifications have been published elsewhere (15, p. 32).

Screening and safety

Although the STF had not been designed for survey purposes, pretest screening was an integral part of the protocol. Shephard (26), reviewing the outcome of approximately 500,000 administrations of the Canadian Home Fitness Test as of 1980, could find no evidence of serious complications, i.e., nothing more serious than a fainting and a sprained ankle. Notwithstanding this record, which included sample surveys of general (10) and employee (8) populations, screening procedures and safety considerations received priority attention in planning the CFS. As a consequence, another 13,000 incident-free test administrations were added to the record.

For screening purposes, three sets of procedures were used. Because these employed an interview format, physiologic assessment, and observation, a variety of risk markers were covered in a fashion that did not rely on a single, fallible method. Screening followed the anthropometry and, if failed, precluded all fitness testing, not just the step test.

It should be noted here that, prior to even the anthropometric measurements, the tests were described in full. Subjects were required to indicate their understanding by signing a consent form that identified as potential risks episodes of transient lightheadedness, fainting, chest discomfort, leg cramps, and nausea. Subjects were told that they could discontinue the tests at any point if they became distressed.

The basic screening tool was a set of seven questions—the Physical Activity Readiness Questionnaire, or PAR–Q (27)—which had accompanied the Canadian Home Fitness Test since its inception. (These questions are shown in appendix A.) Two conditional questions were added to the original seven to increase specificity. Experience with the Canada Health Survey indicated that the question on blood pressure identified past transient hypertension (e.g., during pregnancy) that bore no relation to current status. Similarly, those reporting bone or joint problems were questioned further to confirm that this precluded the climbing of stairs.

The PAR–Q screening questions were asked of all survey participants ages 15 and over by one of the testers. A different set was developed for children because pretesting revealed that the cardiovascular content of the adult questions was not meaningful for younger subjects. A positive answer by a parent to any of the following precluded testing subjects ages 7–14:

(a) Is ______ limited for health reasons from doing strenuous physical activity at school and with friends?
(b) Has ______ now returned to normal activity at school and with friends, with no restrictions? (This question was asked only if the parent indicated that the child had been hospitalized or under a physician's care within the last year.)
(c) Is there any reason why ______ should not do moderately strenuous exercise such as climbing stairs, pushups, and situps?

These questions were administered during a 5-minute rest that followed the anthropometry. At the conclusion of this period, heart rate and systolic and diastolic blood pressure were measured by the tester. One of three cuffs (child, adult, extra-large) was used, as appropriate for the subject, and phases I and IV were recorded. If either blood pressure reading or the heart rate exceeded the allowable limits, a second test of readings was taken after an additional 5-minute rest. If any of the limits was still exceeded, the subject did not proceed to the fitness testing. Limits were as follows:

- Systolic BP—150 mm.
- Diastolic BP—100 mm.
- Heart rate—100 bpm.

The third set of pretest screening procedures was based on observation of signs indicating that the subject would be at risk during testing. Grounds for exclusion were fever, persistent cough, muscular coordination or orthopedic problem, drunkenness, and pregnancy (asked, not merely observed).

Despite the rigorous screening, other precautions were taken to minimize risk to the subject, and some of these were an integral part of the step-test protocol, particularly its submaximal quality and progressive format (26). Testers were instructed to discontinue the stepping if the subject began to stagger; complained of dizziness, extreme leg pain, nausea, or chest pain; or showed facial pallor. No such interruptions were required.

As a final precaution, all testers were required to hold current certification in cardiopulmonary resuscitation and were given a short course in first aid during their training. Emergency health care facilities for each community were identified in advance, and detailed instructions were provided for handling minor, serious, and grave incidents. As it happened, the most serious event associated with 14,000 tests was a fainting.

Quality control

A variety of techniques were used to maximize the quality of the data in addition to the calibration of equipment, narrow tolerances, and repetition of measures already described. These included selection, training, and supervision of staff; quality assurance procedures; and field edits.

Field staff were typically experienced in fitness testing prior to being hired, most being recent graduates of 4-year university programs in the exercise sciences or recreation. To the maximum extent possible, they were hired from the 80 communities in which the CFS was conducted. They worked in pairs, an arrangement that permitted the team approach to anthropometry described earlier and that also contributed to respondent relations because most teams had one member of each sex.

Field staff were assigned to one of 11 geographic groupings and reported to a "regional supervisor," typically a master's level graduate in exercise science. Each of these supervisors had access to a designated university professor who provided guidance as required on scientific matters. Administrative direction was received from the central office of the CFS organization. Supervisors visited each field staff member at least twice during the data collection period to observe a household visit and to permit the fitness tester to carry out the anthropometric procedures on a known subject (the supervisor). Regional supervisors reviewed all data forms received from testers under their guidance before forwarding these to the central survey office.

Training of staff was organized into two phases: (a) The academic resource persons and regional supervisors were intensively briefed over 7 days on all survey procedures by central office staff and expert consultants; (b) they, in turn, trained their own testers on a regional basis over 9 days. This second level of training was observed by central office staff. Among other techniques, videotaped instructions were used to teach the anthropometric procedures, and practice visits to volunteer households were completed. All field staff met on a regional basis 2–3 months after the data collection began in order to review procedures and identify problems.

Careful editing of all data forms was carried out by all testers immediately after each visit and was repeated by the supervisors. This served to ensure that documents were complete and legible, all measurements were in metric units, activities were coded, and the correct documents were matched for each individual. Prior to data capture, these edits were checked in the central office and additional coding was completed. All forms were transferred directly onto computer tape using a key edit process that was 100-percent verified. Range checks and correlation edits were then performed, and duplicate records were removed.

Feedback and counseling

For a variety of reasons, it was seen as desirable and necessary to provide every test subject with a record of test results, suitably interpreted. This began as a natural consequence of adopting the STF as the survey protocol because it concluded with an individual assessment report and an exercise prescription (28). However, there were additional reasons, one based on ethical considerations and the other related to the opportunity for fitness promotion.

The measurement of blood pressure, in particular, placed an ethical obligation on the tester who discovered an unsuspecting hypertensive. For this purpose, testers were issued an interpretive guide (appendix B), adapted from the one used for the Canada Health Survey.

Related to this question of medical ethics was the fact that most subjects were curious to understand their test results and many were anxious to know how they could improve them. (For their part, the testers were eager to provide such advice and many found this the most rewarding part of the household visits.) To accommodate this concern, the STF assessment report was revised and expanded into an informative results booklet completed at the conclusion of each household visit and left with each subject. This booklet contained a record of the raw test scores and estimates of percent body fat and oxygen uptake; provided percentile scores and ratings (excellent through poor), facts on each test parameter, smoking, drinking, and stress; and listed addresses for detailed resource materials. Using this booklet as a guide, testers were prepared, depending on the subject's interest, to discuss a personal fitness program, identify local recreation resources, and supply informative pamphlets on various aspects of health and fitness.

Activity assessment

General issues

In contrast to the testing of fitness, the sponsor of the CFS had relatively general requirements for data describing the leisure-time activity of the Canadian population. These might be described in generic terms as (a) *participation rates,* or counts of individuals involved in a defined activity, and (b) *activity patterns,* especially dimensions of frequency and duration of activity. A third level of activity assessment, namely, the calculation of individual *energy expenditure* values, was initially regarded by the sponsor as desirable but not essential. However, the protocol did ultimately permit such calculation, and a strong argument was made in its favor (2).

Although the protocol for activity assessment was not predetermined, there were two precedents that influenced its design. One was the 1976 Survey of Fitness, Physical Recreation and Sport (1), which had been conducted 5 years earlier for the same sponsor. Although this survey had some methodological shortcomings, which have been documented elsewhere (11), it did produce data on participation rates and served as a useful point of comparison. In the Canada Health Survey, also, information on physical activities had been collected from a general population sample in the context of a comprehensive survey on health (10, 29) and useful logistical and substantive lessons were provided. In both surveys, a self-administered questionnaire format had been used, and data were gathered on frequency and duration for a wide range of leisure-time activities.

The principal constraints facing the designers of the questionnaire were the need to collect data from a wide age range, in the home, within a reasonable time period. As it turned out, a self-administered questionnaire was completed by 21,738 individuals 10–97 years of age. With few exceptions, these were administered during the fitness test visit and consumed an average 20–25 minutes of respondent time.

Protocol for activity assessment

The principal model for the CFS was the Minnesota LTA Questionnaire of Taylor and colleagues (30). The basic approach was to identify leisure-time activities and determine frequency, average duration, and intensity of participation for periods of up to 1 year.

Recognizing the difficulty of reporting accurately for this lengthy a period, the designers modified the Minnesota LTA approach by using not one but three time periods for leisure activities: (a) Weekly activities over the last 3 months, (b) other activities in the past month, and (c) other activities in the past year. In addition, daily work activities were assessed with a single question.

The rationale for this apparent complication was twofold. First, it seemed worth while to identify activities for which a "training effect" could be expected because fitness levels were also being assessed, hence the inclusion of activities done on a regular basis over the last 3 months. Second, it seemed realistic to expect more accurate reports on recent and regular activities and less on distant and irregular ones. Thus, data on occasions per month, average time spent, and intensity were sought for weekly and past month's activities. In contrast, other activities in the last year were described only in terms of total occasions, months done, and average duration (four categories). Appendix C shows the first three pages of the questionnaire.

The flashcard provided to the respondent is shown in appendix D. This was used to establish the set of activities of interest and to provide codes for any not listed on the questionnaire. The inclusion of home-related activities was intended to permit a calculation of *total* energy expenditure.

Daily energy expenditure per unit of body weight can be calculated from questionnaire responses using the formula:

$$\text{kcal/kg/day} = N_i \times D_i \times \text{METS}_i \times 365$$

where N_i = number of occasions of activity i in the past 12 months
D_i = average duration in hours of that activity
METS_i = a constant value for the energy requirement of that activity at a given intensity

One MET is treated as 1 kcal/kg/hour in the calculation—an average of 75 kcal/hour for a man of 75 kg. This approach can be used to calculate leisuretime energy expenditure only, or an analogous method can be extended to work, sleep, and household chores because the duration of each of these and the nature of work and chores were also obtained by the questionnaire.

For this calculation to be meaningful, accurate questionnaire responses and energy constants are both required.

In the absence of any agreed-upon criterion measure, the questionnaire has been only indirectly validated by virtue of a positive association between activity level and: (a) Five fitness test results, (b) three measures of health status, and (c) three other health-related behaviors (31, pp. 23–34). Leisure-time energy expenditure has also been shown as higher among the young, males, and professional and managerial groups (32), which is consistent with the results of several other general population surveys (11) in the United States and Canada using a variety of questionnaire techniques. (See also tables 10 and 14 in this paper.) Finally, there is a positive association between energy expenditure assessed by this questionnaire and cardiorespiratory fitness, even when age and smoking status are controlled (33).

The energy requirements for the 120 activities uniquely coded by the CFS were established by an expert group based on review of the literature, especially a similar undertaking sponsored by Kino-Quebec (34) 3 years earlier. METS values were established by consensus for each activity at three intensity levels; these were reviewed and finalized after some weeks of reflection. The results of this exercise and the principal references used are shown in appendix E.

Other questionnaire topics

Questions for establishing participation rates and calculating energy expenditure values consumed only 3 of 11 pages of the CFS questionnaire. Other topics were also related to activity (reasons, barriers, importance, activities dropped or sought after, place, time, companions) and to other health behaviors (smoking, alcohol use, breakfast, sleep—in the fashion of Breslow (35) and colleagues). Health status was assessed with questions on activity limitation, Bradburn's Affect Balance Scale (36), and a self-rating. An extensive sociodemographic description of respondents was obtained, extending to confirmation of blood relationships among generations and twin relationships of siblings. The complete questionnaire has been published previously (31, 37).

Sample and survey design

Household survey

The 1981 Canada Fitness Survey was intended to provide descriptive statistics for the Canadian population. For this purpose, Statistics Canada was contracted to design a household-based sample that would strike a balance between data quality and cost of collection. The result was a multistage, clustered probability sample very similar to that of the Canada Health Survey (10, appendix II).

At the outset, it was determined that a sample of approximately 25,000 persons was needed for acceptably precise descriptive statistics at the level of detail required for planning. For example, approximations were made for the distribution of step-test results for 12 age-sex groupings, and the sample size requirements were calculated from the smallest of these expected frequencies. At the same time, it became evident that children would have to be oversampled in order to obtain the single-year norms on the fitness tests that were desired; this feature is described in the next section.

Under the Canadian Constitution, the 10 provinces have responsibility for the delivery of fitness and recreation programs. However, their populations range from 160,000 to 7.5 million, indicating that an exact tenth of the sample should probably not be allocated to each province. The compromise was to allocate the total sample to each province proportional to the square root of its population size. The ratio of the sample in the largest compared with the smallest was thus shrunk to 9:1 from a ratio of 47:1 for the populations themselves.

Within each province, the sample was allocated to three strata—major cities, other ubran, and rural—in proportion to their populations. Clusters were then identified for each stratum. These were geographic areas that were sufficiently compact that a tester team could conduct its fieldwork with a maximum of 1½ hours of driving each way and no overnight travel. Each of the 80 clusters nationwide consisted of 168 households, for a total of 13,440 households—the final sampling units—which were selected at random using an area frame.

The fieldwork was designed to span three seasons and ran from February through July 1981. During each of these six months, every tester team was assigned 28 households. This had been established during pilot testing as a reasonable full-time workload.

In order to produce population estimates, each survey respondent was assigned a weight using a post-stratification ratio-estimation procedure. With the 1981 Census of Canada as the point of reference, information on the age, sex, province, and stratum of each respondent was used to calculate the appropriate sampling weight. For questionnaire data, the average weight was 954; for the fitness tests, it was 1,265. This procedure is necessarily based on those persons in the sample for whom this information was obtained. Thus, the residents of the 1,556 dwellings in the sample of 13,440 (12 percent) who were either not contacted or not persuaded to participate are treated as being not systematically different from the survey participants.

As a result of applying the sampling weights, 97 percent of the Canadian population is represented by the survey responses. The remaining 3 percent are individuals who are institutionalized or in collective households. Because this is a small and varied population (residing in hospitals, nursing homes, university dormitories, prisons, and Armed Forces bases), its omission should have slight impact on the accuracy of the population estimates.

Further information on the allocation of the sample, confidence limits for estimated proportions, use of the sample weights, and other details have been published elsewhere (37, 38).

School survey

The sponsor's need for sex-specific fitness norms for single years of age required 7,000 tests for ages 7–18, compared with the actual number of 4,500 obtained in the household survey. To achieve the additional 2,500 tests, a supplemental survey was conducted in 186 schools in nine provinces during April and May 1982. The fitness testing protocol was identical in all important respects to that used in the household survey; no questionnaires were administered to this supplemental group.

Results

Response

Because one of the primary objectives of the CFS was to provide reliable descriptive statistics on the physical activity and physical fitness of the Canadian population, the representativeness of the survey participants is a central concern. Table 3 shows that 88 percent of households contacted agreed to participate in the survey, and 77 percent of the eligible individuals in these households (everyone over age 6) actually completed a questionnaire, agreed to be tested, or both. Part B of table 3, which is restricted to those eligible for both survey components (questionnaire and tests) by virtue of age, shows that persons were more likely to omit the testing if they did not participate fully in the survey. The

Table 3. Response to Canada Fitness Survey

A. Households			
Sample selected	13,440	100%	
Participating	11,884	88.4%	
Not participating	1,556	11.6%	
B. Individuals in participating households			
Eligible (ages 7+)	30,652	100%	
Participating	23,400	76.6%	
Not participating	7,252	23.3%	
	Total	Male	Female
Participants eligible for both components (ages 10–69)	20,692 100%	9,301 100%	11,391 100%
Completed both components	14,365 69.4%	6,777 72.9%	7,588 66.6%
Completed questionnaire only	5,815 28.1%	2,229 24.0%	3,586 31.5%
Completed tests only	512 2.5%	295 3.2%	217 1.9%

questionnaire was completed by 98 percent of eligible participants; in contrast, the tests were agreed to by only 72 percent. Females in these households were more likely than were males to opt out of the tests and to complete only the questionnaire. (Note that these figures for test subjects include those eventually screened out.)

Because there were different age criteria for the two survey components, one can consider that the survey reached two nonindependent samples. Table 4 shows how these two samples compared with the population, as described by the 1981 Census of Canada.

A value of 1.00 in each cell of table 4 would mean that the samples were perfectly self-weighting, and no adjustments would be necessary for estimating population proportions. As it turned out, however, males were underrepresented and females overrepresented, proportionally. For the tests there was a clear tendency for both sexes to opt out in increasing proportions with increasing age. For the questionnaire, however, it was the middle ages of males who were underrepresented and only the oldest age group of females.

Table 4. Survey participants versus population [1]

	Sampling weights, adjusted to 1.00 for total [2]			
	Questionnaire		Tests	
Age	Male	Female	Male	Female
All ages	1.10	.91	1.04	.96
7–9 years	—	—	.70	.76
10–19 years	1.02	.92	.86	.83
20–29 years	1.13	.88	1.08	.96
30–39 years	1.12	.87	1.03	.90
40–49 years	1.20	.91	1.15	1.01
50–59 years	1.29	.96	1.43	1.18
60–69 years	1.04	.91	1.46	1.41
70 years and over	.96	1.07	—	—
Mean weight	1,054.64	872.32	1,319.15	1,214.71

[1] Based on 1981 census.

[2] A value of 1.00 indicates a perfect match between the proportion of survey participants in an age-sex group and the corresponding proportion in the population. A value greater than 1.00 indicates the group was *underrepresented* in the sample; a value less than 1.00 indicates *overrepresentation*.

For eligible members of participating households, there were four possibilities in addition to completing the questionnaire: (a) Undergoing the fitness testing, (b) agreeing to testing but being screened out, (c) refusing the more strenuous testing after submitting to the anthropometry, and (d) refusing all tests and measurements. (As noted earlier, informed consent was required at the outset of the visit, ensuring that participation —and refusal—was reasonably well informed.)

Table 5 shows how household members in several age-sex groups were distributed among these possibilities. The proportion of participants (tested plus screened) ranged from over 95 percent in the youngest groups to 50 percent or less in the eldest. Selective refusals of the aerobic and subsequent performance tests were infrequent (under 2 percent) and greatly overshadowed by blanket refusals of the entire fitness component (27 percent overall). As also seen in table 3, females were

Table 5. Response to fitness tests by participating household members ages 7–69 years

Age and sex	Refused all tests	Refused aerobic test	Screened	Tested	Total *n* (100 percent)
	Percent				
All ages					
Both sexes	26.8	1.9	12.8	58.6	21,841
Male	22.5	1.9	11.6	64.1	9,914
Female	30.1	1.9	13.9	54.1	11,927
7–9 years					
Male	0	3.9	5.4	90.7	614
Female	0	3.9	3.4	92.7	536
10–19 years					
Male	14.8	1.4	3.7	80.1	2,342
Female	19.2	1.6	5.7	73.5	2,439
20–29 years					
Male	22.7	1.7	7.4	68.2	2,131
Female	31.5	1.9	9.8	56.8	2,714
30–39 years					
Male	20.4	2.0	8.8	68.8	1,778
Female	29.0	2.0	12.4	56.6	2,248
40–49 years					
Male	23.7	2.1	15.5	58.6	1,177
Female	33.0	1.9	18.6	46.5	1,502
50–59 years					
Male	33.9	2.1	26.9	37.1	993
Female	39.1	2.0	25.7	33.2	1,356
60–69 years					
Male	47.7	1.6	30.0	20.7	879
Female	51.9	1.4	29.2	17.6	1,132

more likely to refuse the fitness tests than were males; this difference ranged from 4 to 9 percent and was true of all ages. Older women were especially likely to shun the testing.

A striking trend seen in table 5 is the increasing proportion screened out of testing, corresponding to an increase in age. Although only 13 percent of all test subjects were screened out, the proportion approaches one-third in the oldest age group. With the effect of refusals and screening combined, only about 20 percent of those in their sixties actually underwent fitness testing.

Because virtually all survey participants ages 10–69 completed a questionnaire (table 3), it is possible to compare questionnaire responses for those who agreed to be tested with those who refused and thereby to examine whether there is any bias in the pattern of refusals. Table 6 shows the distribution of active (3 + kcal/kg/daily leisure-time energy expenditure) versus sedentary (1.5 kcal/kg/day) individuals across three categories of test outcome. Refusers, by their own description, are much more likely to be sedentary than active, by a ratio exceeding 3:1. There is a similar difference for those screened out, and those tested have a higher proportion classified as active. These relationships are most pronounced in the age group 40–54 and least for those ages 10–19 (data not shown).

In contrast, refusers and test subjects describe their own fitness level in similar terms. Only those screened out give clearly negative ratings to their own fitness level. (In perhaps half of these cases, the questionnaire completion would have followed the actual screening in the sequence of the household visit.)

Because of the organization of the survey protocol, data on blood pressure, resting heart rate, and anthropometric parameters were provided for all but total refusers. Thus, comparisons are possible among test subjects, those screened out, and those who refused the aerobic and subsequent performance tests. These results are summarized in table 7.

The resting heart rates of test subjects are consistently lower than those of persons who refused or were screened out—by 2–6 bpm. Among males and females, younger refusers have resting heart rates that fall between those of the screened and the tested; the heart rates of older refusers are the highest of the three groups.

Table 6. Activity level and self-rated fitness, by response to fitness tests among persons ages 10-69 years

Activity level and self-rated fitness	Total	Refused	Screened	Tested
	Percent			
Activity level				
Sedentary	55.1	66.4	61.5	47.6
Active	24.5	18.2	19.0	29.1
Self-rated fitness [1]				
More fit	15.7	16.4	12.2	16.2
Less fit	20.2	20.4	27.5	18.4

[1] Compared with others of the same age and sex.

A similar pattern exists for both systolic and diastolic blood pressure for both sexes: The refusers are only marginally higher than those tested, but those screened out have substantially higher values, by up to 16 mm for systolic and 10 mm for diastolic blood pressure. The patterns are similar for males and females, and the differences tend to be more pronounced among older age groups.

Resting heart rate and the two blood pressure readings were two of the criteria for screening; anthropometry results did not figure in screening decisions, however. The values for body mass index (weight/height^2) in table 7 reveal very small but consistent intergroup differences: In every case, persons screened out have higher values (more overweight) than the refusers, who in turn have higher values than the test subjects.

Selective refusal of the fitness tests by those who are chronically inactive or ill is one possible source of bias. Another is a systematic relationship between refusing and time of year because the population is more physically active in summer and late spring than in winter and early spring (figure 3). Table 8 shows that, as survey operations proceeded from February to July, the proportion of total test refusals grew steadily. Because the proportion of ineligibles (ages 0–6) was fairly constant, this means that the number of survey participants declined steadily, although it was designed to remain constant. Although true in both urban and rural strata, this pattern was more pronounced in the former, which accounted for two-thirds of the total sample. This decline in survey participation coincided with (and may even have been caused by) an increase in overall physical activity and, presumably, physical fitness.

Screening and incidents

As noted earlier (table 5), approximately 13 percent of eligible survey participants were excluded from the fitness tests on the basis of one or more screening procedures. In the older age groups, the proportion of willing test subjects screened out was substantial—exceeding 20 percent of men and nearly 30 percent of women beginning at ages 40–49.

Of the three screening techniques used, the PAR–Q questions accounted for most of those eliminated (table 9). The most productive of the 17 specific criteria used were self-described hypertensive currently under medication (6 percent), history of heart trouble (5 percent), and frequent heart or chest pains (5 percent). Measured systolic blood pressure exceeding 150 mm also resulted in the screening out of 5 percent of all participants. Despite the variety and number of procedures, most individuals screened out failed only one or two criteria.

By the time the survey was completed, 16,027 individuals ages 7–69 had agreed to participate in the fitness tests. Screening eliminated 2,807 of these, and the balance of 13,220 submitted to the battery of tests. Only three incidents occurred during this experience: One subject fainted after the step test and recovered without complication, another tripped and fell but was not injured, and a third received assistance for a strained muscle.

Fitness and physical activity

Although the focus of this paper is on the methodology of the CFS, some basic descriptive statistics are presented as well. Apart from the fact that these substantive findings are useful for evaluating survey procedures, they may be of interest because they are based on a nationwide probability sample of the free-living population. Some of these data are published here for the first time, but the survey results have in general been amply documented (7, 31–33, 37–43). Tables 10–13 are based on weighted data.

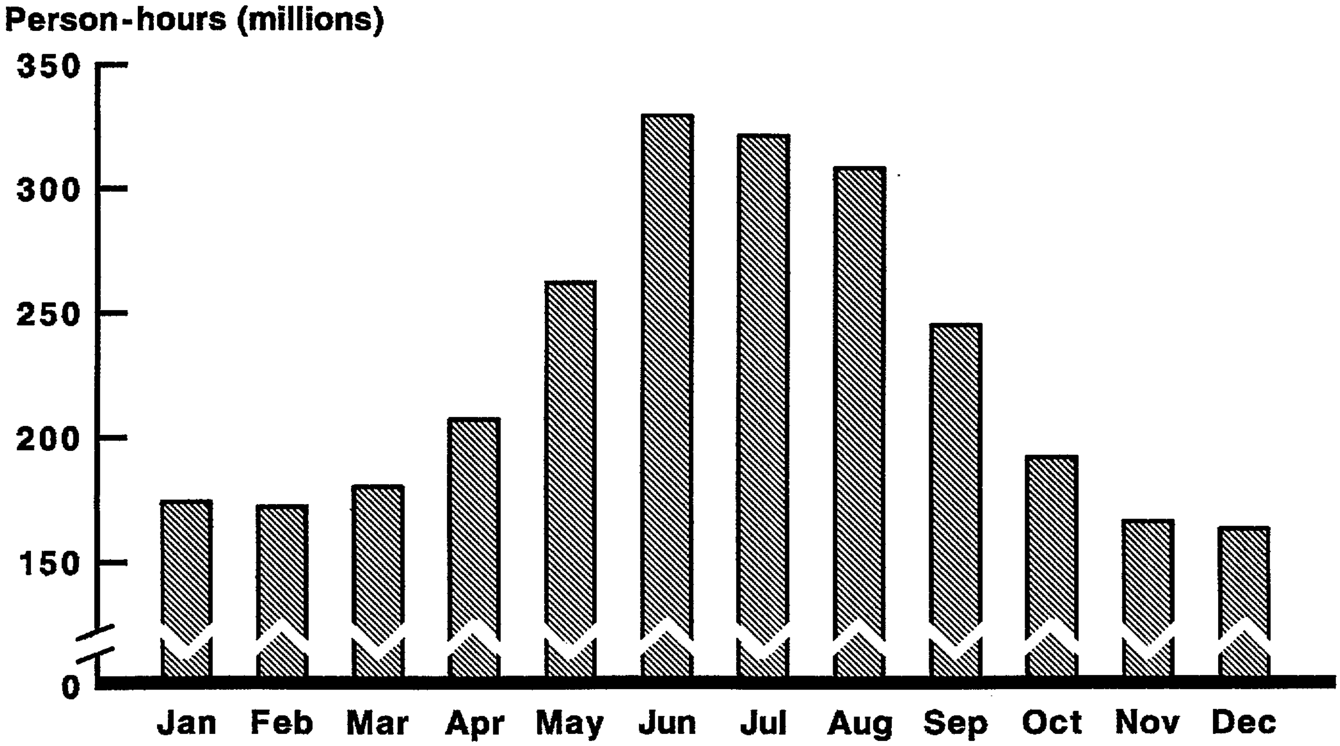
Table 10 shows the values of six fitness parameters for ages 7–69 and energy expenditure values for ages 10 and over. Four percentile values are provided in table 10; more detailed tables with norms are also available (43).

Table 7. Physiologic health indicators, by response to aerobic fitness test

Sex, health indicator and age	Refused		Screened out		Tested		Total	
	x	SD	*x*	SD	*x*	SD	*x*	SD
Male								
Resting heart rate [1]:								
7–19 years	78.7	12.5	83.1	17.6	72.8	11.1	73.3	11.6
20–39 years	71.4	9.2	71.8	13.6	67.2	9.9	67.7	10.3
40–54 years	75.6	9.4	71.3	11.9	67.3	10.0	68.5	10.6
55–69 years	71.1	11.8	69.8	10.8	66.8	8.9	68.4	10.1
Systolic BP [1]:								
7–19 years	107.5	17.2	112.6	20.8	108.2	13.0	108.4	13.5
20–39 years	122.6	16.6	130.7	18.6	120.7	11.9	121.5	13.0
40–54 years	125.0	13.5	137.9	21.0	122.6	11.4	126.1	155.5
55–69 years	128.8	12.4	142.9	21.1	127.2	13.1	134.8	19.0
Diastolic BP [1]:								
7–19 years	69.8	11.9	72.1	10.4	70.9	9.3	70.9	9.4
20–39 years	77.6	11.3	83.6	13.5	78.3	9.4	78.7	9.9
40–54 years	82.7	10.4	90.8	13.8	81.8	8.4	83.8	10.6
55–69 years	80.5	9.7	88.1	12.6	80.8	9.0	84.3	11.5
Body mass index [2]:								
7–19 years	19.1	3.8	19.5	3.8	19.2	3.0	19.2	3.1
20–39 years	24.7	4.0	25.3	3.8	24.6	3.2	24.6	3.3
40–54 years	26.1	3.2	26.9	3.8	26.0	3.2	26.3	3.4
55–69 years	26.5	3.5	26.8	4.0	25.9	3.2	26.4	3.7
Female								
Resting heart rate [1]:								
7–19 years	81.5	12.7	83.5	18.7	76.9	10.8	77.3	11.5
20–39 years	72.5	10.0	74.3	11.6	70.7	9.5	71.2	9.9
40–54 years	72.0	11.9	72.7	11.8	69.2	8.8	70.2	9.9
55–69 years	71.3	6.5	70.5	10.5	68.0	8.5	69.4	9.6
Systolic BP [1]:								
7–19 years	105.8	16.5	110.1	12.6	105.2	11.3	105.4	11.6
20–39 years	111.9	11.0	114.1	16.0	110.7	11.2	111.2	11.9
40–54 years	118.5	13.6	130.0	20.1	117.4	12.9	120.7	16.1
55–69 years	124.7	12.9	141.0	20.6	123.9	12.5	132.7	19.1
Diastolic BP [1]:								
7–19 years	69.3	11.2	73.9	11.1	69.7	9.3	69.9	9.5
20–39 years	73.8	9.1	75.4	12.6	73.0	9.3	73.3	9.8
40–54 years	78.0	10.1	84.5	12.7	76.8	9.4	78.8	10.9
55–69 years	80.2	8.7	86.4	11.5	78.1	8.8	82.4	11.1
Body mass index [2]:								
7–19 years	19.1	3.4	20.7	3.4	19.3	3.1	19.4	3.2
20–39 years	23.4	4.6	24.3	5.2	22.5	3.3	22.8	3.7
40–54 years	25.3	3.9	26.6	5.3	24.3	3.7	25.0	4.4
55–69 years	25.8	3.9	27.0	4.7	25.2	3.5	26.2	4.3

[1] Second reading, if taken.

[2] Weight in kilograms/height in meters ^{2}.



*** walking, swimming (pool), calisthenics, bicycling, jogging, bowling, tennis, golf, softball, dancing-the top ten activities in North America (11)**

Figure 3. Monthly participation in 10 popular activities,* for persons ages 10 years and over: Canada, 1981.

*Walking, swimming (pool), calisthenics, bicycling, jogging, bowling, tennis, golf, softball, dancing—the top 10 activities in North America (11).

A general deterioration in scores is clearly associated with age, beginning in late adolescence (ages 16–19) or ages 20–29 at the latest. Sex differences for the fitness parameters are in the expected direction—females do better in the flexibility test and males perform better in the remainder. Males, except for those ages 10–12, also have higher daily energy expenditure levels.

Energy expenditure values were calculated as described earlier and produced the distribution for age-sex groups shown in table 11. A criterion of 3 + kcal/kg/day, which has been identified as the minimum required to increase aerobic capacity and associated cardiovascular health (44), results in an estimated prevalence of active individuals ages 10 and over of 25 percent. This ranges

Table 8. Response to fitness tests, by month and stratum

Stratum	Total	February	March	April	May	June	July
				Percent			
Urban							
Total [1]	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Participated [2]	67.0	73.8	68.5	64.4	62.8	62.6	56.6
Refused	26.9	20.1	25.3	29.7	29.3	31.6	36.6
Ineligible (age)	6.0	6.0	6.2	6.0	7.9	5.8	6.8
Rural							
Total [1]	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Participated [2]	70.8	75.6	70.4	69.6	67.4	64.6	60.8
Refused	21.5	17.9	21.6	24.0	23.2	25.4	29.2
Ineligible (age)	7.6	6.5	8.1	6.4	9.4	9.9	10.1

[1] Of those participating in some part of the survey.

[2] Tested and screened.

from a low of 15 percent for women ages 40–54 to a high of 45 percent for males 10–19 years. The least active age group among males is also not the eldest but those ages 40–54. This may be evidence of a "survivor phenomenon." The sedentary population (under 1.5 kcal/kg/day) averages 57 percent.

Figure 3 shows the month-to-month variations in person-hours of activity. Person-hours were calculated as the product of the number of participants and their monthly hours of participation. Both of these values show the same variations as seen in figure 3: Most participants and most hours of participation in June through August and least in the winter. Compared with the annual average, monthly figures for hours of participation vary from − 17 to + 22 percent; for numbers of participants, the range is − 16 to + 23 percent (table 12).

Although these are the sort of descriptive data that the CFS was primarily designed to produce, the analytical possibilities in a data set with 23,400 cases and 1,034 variables are also considerable. One example is table 13, which shows the relationship between leisuretime energy expenditure and cardiorespiratory fitness for smokers and nonsmokers in two age groups. This table clearly reveals the beneficial effect of exercise, the detrimental effect of smoking, and the relatively good performance of younger adults. There is also evidence here that smoking attentuates the benefits of exercise, as the differences between the sedentary and most active smokers are

Table 9. Reasons for being screened out of the fitness tests, by age and sex

	All ages		7–19 years		20–39 years		40–54 years		55–69 years	
Reason	Male (*n* = 9,914)	Female (*n* = 11,927)	Male (*n* = 2,956)	Female (*n* = 2,975)	Male (*n* = 3,909)	Female (*n* = 4,962)	Male (*n* = 1,692)	Female (*n* = 2,185)	Male (*n* = 1,357)	Female (*n* = 1,805)
PAR-Q,[1] other questions					Percent					
Doctor ever indicated heart trouble	2.9	2.5	.8	1.0	1.4	1.5	3.7	3.7	10.9	6.3
Frequent heart and chest pains	2.1	2.8	.3	.4	1.7	2.6	3.5	5.2	5.6	4.3
Often feel faint or severely dizzy	1.5	3.2	.3	.8	1.1	3.3	2.3	4.9	4.2	4.9
Hypertensive, under current medication	2.4	3.2	—	.1	.9	.7	4.8	5.6	8.8	12.2
Joint trouble, susceptible to stair climbing	1.7	2.4	.2	.5	.8	1.2	2.7	3.6	6.0	7.2
Over 65, unused to exercise	.6	.7	—	—	—	—	—	—	4.6	4.7
Any other reason for caution	3.0	3.7	1.4	1.7	1.7	2.6	5.0	5.8	7.8	7.6
Children—recently ill and not back to normal	.1	.1	.5	.6	—	—	—	—	—	—
Pregnant	—	.9	—	.2	—	2.0	—	—	—	—
Observations										
Fever	.0	.0	—	.1	.1	—	—	—	—	—
Persistent cough	.1	.1	.1	—	.1	.1	.1	—	.2	.3
Coordination, orthopedic problem	.1	.1	.2	.2	—	.1	.4	.1	—	.2
Impaired from alcohol	.0	.0	—	—	—	—	.1	.1	.1	—
Other observable reason for caution	.2	.2	.2	.2	.1	.1	.2	.2	.3	.3
Measurements										
Heart rate, 100	.6	.6	1.3	1.5	.3	.3	.2	.3	.2	.3
Systolic BP, 150	3.0	1.9	.4	—	1.8	.4	5.0	2.9	9.7	8.1
Diastolic BP, 100	1.7	.8	.1	.2	1.2	.3	4.5	1.5	3.4	2.6
All reasons combined										
0	88.3	86.3	95.9	94.5	91.9	89.5	80.9	80.4	70.7	71.
1	6.4	7.7	2.9	4.0	5.9	7.0	9.6	10.4	11.6	12.3
2	3.3	3.7	.8	1.0	1.8	2.6	7.0	5.7	8.7	8.5
3 or more	2.0	2.4	.3	.5	.5	.9	2.6	3.4	8.9	8.0

[1] Physical Activity Readiness Questionnaire.

smaller than the comparable differences for nonsmokers; this is especially true of the younger age group.

Discussion

Response and bias

The responses at the household level (88 percent) and for eligible individuals within households (77 percent) were both within acceptable limits and should not be a source of concern, although these numbers do raise some questions about the ability of one household member to commit others to survey participation. This would be even less of an issue if only one member per household was being sampled instead of everyone ages 7 and over.

The greater willingness of younger household members to participate presumably resulted from the clear perception of householders that this was a fitness survey. When the field teams changed their introductory remarks at the first household contact to place more emphasis on health rather than fitness, participation by older individuals improved.

Like younger householders, females were more likely to participate in the survey than were males. The relatively greater chance that females would refuse the test component, however, narrowed the difference between the sexes in the representativeness of the test sample compared with the questionnaire sample.

Notwithstanding these patterns of participation that appear to be related to respondent age and sex, the impact on the survey results is minimal because (a) age and sex were taken into account in calculating the poststratification ratio estimators, and (b) the key results are generally presented for specific age-sex groups.

Table 10. Percentile values for fitness test results and energy expenditure

Sex and percentile	7–9 years	10–12 years	13–15 years	16–19 years	20–29 years	30–39 years	40–49 years	50–59 years	60–69 years
Male				Cardiorespiratory endurance: Predicted maximal oxygen intake (ml/kg/min)					
Percentile:									
90	—	—	—	61	58	50	43	39	31
75	—	—	—	59	55	47	41	37	29
50	—	—	—	56	43	43	38	34	28
25	—	—	—	44	40	38	34	31	26
				Flexibility: Sit-and-reach (cm). Touching the toes = 25 cm					
Percentile:									
90	36	35	36	42	40	40	37	38	35
75	32	31	32	36	37	35	32	30	28
50	28	27	27	30	31	29	25	25	22
25	24	22	21	24	25	23	18	16	15
				Muscular endurance (trunk): Situps (number in 60 sec)					
Percentile:									
90	40	47	49	50	45	38	33	28	24
75	34	41	44	44	40	33	29	24	19
50	27	35	40	39	34	28	23	20	13
25	22	29	34	33	29	22	17	13	7
				Muscular endurance (arms and shoulders): Pushups (number completed)					
Percentile:									
90	21	26	32	43	41	32	25	24	24
75	15	20	25	32	32	25	20	15	13
50	9	11	18	24	24	19	13	10	9
25	4	6	10	18	17	12	7	10	5
				Muscular strength (forearm, hand): Grip strength (right plus left, kg)					
Percentile:									
90	38	56	100	119	127	127	123	114	106
75	34	50	86	108	118	117	115	105	96
50	30	43	72	97	107	107	104	97	88
25	26	37	59	84	97	97	94	87	79
				Body fatness: Sum of triceps, biceps, suprailiac, subscapular, calf skinfolds (mm)					
Percentile:									
90	52	65	66	69	82	89	85	81	81
75	39	50	46	51	62	71	72	72	68
50	31	38	36	38	46	55	58	58	57
25	26	30	30	31	34	41	46	46	48
				Energy expenditure: Leisure time (kcal/kg/day)					
Percentile:									
90	—	9.0	12.4	10.5	8.0	6.0	4.5	4.5	[1]5.5
75	—	5.2	6.5	5.2	3.9	2.8	2.2	2.2	2.8
50	—	2.8	3.2	2.1	1.5	0.9	0.8	0.6	1.0
25	—	1.2	1.7	0.8	0.5	0.3	0.2	0.1	0.1

[1] Age 60+ for this variable only.

However, there are serious implications for any survey in which similar demands are made on participants and in which a self-weighting sample is purportedly used. Such a procedure is likely to lead to biased results without adequate controls for age and sex in the analysis.

There may be some cause for concern regarding selective participation in the fitness test component of the CFS. The high rate of refusals among older age groups (table 5), coupled with evidence that refusers were most likely to be sedentary (table 6), especially in the middle age ranges, suggests that a self-selected sample may have presented itself for fitness testing. Some additional support for this conclusion is the consistent elevation of resting heart rate, blood pressures, and body mass index values of the aerobic test refusers compared with test subjects (table 7). The unanswerable question is whether there are any systematic differences between those persons who refused all testing and those who refused the aerobic and subsequent tests after the anthropometry and screening were completed.

There is also some evidence to suggest that the problem of self-selection may not have too severe an impact on the aggregate data. Table 4 shows that refusers are little different from test subjects in their self-ratings of fitness, and there is some indication that these self-ratings are reasonably accurate in this sample (table 14), even though self-ratings of fitness are sometimes biased upward (45). Moreover, the decline in test-participation rates (table 8) coinciding with the month-to-month increase in fitness activities (figure 3 and table 13) means that the data were collected disproportionately from low-activity months, a bias that would counter self-selection of active individuals for testing.

Table 10. Percentile values for fitness test results and energy expenditure—Con.

Sex and percentile	7–9 years	10–12 years	13–15 years	16–19 years	20–29 years	30–39 years	40–49 years	50–59 years	60–69 years
Female	Cardiorespiratory endurance: Predicted maximal oxygen intake (ml/kg/min)								
Percentile:									
90	—	—	—	43	41	38	35	30	26
75	—	—	—	41	39	36	33	28	24
50	—	—	—	38	35	32	28	26	22
25	—	—	—	35	32	29	24	22	20
	Flexibility: Sit-and-reach (cm). Touching the toes = 25 cm								
Percentile:									
90	39	40	43	44	43	42	40	40	37
75	36	36	39	41	39	38	36	36	33
50	32	31	35	35	34	33	31	30	28
25	29	27	28	29	28	27	25	25	23
	Muscular endurance (trunk): Situps (number in 60 sec)								
Percentile:									
90	38	44	45	43	39	31	26	22	18
75	33	38	40	39	32	26	22	16	14
50	26	33	33	33	27	21	16	7	5
25	20	27	27	27	21	15	7	—	—
	Muscular endurance (arms and shoulders): Pushups (number completed)								
Percentile:									
90	29	36	39	38	32	31	28	23	25
75	21	29	30	28	24	22	20	15	13
50	13	19	21	20	16	14	12	9	6
25	7	10	12	12	10	8	5	2	1
	Muscular strength (forearm, hand): Grip strength (right plus left, kg)								
Percentile:									
90	34	51	69	74	74	76	76	69	62
75	31	43	63	67	68	69	69	62	56
50	26	38	56	60	62	62	61	56	52
25	23	33	50	54	55	56	55	51	48
	Body fatness: Sum of triceps, biceps, suprailiac, subscapular, calf skinfolds (mm)								
Percentile:									
90	69	80	89	98	95	110	124	122	122
75	52	59	71	79	75	87	98	101	99
50	41	46	58	61	60	70	77	84	82
25	33	37	44	49	49	55	59	70	68
	Energy expenditure: Leisure time (kcal/kg/day)								
Percentile:									
90	—	10.9	11.1	9.0	5.1	3.8	3.7	4.2	[1]4.2
75	—	5.4	5.8	4.4	2.5	1.8	1.9	2.0	2.0
50	—	2.2	2.6	1.8	1.0	0.7	0.6	0.7	0.7
25	—	0.8	1.0	0.5	0.2	0.2	0.1	0.1	0.0

[1] Age 60+ for this variable only.

Screening and bias

It is axiomatic that a set of procedures to screen potential health problems, especially of a cardiovascular nature, will result in a select sample for fitness testing, one healthier and presumably more fit than the general population of the same age. That this was accomplished is demonstrated by the higher heart rates and blood pressures of screened-out persons as compared to tested subjects (table 7), despite the fact that excessive heart rate and diastolic blood pressure readings were less frequent reasons for being screened out than were the PAR–Q questions (table 9).

It is not possible from these data to give precise estimates of the magnitude of any bias in the fitness results that may have arisen from screening. However, the PAR–Q validation study (27) provides an estimated sensitivity of 37 percent for the self-administered PAR–Q compared with an exercise electrocardiogram and physician examination. Because the CFS version of the PAR–Q was somewhat modified, as described earlier, and additional procedures figured in the screening, it can be assumed that the sensitivity of the CFS screening procedures was significantly higher than 37 percent. However, even if it was double this value, there may still have been a large number of test subjects, especially in older age groups, who passed the screening without being in the best cardiovascular condition. If this was the case, the sample may have been less biased than originally suspected.

Table 11. Leisure-time physical activity in the population: Canada, 1981

Age and sex	Population in thousands	Sedentary (0–1.4 KKD)	Moderate (1.5–2.9 KKD)	Active (3+ KKD)
All ages:				
Male	10,214	53.9%	17.8%	28.3%
Female	10,514	61.6	17.1	21.4
10–19 years:				
Male	2,160	32.4	22.2	45.3
Female	2,062	40.2	20.2	39.6
20–39 years:				
Male	4,089	55.7	16.4	27.8
Female	4,074	64.7	17.0	18.4
40–54 years:				
Male	1,926	67.0	16.2	16.7
Female	1,902	68.6	16.2	15.2
55 years and over:				
Male	2,039	60.5	17.3	22.2
Female	2,476	68.8	15.4	15.8

Note: KKD = kcal/kg/day, 12-month average.

A tentative conclusion on test bias

The fact remains that there were no cardiovascular incidents associated with the fitness tests. This can probably be attributed to a combination of the successful screening of the highest risk cases and the submaximal nature of the step test. A third factor may also have played a role: Refusers may have, in effect, screened themselves out of a risky situation without waiting for the survey procedures to make this decision. Some of the data presented earlier, especially the similar levels of sedentariness among the screened and refused, support this interpretation.

All things considered, the most reasonable interpretation of possible bias in the fitness data may be to regard the refusers and persons screened out as similar in their fitness levels, and to treat these as generally inferior to those of test subjects. This approach is probably most appropriate for older age groups, in which the numbers involved were greatest, and least so for the younger groups, in which refusals were relatively rare. Some anecdotal evidence also suggests that young (ages 15–24) refusers were simply too active to devote an

Table 12. Monthly participation in 10 popular activities for persons ages 10 years and over: Canada, 1981

Month	Hours monthly per participant: Mean	Hours monthly per participant: Variation from annual average (percent)	Number of participants: Estimated (in thousands)	Number of participants: Variation from annual average (percent)
Annual average	19.9	0	11,175	0
January	16.6	−16.6	10,447	−6.5
February	16.9	−15.1	10,200	−8.7
March	17.5	−12.1	10,330	−7.6
April	19.0	−4.5	10,904	−2.4
May	21.3	+7.0	12,307	+10.1
June	24.1	+21.1	13,684	+22.5
July	24.3	+22.1	13,200	+18.1
August	24.1	+21.1	12,797	+14.5
September	21.3	+7.0	11,478	+5.1
October	19.1	−4.0	10,046	−10.1
November	17.8	−10.6	9,350	−16.3
December	17.4	−12.6	9,360	−16.2

Note: Walking, swimming (pool), calisthenics, bicycling, jogging, bowling, tennis, golf, softball, dancing—the top 10 activities in North America (11).

Table 13. Cardiorespiratory fitness, leisure-time energy expenditure, and smoking: Canada, 1981

kcal/kg/day (12-month average)	Percent of subjects achieving recommended fitness level [1]			
	Current smokers		Nonsmokers	
	20–39 years	40–69 years	20–39 years	40–69 years
	Percent			
0–0.9	33.8	32.0	33.6	36.4
1–1.9	38.7	41.8	36.8	47.8
2–2.9	37.7	44.0	49.0	45.2
3+	44.8	40.3	55.7	54.0

[1] Excludes refusals.

evening to fitness testing. The fitness results, e.g., table 10, thus may represent a "best-case" scenario for the population sampled. This qualification would be especially important to data users requiring population prevalence estimates for the fitness parameters. Other data uses would be less affected, such as the use of normative population values for interpreting the fitness results of volunteer subjects screened and tested with the same procedures (e.g., 46). The provision of such norms for the third edition of the Standardized Test of Fitness (43) was, as noted earlier, one of the principal reasons for undertaking the survey.

Questionnaire bias

Because kilocalorie expenditure values for a national population have not been published previously, it is difficult to interpret the findings of the CFS in this respect. The estimate of the active population is higher than that usually found in questionnaire surveys on physical activity (11), but this does not, by itself, invalidate the CFS results. As always, the definition of "active" is crucial. If prevalence were strictly based on activities of at least moderate intensity performed for 30 minutes at a time three or more times weekly (44), the result would obviously be a much lower estimate.

Compared with other major surveys in which similar populations were sampled, the CFS had a longer reporting period and a longer list of activities. A 12-month recall period would clearly produce higher estimates of activity than a 2-week recall in a survey conducted during the off-peak activity months. It is also apparent that prompting a respondent with an activity list produces more reports of activity than an open-ended approach (11). The CFS list comprised 120 activities (appendix D), but other national surveys have been limited to as few as seven activities (47); therefore, comparability is low. Analysis published elsewhere (11) shows that there are sufficient participants in gardening and dancing to place these activities in the top 10 for popularity. These activities were covered in the CFS and were included in the energy expenditure calculation; they were omitted in most other surveys.

One might question the accuracy of reporting for a 12-month period any of the key dimensions of duration, frequency, and intensity. Despite the guidance and probing of the testers, a few individuals reported activities of exceedingly long durations. For the purpose of calculating energy expenditure, no duration exceeding 241 minutes was included, nor was any frequency exceeding 366 times per year or 32 times in the previous month. Unknown durations were set to a minimal value of 13 minutes. Each of these adjustments, although rare, would tend to reduce the estimated prevalence of activity and counter any tendency of individuals to exaggerate their exercise levels. Another conservative influence was the setting of METS values that tended to be on the low side compared with the figures obtained from the literature (appendix E).

Some error may be introduced into kilocalorie calculations such as those reported here by using a single METS value for all individuals performing a given activity at a stated intensity. Thus a midrange bicycling speed of 20 km/h (7.1 METS (34)) may represent a light workload for a fit individual (3 METS, appendix E) and a heavy workload for an unfit person (10 METS, appendix E). This error should be minimized in the CFS because behavioral descriptors (breathing, perspiration) were an integral part of the intensity ratings. These descriptors are particularly important in light of the tendency of fit and unfit individuals to give similar perceived-exertion ratings to the same absolute workload when no behavioral anchor points are provided (48, 49).

The procedures and sample of the Minnesota Heart Health Project (50) probably come closer to the CFS than any other major project. The average daily kilocalorie

Table 14. Self-rated versus measured physical fitness, in persons 10–69 years of age

Self-rated fitness [1]	Total	Aerobic test results [2]				
		Recommended	Minimum	Undesirable	Screened	Refused
Total	100.0	32.2	22.4	2.2	13.3	29.9
More fit	100.0	42.4	15.0	—	10.3	31.2
As fit	100.0	34.3	22.3	1.6	12.1	29.7
Less fit	100.0	18.1	29.0	4.6	18.1	30.2

[1] In response to the question, Comparing yourself to others of your own age and sex, would you say you are. . .

[2] Recommended = 65–70% of average maximum aerobic power for same sex 10 years younger.

Minimum = 65–70% of average maximum for same sex, same age.

Undesirable = 65–70% of average maximum for same sex, 10 years older.

expenditure for that sample was 193 for men and 111 for women (ages 25–74). The values for the age group 25–69 in the CFS are 177 and 104 for men and women, respectively.

The positive relationship between the energy expenditure values estimated from CFS questionnaire responses and cardiorespiratory fitness as estimated from the step test provides indirect validation of the questionnaire. Construct validity may also be inferred from the age-sex distribution of the active population based on energy expenditure values: This is consistent with that found in seven other nationwide surveys of the U.S. and Canadian populations using quite varied questionnaire approaches (11).

Observations on procedures

Like the Canada Health Survey (10), the CFS experience demonstrated the feasibility of in-home fitness testing on a free-living sample, with acceptable levels of data quality being obtained. The assigned workload of 28 households per team per month was a good average, but there was wide variation in the ability of field teams to maintain this schedule. This variation was attributable primarily to travel distance and occasional high nonresponse within one assignment area.

Only a few problems with the protocol are serious enough to warrant discussion here. One, which became apparent from the first assignments, was the difficulty that respondents had in completing the first three questionnaire pages (appendix C). The format turned out to be too complicated, especially for the youngest and oldest participants. As a result, these questions were completed with considerable guidance and probing by the testers.

Although the step test is clearly safe, inexpensive, and operationally feasible (and apparently valid and reliable), it has its limitations, and these must be appreciated by designers of new surveys. As a submaximal test, it is based on three imperfect assumptions: A linear heart rate/oxygen uptake relationship, a constant mechanical efficiency, and a known maximum heart rate. Because the assumptions are more reasonable for groups than individuals, systematic errors of up to 10 percent (especially at the tails of the VO_2max distribution) and random errors of another 10 percent can be expected (51). Because the postexercise heart rate measurement is so crucial and is subject to considerable error without appropriate methods (20, 52), great care must be exercised in this procedure. Even with this precaution, the test is best used for classifying individuals into fitness categories (e.g., table 13) and estimating group means for maximum oxygen consumption rather than calculating individual values of ml O_2/kg/min.

These problems and others apply to children. As noted earlier, the final heart rates are not predictive of oxygen uptake by subjects ages 7–14 years. Number of stages completed must therefore be used as an indication of aerobic endurance (42). The CFS experience provided evidence of the special nature of the child's anatomy and physiology described by Bar-Or (chapter 10): Small children had difficulty with the 20-cm step and had to jump up and down the stairs, and the youngest ones had difficulty maintaining the proper cadence for lack of adequate eye-foot coordination. This latter problem also affected a small number of adults and plays havoc with the assumption that every subject at a given cadence performs the same workload.

There also exists a problem in predicting body composition from the anthropometric measures, a difficulty that is not restricted to the CFS. Bouchard (53) has proposed an approach for the next edition of the Standardized Test of Fitness (43) that will be a definite improvement over existing methods. The proposal is based on a three-level triage: Body mass index values are used to select "overweight" individuals; skinfold values for these individuals separate the obese from the very muscular; and waist/hip ratios and trunk skinfolds of the obese determine the location of fat. Those filtered out by such a process are considered at high risk for health problems, especially cardiovascular disease, because of their body composition and fat structure.

Recommendations for the National Health and Nutrition Examination Survey

On the basis of the Canada Fitness Survey experience, several operational recommendations are offered to the designers of the National Health and Nutrition Examination Survey (NHANES). However, these are preceded by substantive recommendations based on 3-years experience with data analysis and contact with users.

Subject matter suggestions

1. Through NHANES prevalence estimates should be provided of the seriously overweight, and especially those at elevated risk of cardiovascular disease because of excessive abdominal fat.
2. Cardiorespiratory fitness levels of the adult population should be estimated.
3. Trunk flexibility should be measured for adults if it is agreed that this is a reasonable indicator of potential and actual lower back problems.
4. Those unable to undergo exercise testing should be assessed for their capacity to perform the normal activities of everyday life (e.g., grasping, lifting, reaching, walking distances).
5. Total caloric expenditure should be assessed along with caloric intake. This would permit calculation of energy balance values, a unique undertaking for a general population sample. Such values could be usefully related to both cardiorespiratory fitness and body fatness.
6. Activity patterns (type, frequency, intensity, duration) should be identified for leisure time and work time. The relationship of this "exercise dose" to physical health outcomes is insufficiently understood, and NHANES III provides an unprecedented

opportunity to examine this association. This would be especially important and useful if a longitudinal followup were later conducted on a NHANES cohort.

Operational considerations

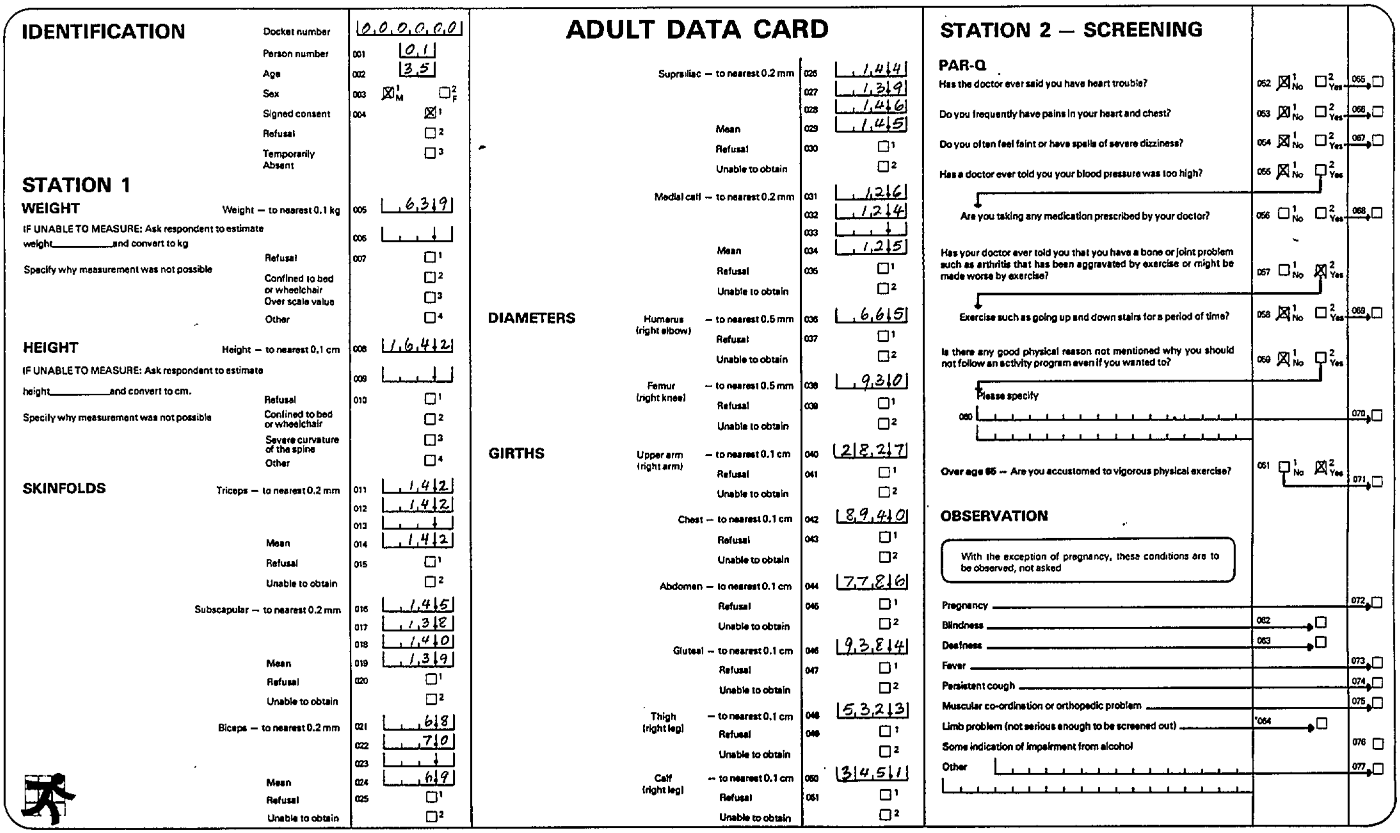
7. Measure waist and hip girths in addition to the usual anthropometric parameters (weight; standing height; biceps, triceps, subscapular, suprailiac, and calf skinfolds). Include abdominal and front-thigh skinfolds for their relevance to the distribution of body fat.
8. Employ a submaximal test of aerobic fitness. This will provide not only an acceptable balance between safety and accuracy but also, if it is equivalent to the Canadian Home Fitness Test, international comparative data.
9. Establish the reliability, validity, and general limitations of such a test for both adults and children. If necessary and desired, develop an appropriate test of aerobic fitness for children.
10. Develop a simple performance test to assess capability to perform the activities of daily life; as a second choice, use a standardized questionnaire to collect this information.
11. Use a multimethod approach to screening (questions, physiologic measures, observations) and determine the sensitivity and specificity of these procedures. As a contribution to the science of fitness testing, develop a screening protocol that can be employed in field settings without elaborate equipment by technicians without medical training.
12. Use the 1985 National Health Interview Survey (NHIS) supplement questionnaire to assess leisure-time activity patterns (appendix F), the recall period and other features being suitably modified to deal with the seasonality problem discussed in recommendation 13. Use an extensive reference card of activities to prompt the respondent; qualify "walking" as "brisk" and avoid ambiguous terms such as "regularly" or "often." Whatever questionnaire is used must be validated and its reliability assessed. If it is different from the NHIS instrument, their comparability must be established so that comparisons can be made between them. (These technical points are discussed further in the concluding section of Chapter 7.)
13. The design of the data collection must take account of the sharp seasonal fluctuations in physical activity. The past practice of "following the good weather" (54) is liable to bias upward any prevalence estimates of activity and fitness and confound regional comparisons. This will be particularly acute if a short (e.g., 2-week) recall period is used. Several possibilities should be considered: (a) A followup at 3-, 6-, 9-month intervals to produce an average annual value for energy expenditure; (b) collection of data in all regions in all seasons; (c) use of a 12-month reporting period. The problem of seasonality is discussed further in Chapter 7 of this report.
14. Initial refusals subsequently persuaded to participate should be identified on the data file for analysis of possible response bias. Similarly, a systematic effort to record observations of age, sex, height, and weight of all refusers would be invaluable.

References

1. Statistics Canada: Culture statistics/recreational activities 1976. Catalog No. 87–501, occasional. Minister of Supply and Services, Ottawa, 1980.
2. Stephens T: Benefits of a fitness survey. *Can Med Assoc J* 126:587–8, 1982.
3. White FMM: The Canada Fitness Survey: implications for health research and public health practice. *Can J Public Health* 74:91–95, 1983.
4. Health and Welfare Canada: *Proceedings of the National Conference on Fitness and Health, 1972.* Information Canada, Ottawa, 1975.
5. President's Council on Physical Fitness and Sports: National Adult Physical Fitness Survey. *Phys Fitness Res Digest* 4:1–27, 1974.
6. Nutrition Canada: *National Report.* Health and Welfare Canada, Ottawa, 1973 (p.121).
7. Canada Fitness Survey: Canada's Fitness: preliminary findings of the 1981 Survey. Canada Fitness Survey, Ottawa, June 1982 (p.3).
8. Jette M: The Standardized Test of Fitness in Occupational Health: a pilot project. *Can J Public Health* 69:431–438, 1978.
9. Jette M, Campbell J, Mongeon J, and Routhier R: The Canadian Home Fitness Test as a predictor of aerobic capacity. *Can Med Assoc J* 114:680–682, 1976.
10. Health and Welfare Canada and Statistics Canada: The health of Canadians: report of the Canada Health Survey. Catalog No. 82–538. Minister of Supply and Services, Ottawa, June 1981 (Chapter 3).
11. Stephens T, Jacobs DR, and White CC: A descriptive epidemiology of leisure-time physical activity. *Public Health Rep* 100:147–158, 1985.
12. Caspersen CJ, Powell KE, and Christenson GM: Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Rep* 100:126–131, 1985.
13. Durnin JVGA and Womersley J: Body fat assessed from total body density and its estimation from skinfold thickness: Measurements on 481 men and women aged from 16 to 72 years. *Brit J Nutr* 32:77–97, 1974.
14. Carter, JEL: The Heath-Carter somatotype method. San Diego State University, October 1975 (mimeo).
15. Fitness and Amateur Sport Canada: *Standardized Test of Fitness, Operations Manual* (2nd edition). Ottawa, 1981.
16. Bailey DA, Shephard RJ, and Mirwald RL: Validation of a self-administered home test of cardiorespiratory fitness. *Can J Appl Sports Sci* 1:67–78, 1976.
17. Jette M: A comparison between predicted VO_2max from the Astrand Procedure and the Canadian Home Fitness Test. *Can J Appl Sports Sci* 4:214–218, 1979.
18. Fitzgerald PI, Knowlton RG, Sedlock DA, Tahamount MV, et al.: A comparison of maximal aerobic power predicted from the Canadian Home Fitness Test and a direct treadmill test. *Med Sci Spts Exerc* 12:88, 1980.
19. Quinney HA and Cottle W: The influence of ambient temperature on performance on the Canadian Home Fitness Test. Report to Fitness Canada, Ottawa, 1982.
20. Shephard RJ, Cox M, Corey P, and Smyth R: Some factors affecting accuracy of Canadian Home Fitness Test scores. *Can J Appl Sports Sci* 4:205–209, 1979.

21. Jette M: The energy requirements of the Canadian Home Fitness Test and their application to the evaluation of work performance. *Can J Public Health* 74:401–403, 1983.
22. Jette M, Ashton NJ, and Sharratt MT: Development of a cardiorespiratory step-test of fitness for children 7–14 years of age. *Can J Public Health* 75:212–217, 1984.
23. Clarke HH: Toward a better understanding of muscular strength. *Phys Fit Res Digest* 3:1–20, 1973.
24. Jackson A, Watkins M, and Patton RW: A factor analysis of twelve selected maximal isotomic strength performances on the Universal Gym. *Med Sci Spts Exerc* 12:274–277, 1980.
25. Kraus H and Raab W: *Hypokinetic disease.* Charles C. Thomas, Springfield, IL, 1961.
26. Shephard RJ: The current status of the Canadian Home Fitness Test. *Brit J Sports Med* 14:114–125, 1980.
27. Chisholm DM, Collis MC, Kulak LL, Davenport W, et al.: *PAR–Q Validation Report.* Health and Welfare Canada, Ottawa, May 1978.
28. Jette M: An exercise prescription programme for use in conjunction with the Canadian Home Fitness Test. *Can J Public Health* 66:461, 1975.
29. Stephens T: Health practices and health status: Evidence from the Canada Health Survey. *Amer J Prev Med* 2(3), 1986, 24–30.
30. Taylor HA, Jacobs DR, Schucker B, Knudsen J, et al.: A questionnaire for the assessment of leisure-time physical activities. *J Chron Dis* 31:741–755, 1978.
31. Stephens T: Fitness and lifestyle in Canada. Canada Fitness Survey, Ottawa, May 1983.
32. Stephens T, Craig CL, and Ferris BF: Adult physical activity in Canada: Findings from the Canada Fitness Survey I. *Can J Public Health* 77, 1986, 285–290.
33. Stephens T, Craig CL, and Ferris BF: Adult physical fitness and hypertension in Canada: Findings from the Canada Fitness Survey II. *Can J Public Health* 77, 1986, 291–295.
34. Groupe d'etude de Kino-Quebec sur le systeme de quantification de la depense energetique (GSQ): Rapport final et annexe au rapport. Editeur officiel du Quebec, Service des impressions en regie, Quebec, fevrier 1980.
35. Berkman LF and Breslow L: *Health and ways of living.* Oxford University Press, New York, 1983.
36. Bradburn, NM: *The structure of psychological well-being.* Aldine Publishing Co., Chicago, 1969.
37. Canada Fitness Survey: Household survey Micro-data tape documentation. Canada Fitness Survey, Ottawa, January 1985.
38. Canada Fitness Survey: A users guide to CFS findings. Canada Fitness Survey, Ottawa, January 1983.
39. Canada Fitness Survey: Fitness and aging. Canada Fitness Survey, Ottawa, November 1982.
40. Canada Fitness Survey: Canadian youth and physical activity. Canada Fitness Survey, Ottawa, October 1983.
41. Canada Fitness Survey: Changing times: Women and physical activity. Canada Fitness Survey, Ottawa, October 1984.
42. Canada Fitness Survey: Physical fitness of Canadian youth. Canada Fitness Survey, Ottawa, March 1985.
43. Fitness Canada: *Canadian Standardized Test A Fitness Operations Manual* (Third Edition). Fitness and Amateur Sport Canada, Ottawa, 1986.
44. Haskell WL, Montoye HJ, and Orenstein D: Physical activity and exercise to achieve health-related physical fitness components. *Public Health Rep* 100:202–212, 1985.
45. Optenberg SA, Lairson DR, Slater DH, and Russell ML: Agreement of self-reported and physiologically estimated fitness status in a symptom-free population. *Prev Med* 13:349–354, 1984.
46. Pfeiffer S, Graham TE, Webb RDG, Wilson BA, et al.: Aspects of physical fitness and health in Ontario dairy farmers. *Can J Public Health* 75:204–211, 1984.
47. Highlights from Wave I of the National Survey of Personal Health Practices and Consequences: United States, 1979. *Vital and Health Statistics.* Series 15, No. 1. DHHS Publication No. (PHS) 81–1162. National Center for Health Statistics, Hyattsville, MD, June 1981.
48. Mihevic PM: Cardiovascular fitness and the psychophysics of perceived exertion. *Res Q Ex Spt* 54(3):239–246, 1983.
49. Morgan WP: Psychophysiology of self-awareness during vigorous physical activity. *Res Q Exerc Spt* 52(3):385–427, 1981.
50. Folsum AR, Caspersen CJ, Taylor HL, Jacobs DR, et al.: Leisure-time physical activity and its relationship to coronary risk factors in a population-based sample: The Minnesota Heart Survey. *Amer J Epidemiol* 121:570–574, 1985.
51. Shephard RJ: Tests of maximum oxygen uptake: a critical review. *Sports Med* 1:99–124, 1984.
52. Morgan K, Hughes AO, and Phillip R: Reliability of a test of cardiovascular fitness. *Int J Epidemiol* 13:32–37, 1984.
53. Bouchard C: Interpretation of body weight and related measurements. Fitness Canada and Canada Fitness Survey, Ottawa (working notes), May 1985.
54. Plan and operation of the Second National Health and Nutrition Examination Survey, 1976–1980. *Vital and Health Statistics.* Series 1, No. 15. DHHS Publication No. (PHS) 81–1317. National Center for Health Statistics, Hyattsville, MD, July 1981.
55. Dunne MF, Selner AJ, and Gregor RJ: Physiological profile of a rollerskating training program (Abstract). *Med Sc Spts Ex* 13(2):103, 1981.
56. Government of Ontario: *FitFive: Awards Scoring Booklet.* Ministry of Culture and Recreation, Toronto, n.d.
57. Amundsen LR, Takahashi M, Carter CA, and Nielson DH: Energy cost of rehabilitation calisthenics. *Physical Therapy* 59(7):855–858, 1979.
58. Durnin JVGA and Passmore R: *Energy, work and leisure.* Heinemann Educational Books Ltd., London, 1967.
59. Fox SM, Naughton JP, and Gorman PA: Physical activity and cardiovascular health. III. The exercise prescription: frequency and type of activity. *Mod Concepts Cardiovasc Dis* 41, June 1972.
60. Montgomery DL: Heart rate response to racketball. *Physician and Spts. Med* 9:59–62, 1981.
61. Montgomery DL, Malcolm V, and McDonnell E: A comparison of the intensity of play in squash and running. *Physician and Spts Med* 9:116–119, 1981.
62. American College of Sports Medicine. Supplement to guidelines for graded exercise testing and exercise prescription. Calculation of the relative energy expenditure for various activities in METS. Lea and Febiger, Philadelphia, 1975.
63. Karen G, Epstein T, Magazanik A, and Sohar E: The energy cost of walking and running with and without a backpack load. *Eur J Appl Physiol* 46:317–324, 1981.
64. Cohen JL, Witriol I, Segal K, and McArdle WD: Cardiorespiratory responses to ballet exercise and physical capacities of elite ballet dancers (Abstract). *Med Sc Spts Ex* 13(2):103, 1981.
65. Weber H: The energy cost of aerobic dancing. *Med Sc Spts. Ex* (Abstract), 5:65–66, 1973.
66. Shaw DK and Deutch DT: Metabolic cost of performing karate kata (Abstract). *Med Sc Spts Ex* 13(2):125, 1981.
67. Pandolf KB, Givoni B, and Goldman RF: Predicting energy expenditure with loads while standing or walking very slowly. *J Appl Physiol: Respirat Environ Exercise Physiol* 43(4):577–581, 1977.

Appendix A. Adult data card



IDENTIFICATION

Docket number		000000
Person number	001	01
Age	002	35
Sex	003	☒1 M ☐2 F
Signed consent	004	☒1
Refusal		☐2
Temporarily Absent		☐3

STATION 1

WEIGHT

Weight — to nearest 0.1 kg	005	63.9
IF UNABLE TO MEASURE: Ask respondent to estimate weight________and convert to kg	006	
Specify why measurement was not possible — Refusal	007	☐1
Confined to bed or wheelchair		☐2
Over scale value		☐3
Other		☐4

HEIGHT

Height — to nearest 0.1 cm	008	164.2
IF UNABLE TO MEASURE: Ask respondent to estimate height________and convert to cm.	009	
Specify why measurement was not possible — Refusal	010	☐1
Confined to bed or wheelchair		☐2
Severe curvature of the spine		☐3
Other		☐4

SKINFOLDS

Triceps — to nearest 0.2 mm	011	14.2
	012	14.2
	013	
Mean	014	14.2
Refusal	015	☐1
Unable to obtain		☐2
Subscapular — to nearest 0.2 mm	016	14.5
	017	13.8
	018	14.0
Mean	019	13.9
Refusal	020	☐1
Unable to obtain		☐2
Biceps — to nearest 0.2 mm	021	6.8
	022	7.0
	023	
Mean	024	6.9
Refusal	025	☐1
Unable to obtain		☐2

Canada Fitness Survey

ADULT DATA CARD

Suprailiac — to nearest 0.2 mm	026	14.4
	027	13.9
	028	14.6
Mean	029	14.5
Refusal	030	☐1
Unable to obtain		☐2
Medial calf — to nearest 0.2 mm	031	12.6
	032	12.4
	033	
Mean	034	12.5
Refusal	035	☐1
Unable to obtain		☐2

DIAMETERS

Humerus (right elbow) — to nearest 0.5 mm	036	66.5
Refusal	037	☐1
Unable to obtain		☐2
Femur (right knee) — to nearest 0.5 mm	038	93.0
Refusal	039	☐1
Unable to obtain		☐2

GIRTHS

Upper arm (right arm) — to nearest 0.1 cm	040	282.7
Refusal	041	☐1
Unable to obtain		☐2
Chest — to nearest 0.1 cm	042	894.0
Refusal	043	☐1
Unable to obtain		☐2
Abdomen — to nearest 0.1 cm	044	778.6
Refusal	045	☐1
Unable to obtain		☐2
Gluteal — to nearest 0.1 cm	046	938.4
Refusal	047	☐1
Unable to obtain		☐2
Thigh (right leg) — to nearest 0.1 cm	048	532.3
Refusal	049	☐1
Unable to obtain		☐2
Calf (right leg) — to nearest 0.1 cm	050	345.1
Refusal	051	☐1
Unable to obtain		☐2

STATION 2 — SCREENING

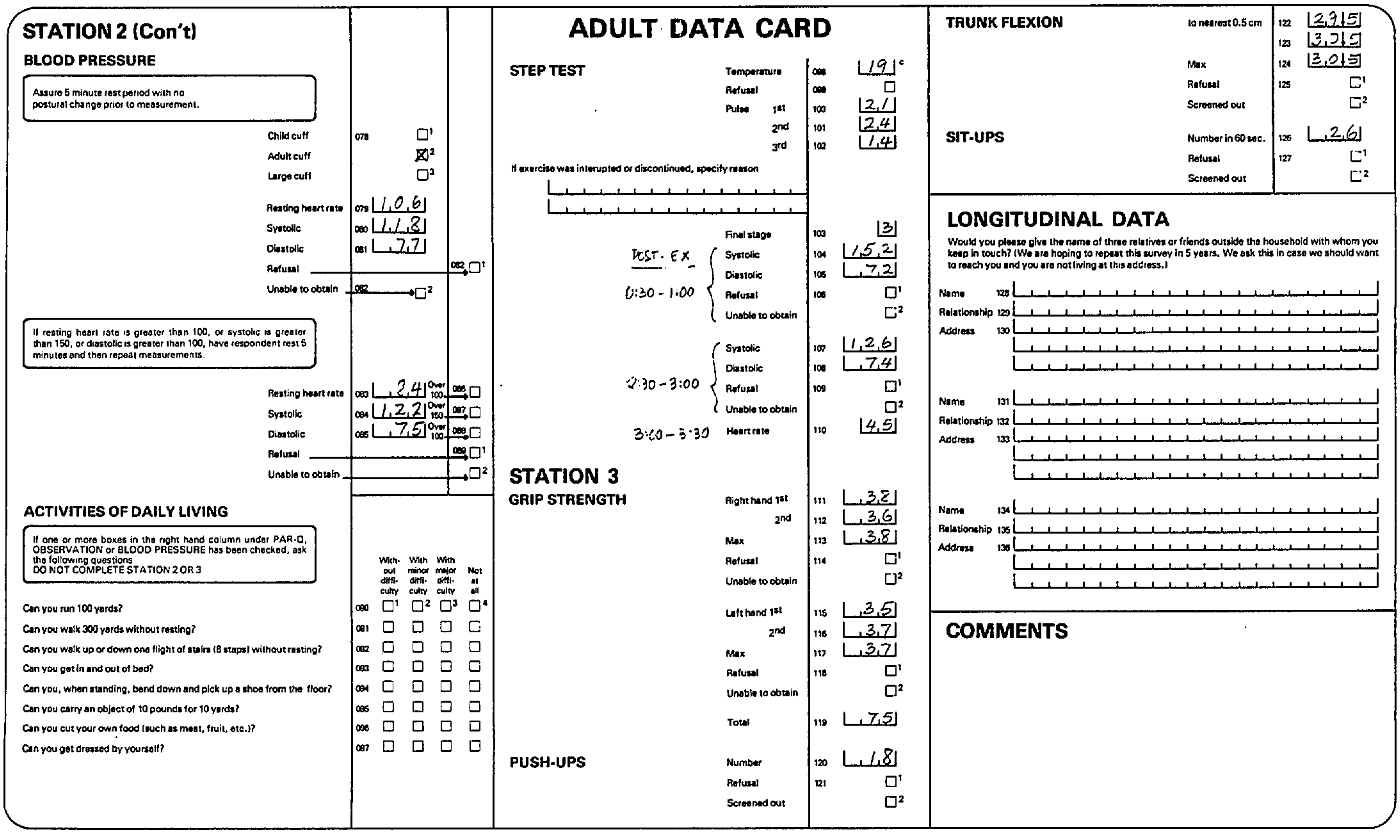
PAR-Q

Has the doctor ever said you have heart trouble?	052	☒1 No	☐2 Yes	065 → ☐
Do you frequently have pains in your heart and chest?	053	☒1 No	☐2 Yes	066 → ☐
Do you often feel faint or have spells of severe dizziness?	054	☒1 No	☐2 Yes	067 → ☐
Has a doctor ever told you your blood pressure was too high?	055	☒1 No	☐2 Yes	
Are you taking any medication prescribed by your doctor?	056	☐1 No	☐2 Yes	068 → ☐
Has your doctor ever told you that you have a bone or joint problem such as arthritis that has been aggravated by exercise or might be made worse by exercise?	057	☐1 No	☒2 Yes	
Exercise such as going up and down stairs for a period of time?	058	☒1 No	☐2 Yes	069 → ☐
Is there any good physical reason not mentioned why you should not follow an activity program even if you wanted to?	059	☒1 No	☐2 Yes	
Please specify	060			070 → ☐
Over age 65 — Are you accustomed to vigorous physical exercise?	061	☐1 No	☒2 Yes	071 → ☐

OBSERVATION

With the exception of pregnancy, these conditions are to be observed, not asked

Pregnancy		072 → ☐
Blindness	062 → ☐	
Deafness	063 → ☐	
Fever		073 → ☐
Persistent cough		074 → ☐
Muscular co-ordination or orthopedic problem		075 → ☐
Limb problem (not serious enough to be screened out)	064 → ☐	
Some indication of impairment from alcohol		076 ☐
Other		077 → ☐



STATION 2 (Con't)

BLOOD PRESSURE

Assure 5 minute rest period with no postural change prior to measurement.

Child cuff	078	☐1	
Adult cuff		☒2	
Large cuff		☐3	
Resting heart rate	079	106	
Systolic	080	118	
Diastolic	081	77	
Refusal			082 → ☐1
Unable to obtain	082 → ☐2		

If resting heart rate is greater than 100, or systolic is greater than 150, or diastolic is greater than 100, have respondent rest 5 minutes and then repeat measurements.

Resting heart rate	083	24 Over 100	086 → ☐
Systolic	084	122 Over 150	087 → ☐
Diastolic	085	75 Over 100	088 → ☐
Refusal			089 → ☐1
Unable to obtain			→ ☐2

ACTIVITIES OF DAILY LIVING

If one or more boxes in the right hand column under PAR-Q, OBSERVATION or BLOOD PRESSURE has been checked, ask the following questions
DO NOT COMPLETE STATION 2 OR 3

		Without difficulty	With minor difficulty	With major difficulty	Not at all
Can you run 100 yards?	090	☐1	☐2	☐3	☐4
Can you walk 300 yards without resting?	091	☐	☐	☐	☐
Can you walk up or down one flight of stairs (8 steps) without resting?	092	☐	☐	☐	☐
Can you get in and out of bed?	093	☐	☐	☐	☐
Can you, when standing, bend down and pick up a shoe from the floor?	094	☐	☐	☐	☐
Can you carry an object of 10 pounds for 10 yards?	095	☐	☐	☐	☐
Can you cut your own food (such as meat, fruit, etc.)?	096	☐	☐	☐	☐
Can you get dressed by yourself?	097	☐	☐	☐	☐

ADULT DATA CARD

STEP TEST

Temperature	098	19 °C
Refusal	099	☐
Pulse 1st	100	21
2nd	101	24
3rd	102	14

If exercise was interupted or discontinued, specify reason

	Final stage	103	3
POST. EX 0:30 - 1:00	Systolic	104	152
	Diastolic	105	72
	Refusal	106	☐1
	Unable to obtain		☐2
2:30 - 3:00	Systolic	107	126
	Diastolic	108	74
	Refusal	109	☐1
	Unable to obtain		☐2
3:00 - 3:30	Heart rate	110	45

STATION 3

GRIP STRENGTH

Right hand 1st	111	38
2nd	112	36
Max	113	38
Refusal	114	☐1
Unable to obtain		☐2
Left hand 1st	115	35
2nd	116	37
Max	117	37
Refusal	118	☐1
Unable to obtain		☐2
Total	119	75

PUSH-UPS

Number	120	18
Refusal	121	☐1
Screened out		☐2

TRUNK FLEXION

to nearest 0.5 cm	122	29.5
	123	30.5
Max	124	30.5
Refusal	125	☐1
Screened out		☐2

SIT-UPS

Number in 60 sec.	126	26
Refusal	127	☐1
Screened out		☐2

LONGITUDINAL DATA

Would you please give the name of three relatives or friends outside the household with whom you keep in touch? (We are hoping to repeat this survey in 5 years. We ask this in case we should want to reach you and you are not living at this address.)

Name 128
Relationship 129
Address 130

Name 131
Relationship 132
Address 133

Name 134
Relationship 135
Address 136

COMMENTS

Appendix B. Interpretation of blood pressure readings and associated messages for subjects

REFERENCE CARD - STATION 2

INTERPRETATION OF BLOOD PRESSURE

BLOOD PRESSURE READING		Refer to Message No.	
		Pregnant Females	All Others
ALL RESPONDENTS 14 YEARS OR YOUNGER			
Systolic	Diastolic		
79 or Less	49 or Less	6	1
80 -- 139	50 -- 89	7	2
140 -- 149	90 -- 94	8	3
150 -- 179	95 -- 119	9	4
180 or Greater	120 or Greater	9	5
MALES 15-49 YEARS -- FEMALES 15 YEARS OR OLDER			
Systolic	Diastolic		
84 or Less	54 or Less	6	1
85 -- 149	55 -- 94	7	2
150 -- 159	95 -- 99	8	3
160 -- 179	100 -- 119	9	4
180 or Greater	120 or Greater	9	5
MALES 50 YEARS OR OLDER			
Systolic	Diastolic		
89 or Less	59 or Less		1
90 -- 159	60 -- 99		2
160 -- 169	100 -- 119		3
170 -- 179	120 or Greater		4
180 or Greater			5

BLOOD PRESSURE MESSAGE

1. Your blood pressure is __ over ___. Although this is below the usual range, it may be entirely normal for you. If so it is not a health problem. Mention this result to your doctor.
2. Your blood pressure is __ over __. This is within the normal range. Have it checked again in one year.
3. Your blood pressure is __ over __. This is slightly above the normal range. It is a minor elevation and there is no cause for alarm, as it may be due to the circumstances of the test. See your doctor within the next few weeks for a recheck unless you are already being treated for high blood presssure.
4. Your blood pressure is __ over __. This is higher than it should be. Arrange to see your doctor for a recheck within the next two weeks.
5. Your blood pressure is __ over __. This is much higher than it should be. See your doctor as soon as possible.
6. Your blood pressure is __ over __. Although this is below the usual range it may be entirely normal for you. Because you are pregnant your doctor should know the reading. Mention it at your next visit.
7. Your blood pressure is __ over __. This is within the normal range for a women of your age who is not pregnant. Because your are pregnant, I cannot be positive that it is normal for you without knowing what your blood pressure was before you became pregnant. Mention the reading to your doctor at your next visit.
8. Your blood pressure is __ over __. This is slightly higher than normal, and because you are pregnant it may be important. Contact your doctor within the next two or three days.
9. Your blood pressure is __ over __. This is much higher than it should be. It is very important that your doctor know about it, particularly because you are pregnant. Contact your doctor as soon as possible.

Appendix C. Questionnaire (pp.1–3)

1

PHYSICAL ACTIVITIES

WHAT YOU DO AT WORK OR AT SCHOOL OR IN THE HOME, PLUS YOUR ACTIVITY IN YOUR LEISURE TIME ALL CONTRIBUTE TO YOUR CURRENT LEVEL OF FITNESS. THE FOLLOWING QUESTIONS WILL PROVIDE A COMPLETE PICTURE OF ALL YOUR ACTIVITIES.

TO HELP YOU DESCRIBE YOUR ACTIVITIES, WE HAVE DESIGNED FOUR QUESTIONS – ONE FOR THOSE YOU DO DAILY, ONE FOR THOSE YOU DO EACH WEEK, ONE FOR THOSE YOU HAVE DONE IN THE LAST MONTH, AND THE FOURTH FOR THOSE ACTIVITIES YOU HAVE DONE IN THE LAST YEAR.

1. DAILY ACTIVITIES

For those activities which you do most days of the week (such as work, school and housework), how much time do you spend. . .

		Almost all of the time	About 3/4 of the time	About 1/2 of the time	About 1/4 of the time	Almost none of the time
Sitting	01	☐1	☐2	☐3	☐4	☐5
Standing	02	☐1	☐2	☐3	☐4	☐5
Walking	03	☐1	☐2	☐3	☐4	☐5
Walking up stairs	04	☐1	☐2	☐3	☐4	☐5
Lifting or carrying heavy objects	05	☐1	☐2	☐3	☐4	☐5

2. WEEKLY ACTIVITIES

Please refer to the reference card for a list of activities. Answer the following for the physical activities you do each week.

Light housework and handywork: washing dishes, ironing, making beds, mowing lawn, etc

Number of occasions each month	Average time actually spent on each occasion	Intensity: Light Slight change from normal	Intensity: Medium Some perspiration Above normal breathing	Intensity: Heavy Heavy perspiration Heavy breathing
J F M (06) A M J (07) J A S (08) O N D (09)	Hrs (10) Mins (11)	☐1 (12)	☐2	☐3

Heavy housework and handywork: washing and waxing floors, painting, etc

Number of occasions each month	Average time actually spent on each occasion	Intensity: Light Slight change from normal	Intensity: Medium Some perspiration Above normal breathing	Intensity: Heavy Heavy perspiration Heavy breathing
J F M (13) A M J (14) J A S (15) O N D (16)	Hrs (17) Mins (18)	☐1 (19)	☐2	☐3

Name of activity ________ 20 ____ 21

Number of occasions each month	Average time	Intensity: Light	Intensity: Medium	Intensity: Heavy	Organized in levels or in a league: Yes	No	Competitive: Yes	No
J F M (22) A M J (23) J A S (24) O N D (25)	Hrs (26) Mins (27)	☐1 (28)	☐2	☐3	☐1 (29)	☐2	☐1 (30)	☐2

Name of activity ________ 31 ____ 32

Number of occasions each month	Average time	Intensity: Light	Intensity: Medium	Intensity: Heavy	Organized: Yes	No	Competitive: Yes	No
J F M (33) A M J (34) J A S (35) O N D (36)	Hrs (37) Mins (38)	☐1 (39)	☐2	☐3	☐1 (40)	☐2	☐1 (41)	☐2

Name of activity ________ 42 ____ 43

Number of occasions each month	Average time	Intensity: Light	Intensity: Medium	Intensity: Heavy	Organized: Yes	No	Competitive: Yes	No
J F M (44) A M J (45) J A S (46) O N D (47)	Hrs (48) Mins (49)	☐1 (50)	☐2	☐3	☐1 (51)	☐2	☐1 (52)	☐2

Name of activity ________ 53 ____ 54

Number of occasions each month	Average time	Intensity: Light	Intensity: Medium	Intensity: Heavy	Organized: Yes	No	Competitive: Yes	No
J F M (55) A M J (56) J A S (57) O N D (58)	Hrs (59) Mins (60)	☐1 (61)	☐2	☐3	☐1 (62)	☐2	☐1 (63)	☐2

2

3. ACTIVITIES IN THE LAST MONTH

Please refer to the reference card for a list of activities. Answer the following for the physical activities you have done at least once in the last month. (**Do not include activities already listed in Weekly Activities.**)

Gardening and cultivating such as spading, digging, weeding

Occasions in the last month	Average time actually spent on each occasion Hrs	Mins	Intensity Light Slight Change from normal state	Intensity Medium Some perspiration Above normal breathing	Intensity Heavy Heavy perspiration Heavy breathing
☐☐ 01	☐ 02	☐☐ 03	☐ 1 04	☐ 2	☐ 3

Shovelling snow

Occasions in the last month	Average time actually spent on each occasion Hrs	Mins	Intensity Light Slight Change from normal state	Intensity Medium Some Perspiration Above normal breathing	Intensity Heavy Heavy perspiration Heavy breathing
☐☐ 05	☐ 06	☐☐ 07	☐ 1 08	☐ 2	☐ 3

Mowing the lawn (pushing a power mower)

Occasions in the last month	Average time actually spent on each occasion Hrs	Mins	Intensity Light Slight Change from normal state	Intensity Medium Some perspiration Above normal breathing	Intensity Heavy Heavy perspiration Heavy breathing
☐☐ 09	☐ 10	☐☐ 11	☐ 1 12	☐ 2	☐ 3

Name of activity ____________________ 13 ☐☐☐☐ 14

Occasions in the last month	Average time Hrs	Mins	Intensity Light	Intensity Medium	Intensity Heavy	Organized in levels or in a league Yes	No	Competitive Yes	No
☐☐ 15	☐ 16	☐☐ 17	☐ 1 18	☐ 2	☐ 3	☐ 1 19	☐ 2	☐ 1 20	☐ 2

Name of activity ____________________ 21 ☐☐☐☐ 22

Occasions in the last month	Average time Hrs	Mins	Intensity Light	Intensity Medium	Intensity Heavy	Organized Yes	No	Competitive Yes	No
☐☐ 23	☐ 24	☐☐ 25	☐ 1 26	☐ 2	☐ 3	☐ 1 27	☐ 2	☐ 1 28	☐ 2

Name of activity ____________________ 29 ☐☐☐☐ 30

Occasions in the last month	Average time Hrs	Mins	Intensity Light	Intensity Medium	Intensity Heavy	Organized Yes	No	Competitive Yes	No
☐☐ 31	☐ 32	☐☐ 33	☐ 1 34	☐ 2	☐ 3	☐ 1 35	☐ 2	☐ 1 36	☐ 2

Name of activity ____________________ 37 ☐☐☐☐ 38

Occasions in the last month	Average time Hrs	Mins	Intensity Light	Intensity Medium	Intensity Heavy	Organized Yes	No	Competitive Yes	No
☐☐ 39	☐ 40	☐☐ 41	☐ 1 42	☐ 2	☐ 3	☐ 1 43	☐ 2	☐ 1 44	☐ 2

Name of activity ____________________ 45 ☐☐☐☐ 46

Occasions in the last month	Average time Hrs	Mins	Intensity Light	Intensity Medium	Intensity Heavy	Organized Yes	No	Competitive Yes	No
☐☐ 47	☐ 48	☐☐ 49	☐ 1 50	☐ 2	☐ 3	☐ 1 51	☐ 2	☐ 1 52	☐ 2

Name of activity ____________________ 53 ☐☐☐☐ 54

Occasions in the last month	Average time Hrs	Mins	Intensity Light	Intensity Medium	Intensity Heavy	Organized Yes	No	Competitive Yes	No
☐☐ 55	☐ 56	☐☐ 57	☐ 1 58	☐ 2	☐ 3	☐ 1 59	☐ 2	☐ 1 60	☐ 2

3

4. ACTIVITIES IN THE LAST YEAR

Please refer to the reference card for a list of activities. Answer the following for the physical activities you have done in the last 12 months.

(Do not include activities you have already listed.)

		Months in which activity was done												Number of occasions in last 12 months	Average number of minutes spent on each occasion			
		J	F	M	A	M	J	J	A	S	O	N	D		15 or less	16 to 30	31 to 60	61 or more
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
	Walking for exercise	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Jogging (using short strides)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Running (using long strides)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Bicycling	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
		J	F	M	A	M	J	J	A	S	O	N	D		15 or less	16 to 30	31 to 60	61 or more
		18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
	Home exercise (push-ups, sit-ups)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Exercise classes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Weight training	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Yoga	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
		J	F	M	A	M	J	J	A	S	O	N	D		15 or less	16 to 30	31 to 60	61 or more
		35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
	Golf (walking and carrying clubs)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Racquetball	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Squash	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Tennis	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
		J	F	M	A	M	J	J	A	S	O	N	D		15 or less	16 to 30	31 to 60	61 or more
		52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
	Baseball	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Softball	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Ice hockey	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Curling	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
		J	F	M	A	M	J	J	A	S	O	N	D		15 or less	16 to 30	31 to 60	61 or more
		69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
	Swimming at a pool	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Cross country skiing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Alpine/Downhill skiing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Ice skating	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
	Names of activities:	J	F	M	A	M	J	J	A	S	O	N	D		15 or less	16 to 30	31 to 60	61 or more
86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
		J	F	M	A	M	J	J	A	S	O	N	D		15 or less	16 to 30	31 to 60	61 or more
105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐
⊔⊔⊔	______	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	⊔⊔	☐	☐	☐	☐

Appendix D. Physical activities reference card for respondent's use

Canada Fitness Survey
Enquête condition physique Canada

PHYSICAL ACTIVITIES REFERENCE CARD

	WALKING--RUNNING--CYCLING
0101	Walking for exercise
0102	Walking upstairs for exercise
0103	Jogging
0104	Running
0105	Bicycling
0106	Roller Skating
0107	Orienteering
0108	Others--Specify

	CALISTHENICS--CONDITIONING
0201	Home Calisthenics
0202	Exercise Classes
0203	Weight Training
0204	Body Building
0205	Rope Skipping
0206	Yoga
0207	Relaxation Exercises
0208	Others--specify

	INDIVIDUAL AND DUAL SPORTS
0301	Archery
0302	Badminton
0303	Equestrian Events
0304	Fencing
0305	Figure Skating
0306	Golf, walking and carrying clubs
0307	Golf, walking and pulling clubs in a cart
0308	Golf, riding a power cart
0309	Gymnastics & Tumbling
0310	Handball (4-wall)
0311	Horseback Riding
0312	Racquetball
0313	Skateboarding
0314	Squash
0315	Table Tennis
0316	Tennis
0317	Weight lifting
0318	Others--Specify

	TEAM SPORTS
0401	Baseball
0402	Basketball
0403	Broomball
0404	Cricket
0405	Curling
0406	Football
0407	Football (touch)
0408	Handball
0409	Floor Hockey
0410	Field Hockey
0411	Ice Hockey
0412	Lacrosse
0413	Officiating (specify)
0414	Ringette
0415	Rugby
0416	Soccer
0417	Softball
0418	Volleyball
0419	Water Polo
0420	Others--Specify

	AQUATICS
0501	Canoeing/rowing-pleasure
0502	Canoeing - camping trip - portage
0503	Diving
0504	Kayaking
0505	Rowing
0506	Sailing
0507	Scuba Diving
0508	Sculling
0509	Snorkeling
0510	Swimming at beach
0511	Swimming at pool
0512	Water Skiing
0513	Others--Specify

	WINTER SPORTS
0601	Alpine Skiing
0602	Cross Country Skiing
0603	Freestyle Skiing
0604	Skating
0605	Snowmobiling
0606	Snowshoeing
0607	Tobogganing
0608	Others--Specify

	OUTING ACTIVITIES
0701	Backpacking
0702	Camping with Backpacking
0703	Camping - prepared grounds
0704	Fishing in boat
0705	Fishing from river bank
0706	Fishing in stream (boots)
0707	Hiking
0708	Hunting - Small Game
0709	Hunting - Large Game
0710	Motor Cycling
0711	Mountaineering/Rockclimbing
0712	Trail Biking
0713	Others--Specify

	DANCE
0801	Ballet
0802	Ballroom Dancing
0803	Disco and popular dancing
0804	Folk Dancing
0805	Modern Dance
0806	Modern Jazz
0807	Square Dancing
0808	Others--Specify (Disco?)

	COMBATIVES & SELF DEFENSE
0901	Wrestling
0902	Boxing
0903	Judo
0904	Karate
0905	Others--Specify

	GAMES OF LOW ORGANIZATION
1001	Spontaneous Ball Games
1002	Street Hockey
1003	Catch (football, baseball, etc.)
1004	Frisby
1005	Tag Games
1006	Climbing - playground equipment

	GAMES
1101	Bowling
1102	Horseshoes
1103	Croquet
1104	Bocce
1105	Others--Specify

	HOME RELATED ACTIVITIES
1201	Light Housework (cooking, ironing, washing dishes, making beds, mowing lawn with power mower)
1202	"Heavy housework and handywork" (washing windows, washing and waxing floors, shopping and carrying grocery bags, painting inside-outside, carpentry, repairing-remodelling)
1203	Spading, digging, filling in, weeding and cultivation of garden, raking lawn
1204	Snow shoveling
1205	Clearing snow with blower
1206	Mowing lawn (pushing lawn mower)
1207	Wood Cutting
1208	Others--Specify

Appendix E. Multiples of the resting metabolic rate

Background

In early 1982, the Canada Fitness Survey organization convened a panel of experienced exercise scientists to establish an algorithm for calculating energy expenditure values based on the questionnaire data. The calculation of kcal/kg/day (described in this chapter in the section on protocol for activity assessment) is based on this methodology. For the purposes of making kilocalorie calculations, the panel also identified the energy requirements in METS of various leisure-time activities and common household chores. These values were obtained from an extensive review of the relevant literature, a task simplified by a related undertaking performed 2 years previously under the auspices of the provincial fitness agency, Kino-Quebec. The report of the earlier work group (34) was the principal reference; other sources were used as required and are shown in the table.

For each activity, METS values at three intensity levels were required. The definitions of intensity on the questionnaire were: "Light—slight change from normal state; medium—some perspiration, above-normal breathing; and heavy—heavy perspiration, heavy breathing." When METS existed for several intensity levels (e.g., five speeds of bicycling), the panel chose values to correspond to these three intensity levels. If only one METS level was found for an activity, its value was interpolated to the three intensities. Table A shows the values obtained from the literature, the sources, and the final values established by the panel after two separate considerations.

Note: This appendix is adapted from a report to the Canada Fitness Survey by Claude Bouchard, Gaston Godin, Fernand Landry, Roy J. Shephard, and James S. Skinner.

Table A. METS values used in Canada Fitness Survey

				Canada Fitness Survey value		
Activity	Intensity	METS	Source	Light	Medium	Heavy
101. Walking for exercise	3 km/h	1.8	34	3	4	5
	5 km/h	3.2	34			
	7 km/h	5.3	34			
102. Walking upstairs for exercise		4.7	34	4	6	8
103. Jogging (level slope)	9 km/h	8.8	34	7	10	12
	11 km/h	11.2	34			
104. Running (level slope)	13 km/h	12.9	34	12	14	16
	15 km/h	14.6	34			
105. Bicycling	10 km/h	4.8	34	3	7	10
	15 km/h	5.9	34			
	20 km/h	7.1	34			
	25 km/h	8.4	34			
	30 km/h	9.8	34			
106. Roller skating	12.9 km/h	5.7	55	5	6.5	8
	13.9 km/h	7.6	55			
	16.1 km/h	9.5	55			
	17.7 km/h	10.5	55			
107. Orienteering		7.0	56	8	10	12
201. Home calisthenics		1.75–5.87	57	3	5	8
202. Exercise classes				4	6	9
203. Weight training		10.9	58	3	5	7
204. Body building				3	5	7
205. Rope skipping	66/min	9.8	34	7	10	12
	84/min	10.5	34			
	100/min	11.0	34			
	120/min	11.4	34			
	125/min	11.7	34			
	130/min	11.8	34			
	135/min	12.0	34			
	145/min	12.1	34			
206. Yoga		3.2	56	2	2	2
207. Relaxation exercises				2	2	2
209. Gym classes				2	2	2
210. Health, exercise club				3	5	7
301. Archery		4.3	34	3	3	3
302. Badminton		5–9	34	3	6	9
	Doubles	3–4	59			
	Singles	4–5	59			
	Competitive	6–7	59			
303. Equestrianism		7	56	3	5	7
304. Fencing		6–10	34	5	7	10
		8–9	59			
305. Figure skating		12.9	34	4	6	10
306. Golf	Carrying clubs	5.1	34	4	4	4
307. Golf	Pulling cart	3–4	59	4	4	4
308. Golf	Riding cart	2–3	59	3	3	3
309. Gymnastics		7	56	5	7	10
310. Handball (4 wall)		7.7	34	6	9	12
	Social	8–9	59			
311. Horseback riding	Walk	3.2	34	3	5	7
	Trot	6.9	34			
	Gallop	8.6	34			
312. Racquetball		8–12	34	6	9	12
		12.3	60			
313. Skateboarding				5	6.5	8
314. Squash		8.5	34	6	9	12
		12.3	61			
	Social	8–9	59			
315. Table tennis		4.7	34	4	6	9
316. Tennis		6.8	34	4	6	10
	Singles	6–7	59			
	Doubles	4–5	59			
317. Weight lifting	Light	2.7	56	3	5	7
		7.2	34			
319. Track and field	Marathon	13.3	34	4	6	8

Table A. METS values used in Canada Fitness Survey—Continued

Activity	Intensity	METS	Source	Canada Fitness Survey value		
				Light	Medium	Heavy
	High jump	04.1	34			
	Long jump	15.0	34			
	Shot put	03.8	34			
320. Hang gliding				3.5	3.5	3.5
401. Baseball		4–7	34	3	4	5
		4	62			
402. Basketball		11.1	34	6	8	11
		8.9	59			
403. Broomball		6.3	56	5	7	9
404. Cricket		6.1	34	3	4	5
405. Curling		7.4	34	4	5	6
406. Football		6–7	34	5	6	7
407. Football (touch)		7–8	56	5	6	8
408. Handball		9.9	34	6	8	11
409 Floor hockey	Forwards	10.3	56	6	8	10
410. Field hockey	Goalie	6	56	6	8	10
411. Ice hockey		12.9	34	6	8	10
412. Lacrosse	Forwards	12.6	56	6	8	10
413. Officiating				2	4	7
414. Ringette	Forwards	12.6	56	5	7	9
415. Rugby		12.6	56	6	8	11
416. Soccer		10.3	34	5	7	11
417. Softball		3–6	34	3	4	5
		4	62			
418. Volleyball		6	34	5	6	8
419. Water polo		9.8	56	6	8	11
501. Canoeing/rowing	3.0 km/h	2–4	34	3	4	6
	4.0 km/h	2–3	59			
	5.0 km/h	4–6	34			
	6.4 km/h	5–6	59			
	7.0 km/h	9–11	34			
	8.0 km/h	7–8	59			
502. Canoeing (camping)				4	6	9
503. Diving				4	4	4
504. Kayaking	12.5 km/h	7.8	34	6	8	11
	15.0 km/h	11.0	34			
505. Rowing (competitive)	4 km/h	5.5	34	7	10	13
	8 km/h	10.3	34			
	12 km/h	13.5	34			
	16 km/h	16.4	34			
	20 km/h	19.1	34			
506. Sailing (small boat)		3–4	59	3	4	6
507. Scuba diving		11	34	4	5	6
508. Sculling		3.7–9.8	58	4	6	10
509. Snorkeling				4	5	6
510. Swimming (beach)				2	3	4
511. Swimming (pool)—average for breast stroke, forward and back crawl	2.0 km/h	4.3	34	3	5	9
	2.5 km/h	6.8	34			
	3.0 km/h	8.9	34			
	3.5 km/h	11.5	34			
	4.0 km/h	13.6	34			
512. Water skiing		7.9	34	5	7	9
		6–7	59			
514. Windsurfing				4	5	7
515. Synchronized swimming, aquasize, etc.	Legs only	8.7	34	4	6	8
	Arms only	9.8	34			
601. Alpine skiing		5–9	34	4	6	8
	Light	6–7	59			
	Vigorous	7–8	59			
602. Cross-country skiing	4 km/h	5.5	34	5	9	13
	6 km/h	7.7	34			
	8 km/h	9.9	34			
	10 km/h	12.2	34			
	12 km/h	14.3	34			
	14 km/h	16.5	34			

Table A. METS values used in Canada Fitness Survey—Continued

Activity	Intensity	METS	Source	Canada Fitness Survey value		
				Light	Medium	Heavy
603. Freestyle skiing				4	6	9
604. Skating	18 km/h	4.0	34	4	7	13
	25 km/h	4.8	34			
	28 km/h	9.2	34			
	32 km/h	10.8	34			
	36 km/h	15.2	34			
605. Snowmobiling		2–3	34	3	3.5	5.0
606. Snowshoeing	4 km/h	9.5	34	5	7	10
607. Tobogganing		7	62	5	6	7
701. Backpacking (5% slope, 20 kg pack)	6.4 km/h	8	63	6	8	10
	7.2 km/h	9.6	63			
	8.0 km/h	11.6	63			
	9.6 km/h	13.1	63			
	11.2 km/h	15.5	63			
702. Camping with backpacking				3	3	3
703. Camping (prepared grounds)				2	2	2
704. Fishing from boat		2–3	34	2.5	2.5	2.5
		2.5	62			
705. Fishing from bank		2–3	34	2.5	2.5	2.5
		2.5	62			
706. Fishing in stream		4–7	34	3.5	3.5	3.5
	Standing	3–4	59			
	Walking	5–6	59			
707. Hiking, beachcombing		6	62	3	6	8
708,9. Hunting (small and large game)		3–7	34	3	5	7
		6	62			
710. Motorcycling		2.2	34	2.5	4	7
		1.5–2	59			
711. Mountaineering, rock climbing		7–8	59	7	8	10
		8.6	56			
712. Trail biking		6–8	34	4	5	7
713. Car driving		1.5–2	59	2.5	2.5	2.5
714. Trapping				5	5	5
801. Ballet		6–8	34	5	6	8
		5–6.7	64			
802. Ballroom dancing		3–5	34	3	4	5
		4–5	59			
803. Disco, popular dancing		3–8	34	3	5	7
	Rhumba	5.7	59			
804. Folk dancing		4.8	34	3	5	7
		6–7	59			
805. Modern dancing		6–8	34	5	6	8
806. Modern jazz		6–8	34	5	6	8
807. Square dancing		4.8	34	3	5	7
		6–7	59			
809. Dancercize	Low	3.9	65	4	6	9
	Medium	6.0	65			
901. Wrestling		8–12	34	6	9	12
902. Boxing		13.4	34	6	9	12
903. Judo		10.5	34	6	8	12
904. Karate		8–12	34	5	8	12
		5–8.8	66			
1001. Ball games				3	4	5
1002. Street hockey, ball hockey				3	4	5
1003. Catch (ball)				3	4	5
1004. Frisbee				3	4	5
1005. Tag				3	4	5
1006. Climbing (playground)				3	4	5
1007. Other unorganized games				3	4	5
1101. Bowling		2–4	34	2	2.5	3
		2–3	59			
1102. Horseshoes		2–3	34	2	2.5	3
		3–4	59			
1103. Croquet		2–3	34	2	2.5	3
1104. Bocce				2	2.5	3

Table A. METS values used in Canada Fitness Survey—Continued

Activity	Intensity	METS	Source	Canada Fitness Survey value		
				Light	Medium	Heavy
1201. Light housework	Cooking	2.5	34	2	2.5	4
	Ironing	2.0	34			
	Dishes	2.1	34			
	Making beds	3–5	34			
	Mowing lawn (power mower)	3–5	34			
1202. Heavy housework	Washing windows	4.9	55	3	3.5	5
		3–4	59			
	Washing, waxing floors	3.3	55			
	Shopping for groceries	2.2–7.4	67			
	Painting	3.5–5.0	34			
		4–5	59			
	Carpentry	5–6.5	34			
		4–5	59			
	Repairing, remodeling	3.5–5	34			
		2–3	59			
1203. Gardening	Digging	4.4	34	3	5	7
		5–6	59			
	Weeding	3.5	34			
	Raking	3.5	34			
1204. Snow shoveling		5.1	34	4	6	8
		6–7	59			
1205. Snow blowing				4	4	4
1206. Mowing (push mower)		5–7	34	3	4	5
		6–7	59			
1207. Wood cutting		5–6.5	34	4	5	7
		6–7	59			

Appendix F. Excerpt from 1985 Supplement to National Health Interview Survey

Section R. EXERCISE — Continued

NOTE *— ASK ALL OF 2a BEFORE GOING TO 2b–d.*

NOTE*: ASK 2b–d FOR EACH ACTIVITY MARKED "YES" IN 2a.*

Read to respondent: **These next questions are about physical exercise.** *Hand calendar.* **2a. In the past 2 weeks (outlined on that calendar), beginning Monday, *(date)*, and ending this past Sunday, *(date)*, have you done any (of the following exercises, sports, or physically active hobbies) —**	YES NO	7	**b. How many times in the past 2 weeks did you [play/go/do] *(activity in 2a)*?**		**c. On the average, about how many minutes did you actually spend *(activity in 2a)* on each occasion?**		**d. {What usually happened to your heart rate or breathing when you *(activity in 2a)*?} Did you have a small, moderate, or large increase, or no increase at all in your heart rate or breathing?**	
(1) Walking for exercise?	1 ☐ 2 ☐		(1) ______ Times	8–9	______ Minutes	10–12	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	13
R2	Refer to age. 1 ☐ SP is 75+ *(23)* 8 ☐ Other *(2)*	14						
(2) Jogging or running?	1 ☐ 2 ☐	15	(2) ______ Times	16–17	______ Minutes	18–20	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	21
(3) Hiking?	1 ☐ 2 ☐	22	(3) ______ Times	23–24	______ Minutes	25–27	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	28
(4) Gardening or yard work?	1 ☐ 2 ☐	29	(4) ______ Times	30–31	______ Minutes	32–34	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	35
(5) Aerobics or aerobic dancing?	1 ☐ 2 ☐	36	(5) ______ Times	37–38	______ Minutes	39–41	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	42
(6) Other dancing?	1 ☐ 2 ☐	43	(6) ______ Times	44–45	______ Minutes	46–48	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	49
(7) Calisthenics or general exercise?	1 ☐ 2 ☐	50	(7) ______ Times	51–52	______ Minutes	53–55	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	56
(8) Golf?	1 ☐ 2 ☐	57	(8) ______ Times	58–59	______ Minutes	60–62	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	63
(9) Tennis?	1 ☐ 2 ☐	64	(9) ______ Times	65–66	______ Minutes	67–69	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	70
(10) Bowling?	1 ☐ 2 ☐	71	(10) ______ Times	72–73	______ Minutes	74–76	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	77
(11) Biking?	1 ☐ 2 ☐	78	(11) ______ Times	79–80	______ Minutes	81–83	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	84
(12) Swimming or water exercises?	1 ☐ 2 ☐	85	(12) ______ Times	86–87	______ Minutes	88–90	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	91
(13) Yoga?	1 ☐ 2 ☐	92	(13) ______ Times	93–94	______ Minutes	95–97	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	98
R3	Refer to age. 1 ☐ SP is 65–74 *(23)* 8 ☐ Other *(14)*	RT77 3–4 5						
(14) Weight lifting or training?	1 ☐ 2 ☐	6	(14) ______ Times	7–8	______ Minutes	9–11	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	12
(15) Basketball?	1 ☐ 2 ☐	13	(15) ______ Times	14–15	______ Minutes	16–18	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	19
(16) Baseball or softball?	1 ☐ 2 ☐	20	(16) ______ Times	21–22	______ Minutes	23–25	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	26
(17) Football?	1 ☐ 2 ☐	27	(17) ______ Times	28–29	______ Minutes	30–32	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	33
(18) Soccer?	1 ☐ 2 ☐	34	(18) ______ Times	35–36	______ Minutes	37–39	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	40
(19) Volleyball?	1 ☐ 2 ☐	41	(19) ______ Times	42–43	______ Minutes	44–46	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	47
(20) Handball, racquetball, or squash?	1 ☐ 2 ☐	48	(20) ______ Times	49–50	______ Minutes	51–53	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	54
(21) Skating?	1 ☐ 2 ☐	55	(21) ______ Times	56–57	______ Minutes	58–60	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	61
(22) Skiing?	1 ☐ 2 ☐	62	(22) ______ Times	63–64	______ Minutes	65–67	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	68
(23) Have you done any (other) exercises, sports, or physically active hobbies in the past 2 weeks (that I haven't mentioned)? Anything else? ☐ Yes — **What were they?** ☐ No ______________		69–70	(23) ______ Times	71–72	______ Minutes	73–75	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	76
______________		77–78	(23) ______ Times	79–80	______ Minutes	81–83	1 ☐ Small 3 ☐ Large 2 ☐ Moderate 4 ☐ None	84

Page 12

FORM HIS-1(SB) (1985) (10-15-84)

Chapter 18

An International Perspective on Critical Issues in Fitness Testing of U.S. Adults

Roy J. Shephard, M.D., Ph.D.
University of Toronto

The purpose of the present chapter is to discuss some of the lessons that can be learned from previous surveys of fitness and physical activity in other nations. Particular emphasis will be placed on the Canadian scene, partly because I have greater familiarity with this work and partly because Canada has been very active in the study of national fitness and activity patterns. Specific issues to be addressed include survey objectives; decisions on methodology and statistical design; questions of human experimentation; matters of organization and field administration; and considerations of data processing, analysis, and interpretation. Based on a critique of the studies reviewed, recommendations for action by the U.S. Government will be made, with an indication of options, where appropriate.

Survey objectives

Hidden agendas

A clear definition of survey objectives is vital to an effective study. Unfortunately, previous surveys have often failed to articulate their purpose clearly, and at times a "hidden agenda" has dominated both methodology and the presentation of results.

In an earlier era, Canadian interest in fitness was stimulated by the poor physical condition of military recruits (Shephard, 1977). More recent research was provoked by comments of H.R.H. the Duke of Edinburgh on the fitness of young Canadians made to the annual meeting of the Canadian Medical Association in 1959, a popular taunt that the 25-year-old Canadian was less fit than the 65-year-old Swede, and a national concern over poor performances in successive Olympic Games. There was thus substantial political pressure to prove Canadians were not as unfit as had been proclaimed and to document gains of national fitness in response to corrective programs initiated by the federal government. Other administrative concerns included lagging national productivity and the ever-rising cost of prepaid medical care. It was further hoped that appropriately organized testing might demonstrate an effect of new fitness programs in boosting industrial efficiency and in containing medical expenditures.

Overt agendas

To these various unspoken objectives must be added many reasons for surveying fitness. Information on activity patterns may be needed by those planning facilities or publicity. In children, the intent may be to supplement anthropometric observations on normal growth and development with norms of effort tolerance. In adults, a knowledge of the population distribution of fitness at various ages is becoming vital information in legal battles over gender and age discrimination in employment. The study of activity patterns is also an important (and often overlooked) factor in determining the nutritional needs of the individual and of the community. Only testing can establish national status relative to other populations and allow intragroup comparisons based on reported activity patterns, self-perceptions of fitness, or socioeconomic status. A plan for repeated testing implies setting a benchmark of national condition. Changes of score over periods of 5 years or longer can then be related to program changes, differences in reported activity patterns, differences in hospital operating costs, and so on.

Articulation of objectives for the NHANES will thus influence a number of important choices in survey design and methodology. Specifically:

a. Is the prime intention to make a "once-only" characterization of adult fitness, filling a gap in the U.S. health information system, or will measures be repeated?
b. Is the intent to relate information to the health field (perceived health, incidence of disability and disease, medical costs) or to the industrial field (where tests might emphasize mental alertness, physical working capacity and life satisfaction rather than health)?

Definition of fitness

The definition of fitness applicable to the NHANES will be greatly influenced by the overt and hidden agendas noted above. It was long argued that, in terms of improved industrial performance, controlling obesity, and reducing the risk of future ischemic heart disease, the prime variable of interest was cardiorespiratory endurance, assessed by the measurement or prediction of maximum oxygen intake (Shephard, 1977). However, recent studies of military personnel have suggested that the main factor limiting employment of women in "heavy" work is a lack of muscular strength and endurance (Nottrodt and Celentano, 1984); and in old people, participation in ordinary daily activities such as traveling by bus may be handicapped by lack of flexibility.

Some large surveys of fitness (for example Métivier and Orban, 1971) have looked at a cardiorespiratory variable such as the physical working capacity measured on a cycle ergometer. However, to allow a broader application of its findings, the Canada Fitness Survey (1983) included observations on aerobic power (the Canadian Home Fitness Test (CHFT)), muscle strength and endurance (handgrip force, situps, pushups), obesity (height, body mass, skinfold readings), flexibility (sit and reach test), and blood pressure (table 1).

Moreover, subjects completed a detailed questionnaire covering physical activity at work, at home and at play. There is still much disagreement as to how far fitness reflects regular daily activity, and how far it can be attributed to a constitutional advantage. However, it seems probable that at least a third, and possibly a half, of the variance in fitness scores is the result of habitual exercise rather than genetic factors (Shephard, 1978). The CFS thus sought to link fitness scores with activity data; in this way, it was hoped to determine both patterns of activity that maximized human potential and the community prevalence of such activity patterns. Given this information, CFS reasoned it would become possible to predict the change of fitness (and related health variables) resulting from an optimization of physical activity patterns.

Table 1. Measurements made in Canada Fitness Survey

Variable	Comment
Questionnaire data:	
Lifestyle	
Physical activity habits	
Anthropometric data:	
Standing height	Gentle traction, estimated if measurement not possible
Body mass	Estimated if necessary
Skinfolds (biceps, triceps, subscapular, suprailiac, medial calf)	All measures repeated once, twice if difference exceeds 1 mm.
Girths (relaxed upper arm, chest, waist, hips, right thigh, right calf-maximum)	Steel tape used
Bone diameters (biepicondylar humerus-elbow, biepicondylar femur-knee)	
Performance test items:	
Blood pressure (for screening)	Systolic and diastolic (phase 4) after 5 min of rest before step test, also at 30–60 sec and 2.5–3.0 min after completing the step test.
Step test (cardiovascular)	Canadian Home Fitness Test (advanced version — up to 3 bouts of 3 min each) Results in final heart rate or estimated $\dot{V}O_2$max. Also resting, intermediate and 3 min post-test heart rate.
Grip strength (muscular strength)	Combined maximum, right and left hands, after two attempts each.
Push ups (muscular endurance)	Males from toes, females from knees
Sit and reach (trunk flexion)	Using a modified Wells and Dillon flexometer, after stretching.
Sit ups (muscular endurance)	60 sec maximum.

The ideal recommendation to the NHANES would plainly seem a broad, balanced battery of fitness tests, coupled with a simple but reliable and carefully validated index of daily habitual activity.

Test methodology

Test reliability and validity

If test scores are to have meaning for survey organizers, it is axiomatic that measurements should show a high test-retest reliability and that scores indicate what they are supposed to reflect. CFS thus devoted considerable time to development of a standard protocol (Standardized Test of Fitness, 1981).

Poor reliability may reflect subjective factors, such as learning, habituation and variations of motivation, observer errors (failure to follow the required protocol or differences in its interpretation); technical factors (for example, change in calibration of a dynamometer, ergometer, or caliper); and environmental factors (heat, cold, anxiety and so on). Lack of validity may reflect a technical error (for example, a method of gas analysis may systematically underestimate oxygen concentrations, thereby yielding an excessive estimate of oxygen consumption); alternatively, the nature of the measurement may be misunderstood (for example, a test may be thought to reflect fitness when in fact the major influence is stature, body mass, or motivation). Partly from ignorance and partly (one suspects) from a fear of what might be discovered, the majority of surveys have ignored most of these factors.

Age, height, and mass

In Canada and the United States, there is usually little difficulty in assigning a valid age to subjects. However, in some more exotic surveys, a lack of public records has led to approximations of age based on first school attendance (Shephard, 1978), with 30–40 year exaggerations in the oldest age categories, because of a veneration of the elderly (Mazess & Mathisen, 1982). The CFS and several other large Canadian surveys reported age as the number of completed years (for example, 13 years = 13.00–13.99 years). However, many longitudinal anthropometric surveys have collected information at or near the individual's birthday (a discrepancy averaging 0.5 years).

The accurate measurement of stature requires a sturdy stadiometer, where a tray of known mass, sliding at a right angle to a vertical scale, is allowed to rest on a carefully positioned head. Portable anthropometry kits are available, but many field surveys have relied on very simple expedients. For example, in the CFS a set square was moved over tape-measured marks on a vertical wall; and in some instances, subjects' recollection of their height was substituted. Recollections may be biased by measurements made while wearing shoes, and they may underestimate the decrease of stature with age. However, we saw no evidence that population averages in the CFS were grossly inaccurate.

Body mass can be approximated by use of a relatively cheap bathroom scale, provided that readings are cross-checked against values for an observer, recently weighed on a precision instrument at the same time of day. Under field conditions, precise instructions should be given regarding allowances made for the mass of any clothing that is not removed. In the CFS, specific instructions included placing a board under the scale if the floor was carpeted. However, a proportion of subjects were asked to recollect their "weight," an estimate that often differs by several kg from the actual value (Kreitler and Kreitler, 1970; Sidney and Shephard, 1977, Stunkard and Albaum, 1981). Furthermore, body mass varies systematically over the course of a day because of the ingestion of food and fluids. However, most survey data have been collected at random throughout normal working hours. Average values are thus likely to overestimate the true fasting body mass by 1–2 kg.

Performance tests

Many large-scale surveys of "fitness," both in Canada (CAHPER, 1965; 1980) and the United States (AAHPER, 1976) have used a battery of performance tests. Attempts have been made to standardize technique, both by centralized observer training and provision of detailed instruction booklets, but there has been no post-hoc analysis of procedural variations that may have developed subsequent to training of the survey teams. It is also suspected that subjects show substantial "learning curves" for a number of items in the test battery, and there has been a strong suspicion that regular repetition of the test items accounted for much of the apparent "improvement of fitness," both in Canada (from 1965 to 1980) and in the United States (from 1958 to 1965). Unfortunately, most adults lack recent experience with gymnasium-type tests (Drake et al., 1968). Scores are also much dependent on environment (temperature if performed indoors, as would be likely in the NHANES, wind-speed and ground surface if performed outdoors) (Strydom, 1978). Confines of space may require alterations of movement patterns, with gross changes in average scores (Drake et al., 1969). Home testing largely eliminates the element of competition important to a maximal score. Finally, the results of most performance tests depend largely on stature and body mass (Cumming et al.; Drake et al., 1969). Indeed, it is uncertain whether such measurements contribute knowledge about any aspect of fitness other than an understanding of technique (how to run or how to jump) once statistical allowance has been made for the influence of height and body mass.

Skinfold readings

Technical problems of skinfold measurement can arise from the use of a nonstandard gauge (as in the early

Table 2. Errors in plastic skinfold caliper (Ross) relative to standard Harpenden instrument

Variable	Real aperture 5mm		Real aperture 25 mm		Real aperture 40 mm	
	Ross	Harpenden	Ross	Harpenden	Ross	Harpenden
Aperture (mm)	6.0	5.1	29.0	26.2	46	41.5
Pressure (g·mm^{-2})	—	—	11.9	8.2	—	—

Source: Based on data of Léger et al., 1982.

observations of Pãrízková, 1977). The Cambridge and Harpenden calipers each give comparable readings, but they require periodic checking to ensure that the zero position of the needle has not altered and that spring tension has not changed. It is probably desirable to purchase standard calipers for the NHANES, although surprisingly good results have been reported for three recent low-cost plastic designs of caliper (Ross, McGaw, and Ponderal instruments; Rombeau et al., 1977; Leger et al., 1982) (table 2).

Interobserver errors are a larger problem in a big survey. Discrepancies of 25% or more can arise from differences of measuring site, differences in support given to the skinfold, differences in the period of skin compression that is allowed, variations of subject position, and nonuniform training of observers (Kemper and Pieters, 1974; Jamison and Zegura, 1974; Jackson et al., 1978; Housh et al., 1983). Weiner and Lourie (1981) have provided precise instructions on most of these points, but nevertheless it is desirable that all observers check their ability to obtain comparable readings on a standard panel of subjects. Surveys to date have generally ignored this precaution.

If skinfolds are to be used to predict body fat, there remains much disagreement on the equations to be used for this purpose. Depending on the formula that is chosen, the same skinfold data can yield estimates of body fat that differ by a factor of 2. The CFS measured the four folds recommended to the International Biological Program (biceps, triceps, subscapular, and suprailiac) plus one other fold (medial calf). The present author has thus used the familiar Durnin and Womersley (1974) formulas for interpretation of CFS body fat readings:

$$\text{Body density: Men} = 1.1631 - 0.0632\ \Sigma\ 4$$
$$\text{Women} = 1.1599 - 0.0717\ \Sigma\ 4$$

where $\Sigma\ 4$ is the sum of the four skinfolds

$$\text{Body fat \%} = (4.570/D - 4.142)\ 100$$

where D is the estimate of body density.

Other CFS participants have continued to use formulas hallowed by the practice of their particular laboratory. The Durnin and Womersley (1974) formula now has rather wide international acceptance, but one drawback in the NHANES context is that subjects (particularly the men) were thinner than the average American. Unfortunately, obesity seems to influence both the relative distribution of subcutaneous fat and the balance between deep and superficial fat. Jackson et al. (1978) have argued that summing over multiple folds reduces interobserver errors. Possible formulas are summarized in tables 3–5. Lohman (1981) has argued that the optimum formula for young U.S. adults is a quadratic but nonlogarithmic summation of triceps, abdominal, and subscapular folds:

$$D = 1.0982 - 0.000815\ (\Sigma s) + 0.0000084\ (\Sigma s)^2$$

If his suggestion is adopted, it would seem wise to measure also the biceps and suprailiac folds, to allow comparisons with European and Canadian data. Some

Table 3. Prediction formulas for body density based on typical male subject

Formula	Calculated body density	Age of population originally tested	Author
$1.1533–0.0643\ (\log_{10} \Sigma\ S)$	1.056	15	Durnin & Rahaman (1967)
$1.1610–0.0632\ (\log_{10} \Sigma\ S)$	1.066	22	Durnin & Rahaman (1967)
$1.1447–0.0612\ (\log_{10} \Sigma\ S)$	1.052	Adult	Durnin (personal communication)
$1.130–0.055\ (\log_{10} S_t)–0.026\ (\log_{10} S_s)$	1.053	13–16	Parizkova (1961)
$1.0923–0.00202\ (S_t)$	1.077	22	Pascale et al (1956)
$1.0896–0.00179\ (S_s)$	1.075	22	Pascale et al (1956)
$1.0962–0.0027\ (S_t)$	1.075	Young men	Sloan (1967)
$1.0955–0.00236\ (S_t)$	1.066	Young men	Katch & McArdle (1973)
$1.0811–0.00195\ (S_t)$	1.066	Young men	Wilmore & Behnke (1968)
$1.0936–0.00186\ (S_t)$	1.079	Young and middle aged	Pollock et al (1976)
$1.0842–0.00163\ (S_i)$	1.072	Adult	Lohman (1981)
$1.1613–0.0632\ (\log_{10} S^1)$	1.062	Young and middle aged	Durnin & Womersley (1974)

S_b biceps = 5.0 mm, S_t triceps = 7.8 mm, S_s subscapular = 11.9 mm, S_i suprailiac = 12.7 mm

$\Sigma\ S = S_t + S_s + S_i$

$\Sigma\ S^1 = S_b + S_t + S_s + S_i$

Table 4. Prediction formulas for body density based on typical female subject

Formula	Calculated body density	Age of population originally tested	Author
$1.1369–0.0598\ (\log_{10} \Sigma S)$	1.040	15	Durnin & Rahaman (1967)
$1.1581–0.0720\ (\log_{10} \Sigma S)$	1.042	22	Durnin & Rahaman (1967)
$1.1309–0.0587\ (\log_{10} \Sigma S)$	1.036	adult	Durnin (personal communication)
$1.114–0.031\ (\log_{10} S_t) - 0.041\ (\log_{10} S_s)$	1.034	13–16	Pařízková (1961)
$1.0764–0.00088\ (S_t) - 0.00081\ (S_s)$	1.054	20	Sloan, Burt & Blyth (1962)
$1.0692–0.00138\ (S_t)$	1.048	Young women	Katch & McArdle (1973)
$1.0508–0.00080\ (S_t)$	1.038	Young women	Wilmore & Behnke (1970)
$1.0772–0.00183\ (S_t)$	1.049	Young adult, middle aged adult	Pollock et al., (1975)
$1.0524–0.00072\ (S_t)$	1.041	Young women	Young (1963)
$1.1599–0.0717\ (\log_{10} \Sigma S^1)$	1.038	Young and middle aged	Durnin & Womersley (1974)

S_b biceps = 8.0 mm, S_t triceps = 15.6 mm, S_s subscapular = 11.3 mm, S_i suprailiac = 14.6 mm
$\Sigma S = S_t + S_s + S_i$
$\Sigma S^1 = S_b + S_t + S_s + S_i$

authors have suggested that, because of the wide range of options and the limited precision of all formulas, data should be presented simply as skinfold readings. The main reason for attempting to estimate body fat is that the lean body mass can then be calculated. This provides an objective measure of body muscularity that requires no cooperation from the subject. However, skinfold readings and density estimates should also be reported.

Girths

The CFS included measurements of six girths (relaxed upper arm, chest, waist, hips, right thigh, and right calf). Several of the potential formulas for body fat prediction included one or more girths than from skinfold readings. The obvious difficulty is that girth readings, and there have been suggestions that untrained observers make better estimates of body fat from girth, reflect a combination of fat, muscle, and bone; and in our experience (Murray and Shephard, 1984), correlations with hydrostatic estimates of body fat are marginally better for skinfolds than for circumferences.

Blood pressure

Blood pressures are generally somewhat lower under the conditions of a mass survey than in a clinical office, and we have found normal readings in a high proportion of subjects who had been told by their doctors that they were hypertensive (Shephard, Cox, and Simper, 1981). This has some practical importance in that a report of prior hypertension is often used as one criterion for exclusion from a paramedical exercise test (Chisholm et al., 1975). Thus, if blood pressure is not checked by the survey investigator, as many as 20% of adults may be excluded from testing on this basis alone (Shephard, Cox, and Simper, 1981). The drop of pressure relative to medical reports probably reflects a lower level of anxiety away from the doctor's office.

Other technical factors that can influence blood pressure readings are the duration of rest prior to measurement, environmental temperature, the posture of the subject, the arm used for measurement, an appropriate choice of cuff size, recording of the 4th or 5th phase of the Kortkov sounds, and digit preference on the part of the observer. The CFS gave precise instructions on posture, use of left arm, and choice of the 4th sound, but some surveys have ignored many of these points.

Apart from the obvious health importance of hypertension, the pulse pressure gives a crude indication of cardiac stroke volume. In general, fit individuals with a large stroke volume also have a large pulse pressure.

Table 5. Percent body fat if density = 1.060

Formula	Percent fat	Author
$(5.548/D - 5.044)\ 100$	19.0	Rathburn & Pace (1945)
$(4.971/D - 4.519)\ 100$	17.1	Brozek et al. (1963a)
$(4.570/D - 4.142)\ 100$	16.9	Brozek et al. (1963a)
$(4.0439/D - 3.6266)\ 100$	18.8	Grande (1961)
$(1.10 - D)\ 500$	20.0	MacMillan et al. (1965)
$(4.95/D - 4.50)\ 100$	17.0	Siri (1961)

Aerobic power

Aerobic power, or maximum oxygen intake, is perhaps the most important aspect of fitness in the average adult (Shephard, 1977). In the laboratory, measurement by a progressive treadmill test is relatively unequivocal. Direct determinations of oxygen intake are continued at ever-increasing work rates until a plateau of oxygen consumption is demonstrated. The resultant "maximum" cannot be boosted appreciably by performing other types of exercise simultaneously, and scores obtained during this type of activity seem generalizable to other types of exercise (Shephard, 1982). The only limitation on the validity of the data may arise from failure to standardize methods of gas analysis (Cotes and Woolmer, 1962).

Cooper (1968) proposed an all-out field test of aerobic power, but there has been some concern regarding the safety of 12 minutes of all-out exercise by older untrained adults in the absence of medical supervision. Moreover, the construct validity of scores for nonathletic adults is often compromised by lack of knowledge of an appropriate running pace and poor motivation (Shephard, 1982).

Submaximum exercise tests may be interpreted in their own right or used to predict maximum oxygen intake (generally through a manipulation of the heart rate-oxygen consumption relationship such as the Åstrand (1960) nomogram). Three large-scale Canadian surveys (Howell and MacNab, 1968; Métivier and Orban, 1971; Gauthier et al., 1983) used a mechanically braked cycle ergometer to predict the physical working capacity at a heart rate of 170 beats·min^{-1}. Although instructions were provided for calibrating the Monark ergometer used in these trials, the energy loss because of friction in the chain and pedal bearings was ignored. Such friction is variable, averaging about 8% of the work performed (Cumming and Alexander, 1968). A more serious difficulty when applying the PWC$_{170}$ index to adults is that the task is relatively easy for a young individual, but may exceed maximum effort in an old person. An alternative approach is to set a power output that will generate an equivalent heart rate in all subjects (for example, 70% of the heart rate reserve) and to use a computer solution of the Åstrand nomogram (Shephard, 1970) to predict maximum oxygen intake. If the attained oxygen consumption is estimated from the work rate, this imposes a random error of at least 5% (Shephard et al., 1968).

In older individuals, the mechanical efficiency of operating a cycle ergometer may also depart from the 23% figure assumed in the nomogram calculations (Sidney and Shephard, 1977b). It is well-recognized that quadriceps weakness leads to underestimation of the directly measured maximum oxygen intake on a cycle ergometer (Shephard et al., 1968). Unfortunately, this same problem appears to increase the heart rate response to submaximum cycle ergometer effort in older individuals, yielding low predictions of maximum oxygen intake relative to other modes of exercise (Bailey et al., 1976). Increases of heart rate from anxiety or a high environmental temperature cause similar problems. A further problem in field application of the cycle ergometer is that it can only be moved with difficulty. Although it can be transported in a truck to a central testing station and recalibrated on site, it is not practicable to carry the apparatus to individual homes.

An alternative basis for a submaximal exercise test is the repeated climbing of a double step (Shephard et al., 1968). The concept of stair climbing is familiar to everyone, so that there is little change of mechanical efficiency with test learning and little spurious increase of heart rate because of anxiety. The work rate is given very simply as the product of step height × stepping rate × body mass, avoiding the need for any complicated calibration. The only precautions are to weigh the subject in the clothing used during testing; to ensure that subjects place both feet flat on the ground between ascents and that they stand erect on the top step; and to maintain the required stepping pace. Although it is possible for subjects to step with an incorrect rhythm (Bonen et al., 1977), in general the attained oxygen consumption conforms quite closely with the intended value, whether the pace be set by a metronome (Shephard, 1967) or by tape-recorded music, as in the Canadian Home Fitness Test (CHFT, Bailey et al., 1976); Jetté, 1983). The heart rate is best measured by electrocardiogram, although if the pulse is counted in the interval 5–15 seconds after a given exercise stage, it can be palpated without systematic error (Bailey et al., 1976). This provides an interesting contrast with cycle ergometry, where the heart rate decreases significantly within 5 seconds of halting a test (Yamaji and Shephard, 1985). Step test data can be converted to a predicted maximum oxygen intake, using the Åstrand nomogram; if the oxygen consumption is predicted from the rate of climbing, this increases the variability of data by about 7%. If the CHFT protocol has been followed, an alternative approach is to predict maximum oxygen intake from the attained stepping rate and heart rate, using a multiple regression formula proposed by Jetté et al. (1975):

$$\dot{V}O_2max = 42.5 + 16.6(E) - 0.12(M) - 0.12(f_h) - 0.24\,A.$$

where $\dot{V}O_2max$ is the maximum oxygen consumption (ml·kg^{-1}min^{-1}), E is the average cost of stepping (l·min^{-1} STPD), M is the body mass (kg), f_h is the heart rate 5–15 sec after exercise (beats·min^{-1}, and A is the age (years). The original equation was developed on quite a small sample of young adults, but subsequent trial has suggested this equation gives as valid answers as the Åstrand nomogram in a single survey, and a rather better estimate of the training response if a survey is repeated (Shephard and Cox, 1982; Jetté et al., 1982; Léger, 1984). A major part of the variance of data is attributable to the first two terms in the equation (in other words, the attained rate of working), in contrast with the Åstrand procedure (which relies entirely on the attained heart rate). For this reason, the Jettė equation is less vulnerable to factors

Table 6. The relationship between the attained stage of exercise on the Canadian Home Fitness Test and the maximum oxygen intake of the subjects.

	Undesirable (3-min stepping)		Minimum (6-min stepping)		Recommended (9-min stepping)	
Age	M	F	M	F	M	F
15–19 years	47	38	47	40	51	44
20–29 years	43	38	47	38	47	40
30–39 years	38	35	43	38	47	38
40–49 years	33	30	38	35	43	38
50–59 years	27	24	33	30	38	35
60–69 years	24	22	27	24	33	30

Note: All values ml·kg^{-1}min^{-1} STPD.

distorting the heart rate response to exercise, such as a high environmental temperature (Quinney et al., 1983) and anxiety; however, it may also lack the potential to predict a very high $\dot{V}O_2$max in an endurance athlete (Léger, 1984). An even simpler method of analysis of CHFT data is to record the attained exercise stage and relate this directly to maximum oxygen intake (table 6).

Jetté and his associates (1984) have proposed additional multiple regression equations for the prediction of maximum oxygen intake in children as follows:

Sex	*Age*	*Regression equation (years)*
M & F	7–10	$\dot{V}O_2\max = 0.49 + 1.64\,(\dot{V}O_2) - 0.015 \cdot (\Sigma S^1) - 0.0015\,(f_b)$
M	11–14	$= 1.56 + 1.14\,(\dot{V}O_2) - 0.012 \cdot (\Sigma S^1) - 1.0062\,(f_b)$
F	11–14	$= 0.053 + 1.149\,(\dot{V}O_2) - 0.0035 \cdot (\Sigma S^1) - 0.00084\,(f_b)$

where $\Sigma S^1 = S_b + S_t + S_s + S_i$ (see table 3), $\dot{V}O_2$ is the oxygen cost of the final exercise stage and f_b is the heart rate attained in the final stage of the Canadian Home Fitness Test. However, trial of these equations in the CFS has shown them to yield an unrealistically large jump of $\dot{V}O_2$max scores between age groups. In contrast, the Åstrand nomogram approach and the attained exercise stage approach both give reasonably smooth and satisfactory age-related curves for predicted maximum oxygen intake, except that the Åstrand predictions for young children seem too high (presumably because of difficulties in palpating the heart rate at this age) (table 7).

There has been much criticism of attempts to predict maximum oxygen intake from submaximal data (Rowell et al., 1964; Davies, 1968) on the grounds that the information content of the data is not increased by extrapolation, and that the error of predictions is too large to attach meaning to the calculations. The first of these arguments may be mathematically correct, but use of either the Jetté or the Åstrand nomogram allows some to visualize the interrelationship of age, heart rate, and work rate in a fashion that is not possible when using either the PWC_{170} or its homologue (the heart rate at an estimated oxygen consumption of 1.5 $l \cdot min^{-1}$) (Cotes, 1966). The second argument of inaccuracy also has some validity as far as the individual subject is concerned, but because the error of prediction is random rather than systematic (Shephard, 1977), population averages are correct under the conditions of a large-scale survey.

Muscle force

Some early surveys of muscle force used nonstandard dynamometers, so that leverage varied with the dimensions of the hand and the dynamometer. The widely adopted Stoelting apparatus allows the size of handgrip to be adjusted for body build, and the results obtained with this portable mechanical device do not differ significantly from those obtained by standard strain gauge techniques (Shephard et al., 1968b). Nevertheless, care is required with the zero setting, and periodic calibration is required against standard loadings. These points have not been emphasized in most of the surveys conducted to date.

There are differences of 10–50 N (2–10%) in the force developed by dominant and nondominant hands. Unfortunately, many surveys have overlooked this issue. The CFS originally presented data as the *sum* of forces for right and left hands, but it proved possible to approximate averages for dominant and nondominant hands by searching for the higher and lower values of each pair. The dominant value is often influenced by local occupational or athletic demands, and the nondominant reading may be more representative of general muscular development. Debate continues on the generality of handgrip information. Clarke (1966) claimed a moderately close correlation with overall muscular force, as determined by a series of tensiometer readings ($r = 0.79$); and except where there has been specific training of the wrist muscles, this still appears a reasonable conclusion.

Many types of tensiometer and dynamometer are too clumsy for transport to individual households. However,

Table 7. Measures of cardiovascular function for survey subjects

	Systemic blood pressure (mm Hg) [a]		Percent with recommended fitness [b]		Åstrand Predicted $\dot{V}O_2$max ($ml \cdot kg^{-1}\ min^{-1}$)		Jetté Predicted $\dot{V}O_2$max ($ml \cdot kg^{-1} min^{-1}$)	
Age	M	F	M	F	M	F	M	F
7–9 years	100/68	99/66	72.0	72.6	70.4	53.4	53.5	53.4
10–12 years	105/70	103/69	85.3	74.1	65.2	49.0	45.3	52.8
13–14 years	110/71	108/71	92.2	67.6	63.0	45.5	41.2	52.1
15–19 years	117/75	110/72	56.3	31.9	55.1	41.1	51.8	37.4
20–29 years	122/78	110/72	35.1	40.9	47.1	40.1	46.3	34.9
30–39 years	121/80	111/74	53.4	30.7	43.4	36.3	42.0	31.9
40–49 years	125/84	118/77	54.5	33.8	42.2	36.2	36.7	27.8
50–59 years	128/83	123/79	43.6	35.3	39.7	33.8	32.8	24.6
60–69 years	137/83	136/82	38.0	24.2	44.7	37.3	27.0	21.6

Source: Canada Fitness Survey, 1983.

[a] Resting systemic blood pressure, excluding subjects with triceps skinfold > 30 mm.

[b] Assuming those screened out or refused have low fitness.

if a car seatbelt and a cable tensiometer are attached to the stepping box, this allows some assessment of knee extension force in the field (Rode and Shephard, 1971). Repeated measures suggest that determinations of knee extension and back extension force have an extended learning curve (Shephard et al., 1977); and if such measures are to be introduced into a test battery, care must be taken that subjects are allowed sufficient practice to attain their maximum potential score.

Muscular endurance

In the home environment, it is difficult to find support bars for pullups or chinning, and the two most suitable simple tests of muscular endurance are bent-knee situps and pushups. There have been minor differences of technique for each of these procedures between U.S. and Canadian surveys, and these differences must be noted when comparing scores. In the CFS, the hands were clasped behind the head, the knees were bent to 90°, and the ankles were held by the observer. A count was kept of the number of situps completed within 60 seconds. The men performed pushups from the toes, in the conventional manner. However, it was doubted that older women could perform this type of maneuver; and, accordingly, they were allowed to carry out the test with the knees touching the ground. In both sexes, the reported score was the maximum number of pushups performed.

Muscular endurance tests depend greatly on recent experience and motivation. Unfortunately, most adults have not carried out any performance tests for many years. The CFS observers used "moderate" verbal encouragement, but this is obviously very difficult to standardize from one situation to another. A further factor, rarely considered in the interpretation of the data, is that muscular endurance is measured relative to body mass; this places the obese person at a substantial disadvantage. The handicap may be warranted when assessing aptitude for tasks where body mass must be displaced, but in much of the heavy physical work, where employment of women and the elderly is an issue, loads must be held 1–2 meters above the ground; in such postures, obesity is not a major handicap.

> The U.S. and the Canadian Armed Forces have recently been evaluating a semi-portable isotonic lifting machine, where subjects are required to make repeated lifts of a load that slides upwards in a steel frame. While the results have been interesting, to date they have shown no superiority over total mass or lean mass as predictors of performance "on the job." (Nottrodt and Celentano, 1984)

The CFS measured the maximum grip force of both hands. That of the dominant hand may be modified by local factors such as occupation or involvement in specific sports, but the figure for the nondominant hand offers a simple and reproducible estimate of general muscularity. Although there do not seem to have been any trials under field conditions, a measure of muscle endurance that was independent of body mass could be obtained by timing the grip at 50% of maximum force.

Flexibility

The CFS used a single widely accepted measure of "static" flexibility, the Dillon sit and reach test. One problem in the interpretation of scores from this apparatus is that the various surveys have used differing scale readings. The CFS assigned a score of 25 cm if a subject was able to touch the toe board with the tips of the fingers. Taller subjects, especially those with long legs, are probably at a disadvantage in the Dillon test, but there is no agreed basis of correcting scores for body build, nor is it known how far results can be generalized to the flexibility at other joints. In many types of sports, dynamic flexibility is more important than static flexibility, but for the average sedentary adult, the more important measure is probably "static" flexibility (where ample time is allowed to complete any given movement).

Other tests

The test battery used in the CFS is probably almost as lengthy as is desirable for a household visit. Other possible test items might include the following:

- *Perception of exertion.* An unfit subject rates a given bout of exercise as more exhausting than a person who is fit. Standard scales of perceived exertion are available (Borg, 1971), and it has been suggested that these scales may provide as accurate a measure of fitness as many submaximal tests of cardiorespiratory performance. There seems scope for linking this measure to the Canadian Home Fitness Test.
- *Perception of fitness.* Are subjects able to classify their personal fitness without completing a lengthy battery of tests? One early survey (Shephard and McClure, 1965) showed that, in service personnel, a simple activity/fitness rating accounted for 36% of the variation in exercise response. In a more recent population survey (Bailey et al., 1976), a 30% difference of CHFT scores was found between men who perceived themselves as fit and those who did not. However, in women, there was little relationship between fitness and CHFT score, suggesting that the women perceived fitness in terms of a slim figure and good posture rather than the development of aerobic fitness. Ratings could probably be improved by giving a more extended description of endurance fitness.
- *Anaerobic measurements.* Margaria (1966) developed a simple staircase sprint as a test of anaerobic power. This can be used for survey work if subjects report to a central location and timing lights can be installed on a convenient staircase. However, it is inappropriate both for a home survey and for an examination survey conducted in a mobile trailer.

Bar-Or (1984) has developed cycle ergometer tests of anaerobic power and capacity. If it is decided to carry a cycle ergometer to individual homes or to use one in a mobile trailer setting, the Bar-Or procedures could be used in a large-scale survey, although more study is needed of learning curves for this type test.

- *Lung volumes.* A variety of electronic spirometers now provide reasonably accurate field assessments of forced vital capacity, 1-second forced expiratory volume, and peak expiratory flow rate (Shephard and Cox, 1980). In most cases, the weight of the equipment is such that it must be left in a central test laboratory, although the McKesson-Scott Vitalor (Shephard, Thompson et al., 1958) has been used for domestic surveys. Measurement of lung volumes is particularly useful where testing is to be linked to a program of health education; it provides a useful entrée to a discussion of smoking habits and cigarette-related diseases.

Activity measurements

Activity questionnaires

The CFS used a relatively complicated three-page recall questionnaire to establish patterns of occupational, domestic, and leisure activities. One problem, perhaps more acute in Canada than in other nations, is a seasonal complexity of activity patterns. The intensity of effort in many occupations, such as farming, construction, and tourism, shows large seasonal extremes. Among vigorous domestic occupations, snow shovelling occurs irregularly for perhaps 3 months of the winter, and gardening and outside domestic repairs are limited to 3–4 months of summer. Snow conditions are good for cross-country skiing for perhaps 6 weeks in the winter, and lake conditions are comfortable for swimming during 8 weeks of summer; and in the fall, the number of outdoor hikers increases substantially. The CFS attempted to respond to this bewildering variety of patterns by recall questions based on the previous week, the previous month, and the previous year. However, there is a real danger that this pattern of questioning was confusing to the subjects, adding to the problems inherent in a long inventory of distant events. Although responses gave some guidance to those planning publicity and developing facilities, they were unsatisfactory from the fitness point of view, because the measures of intensity (light, medium, heavy, organized competitive) did not give a clear indication of vigor. Delphic inferences had to be drawn from the general description (for example, swimming), the frequency of participation, and the degree of organization (on the assumption that frequent participants in organized sports were more vigorous than occasional unorganized participants). Many questionnaires were returned only partially completed, and these required extensive editing. Algorithms also had to be developed for "days" that were too short or too long by several hours. Having applied these various corrections and adjustments, scores for daily energy expenditures bore some resemblance to anticipated figures, but the gradient of various fitness measures with reported vigorous activity was still quite slight (table 8). Given this experience, the present author's recommendation would be to revert to a much simpler activity questionnaire, possibly of the type recommended by Godin (1983). This explores the number of sessions per week of activity at three clearly specified intensities over the last 3 months (figure 1). A discriminant function score derived from this data is clearly linked to personal fitness (Godin, 1983).

Movement sensors

The physical impulses and accelerations resulting from physical activity can be measured by devices such as tilt meters, pedometers, and accelerometers that indicate movement in 1, 2, or 3 planes (Kemper and Verschuur, 1977; Saris and Binkhorst, 1977; McPartland et

Table 8. Influence of reported activity on some key variables

	Active						Moderate and sedentary					
Variable	Male			Female			Male			Female		
Age (years)	10–12	20–29	60–69	10–12	20–29	60–69	10–12	20–29	60–69	10–12	20–29	60–69
Body mass/height	0.270	0.417	0.455	0.272	0.352	0.416	0.264	0.423	0.456	0.270	0.354	0.40
Σ4 skinfolds (mm)	34.8	43.2	49.6	39.8	51.2	70.2	33.5	41.7	51.6	39.1	48.1	64.9
Percent body fat	18.4	16.2	18.2	23.9	25.4	29.9	17.8	15.7	18.5	23.8	24.8	28.8
Lean mass ($g \cdot cm^{-1}$)	0.219	0.346	0.371	0.205	0.260	0.287	0.215	0.354	0.370	0.204	0.264	0.28
Blood pressure (mmHg)	$\frac{104.0}{69.8}$	$\frac{121.5}{77.8}$	$\frac{135.3}{82.6}$	$\frac{103.1}{69.7}$	$\frac{110.1}{72.1}$	$\frac{139.3}{81.3}$	$\frac{104.4}{70.7}$	$\frac{121.7}{77.7}$	$\frac{138.3}{84.1}$	$\frac{103.0}{68.5}$	$\frac{110.2}{72.4}$	$\frac{133.7}{82.4}$
Predicted maximum oxygen intake ($ml \cdot kg^{-1} min^{-1}$)	—	45.6	26.9	—	34.3	21.3	—	46.7	27.1	—	35.2	21.7
Handgrip force (sum of two hands, N)	446	1041	839	395	596	503	433	1056	868	384	600	508
Pull ups (number)	11.4	22.2	7.5	22.0	16.3	6.9	13.8	26.0	10.4	19.2	18.2	9.9
Sit ups (number)	33.1	31.7	12.8	32.3	24.6	6.8	35.7	34.8	11.7	33.3	27.2	7.6
Sit and reach (cm)	26.0	29.8	20.7	30.8	31.7	24.8	26.3	30.4	21.4	31.1	33.4	28.9

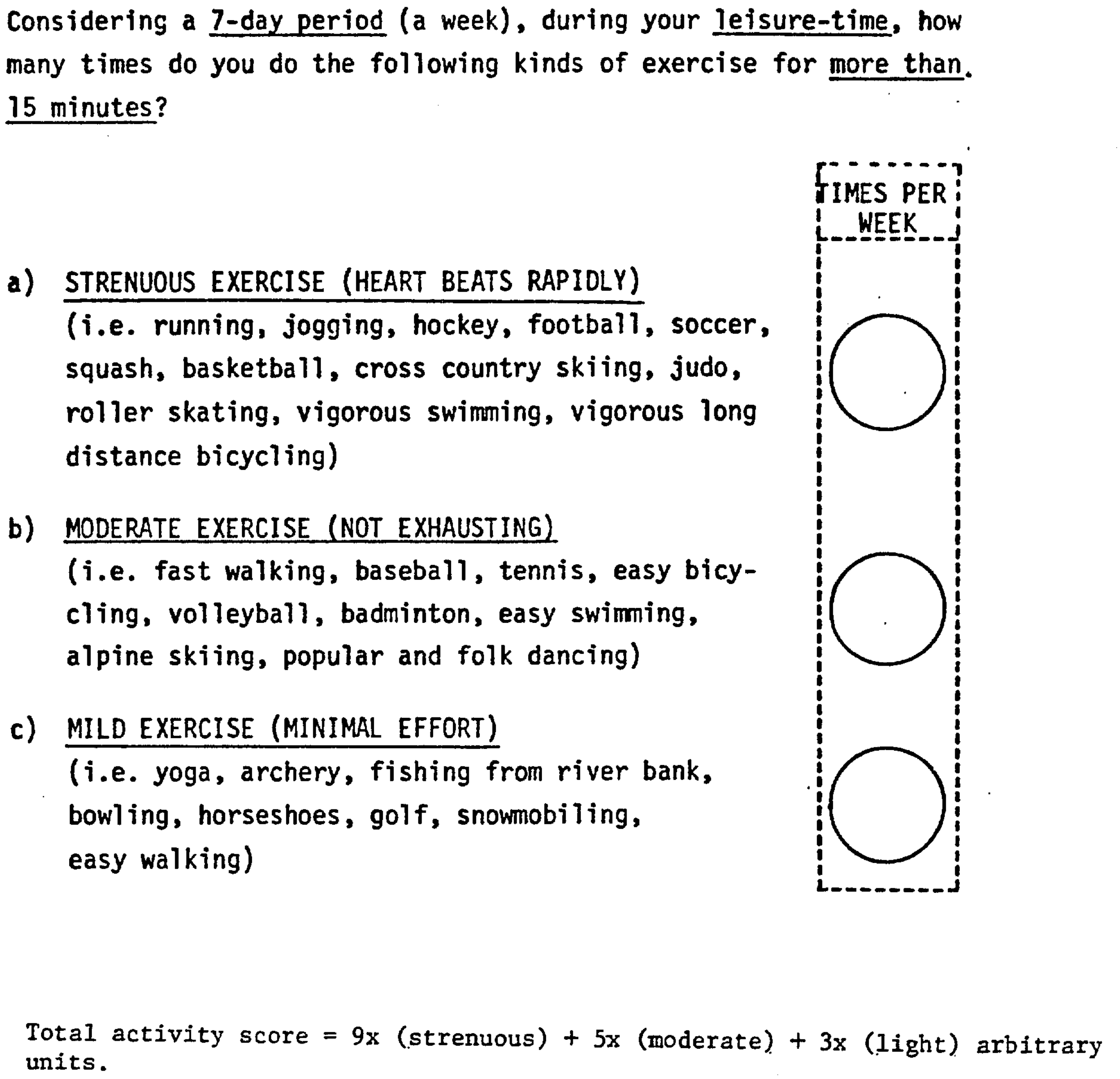
Note: Based on standard method of classifying active, moderately active, and sedentary individuals.
Source: (Canada Fitness Survey, 1983)

al., 1976; Montoye et al., 1983; Renfrew et al., 1984). This type of instrumentation is most effective when the daily activity follows a rather regular pattern (such as walking a postal carrier's route or jogging each evening). At best, the precision of results is rather limited, and the cost of such devices would preclude their use by more than a subsample of any large survey population.

Heart rate measurements

Heart rate may be recorded on the supposition that there is a linear relationship between heart rate and oxygen consumption. This assumption is theoretically weak, because the linear relationship holds only between 50% and 90% of maximum oxygen intake, and most daily activities involve an energy expenditure that is 10–50% of maximum (Shephard, 1967). However, activities that will induce a training effect (and are thus of interest to a fitness survey) do stimulate a heart rate that is 60–90% of maximum. Another problem of interpretation is that many factors other than endurance exercise can induce a tachycardia, including emotional stress; a high environmental temperature; and isometric muscle contraction. Moreover, the relationship between heart rate and oxygen consumption differs widely between individuals and between arm and leg work.

Recording devices are based on an electrocardiograph signal that is stored on a tape recorder for later playback (Holter, 1961; Shephard, 1967; Sidney and Shephard, 1977), or is integrated electrochemically (Wolff, 1966; Baker et al., 1967), mechanically (Glagov et al., 1970), or electronically (Saris et al. 1977). The main difficulty with tape recorders is to maintain a constant recording speed under field conditions; some of the more modern devices thus superimpose a 60 Hz signal, allowing a post-hoc correction of heart rates. It still remains necessary to inspect individual records for artifactual low or high counts caused by poor electrode contacts. Single-level integrators also give misleading



Considering a 7-day period (a week), during your leisure-time, how many times do you do the following kinds of exercise for more than 15 minutes?

TIMES PER WEEK

a) STRENUOUS EXERCISE (HEART BEATS RAPIDLY) ○
(i.e. running, jogging, hockey, football, soccer, squash, basketball, cross country skiing, judo, roller skating, vigorous swimming, vigorous long distance bicycling)

b) MODERATE EXERCISE (NOT EXHAUSTING) ○
(i.e. fast walking, baseball, tennis, easy bicycling, volleyball, badminton, easy swimming, alpine skiing, popular and folk dancing)

c) MILD EXERCISE (MINIMAL EFFORT) ○
(i.e. yoga, archery, fishing from river bank, bowling, horseshoes, golf, snowmobiling, easy walking)

Total activity score = 9x (strenuous) + 5x (moderate) + 3x (light) arbitrary units.

Figure 1. Simple activity questionnaire recommended by Godin (1983)

information because of such artifacts; but with multilevel integration, most of the "dropped counts" and high frequency noise can be discarded into the two extreme storage bins. The integrators used for heart rate counting are quite small, being worn on the belt or the wrist, and they are thus "socially acceptable" in the context of a survey. Their main drawback is again cost, which would probably restrict observations to quite a small subsample of the survey population.

Oxygen consumption measurements

Oxygen consumption measurements generally provide an accurate integral of energy expenditures over a specified interval, but they may flatten out peaks of activity that are important from the viewpoint of developing cardiorespiratory fitness. They are also relatively large and heavy (2–3 kg), and they require use of a facemask and harness or a mouthpiece and noseclip. Earlier devices provided a mechanical or an electrical integration of ventilation, with proportional sampling of gas for later chemical analysis (Kofranyi and Michaelis, 1949; Wolff, 1958); in consequence, each "experiment" yielded only a single estimate of energy consumption for a period of 10–66 minutes. A more recent development (the "Miser") (Humphrey and Wolff, 1976) allows a continuous measurement of ventilation and oxygen consumption. Because of cost factors, this type of equipment is again generally reserved for checking and amplifying information obtained from activity questionnaires.

Statistical design

Sample recruitment

When testing schoolchildren, it is a fairly simple matter to identify representative schools and to request the participation of randomly selected pupils from within those schools (CAHPER, 1965). However, care must be taken to identify students bypassing the normal State school system (private schools, schools for the handicapped, reformatories, etc.). If a simple field test of physical performance is required, the student volunteer rate may be close to 100%, although the obtaining of written parental consent for a series of laboratory tests reduces the proportion of volunteers to 90% in a small town (Shephard, Lavallee, et al., 1977) and to 33% in a major metropolis (Shephard et al., 1968b).

When dealing with large populations of adults, it is much more difficult to solicit recruits. An attempt to "synthesize" a representative sample on the basis of the reported occupations (Métivier and Orban, 1971) plainly overestimated the PWC_{170} of Canadian men. Occasionally, it may prove possible to test 60–70% of the adults in an isolated community (Shephard, 1978); but in an average office or factory setting, even a well-publicized survey organized on company time may attract no more than a third of employees (Cox, Shephard, and Corey, 1981). In the city of Saskatoon, with a population of about 120,000, volunteers for a laboratory fitness test were sought by telephone. We recognized that this approach excluded the institutionalized, those with unlisted telephones, and those too poor to rent a telephone line. About two-thirds of those contacted expressed a willingness to be examined, but as many as a half of these individuals failed to attend the laboratory at the scheduled hour. A further small proportion of subjects were excluded on medical grounds, so that of 2,648 contacts, only 778 began and 713 completed a simple step test (Bailey et al., 1976).

The CFS started from the vantage point that Canadian citizens are legally required to respond to questions posed by Statistics Canada. Nevertheless, Statistics Canada had no part in the data collection, and the CFS had a substantial "selling" task in recruiting volunteers. The normal approach in government surveys is to visit "representative households," which immediately bypasses service personnel, prisoners, hospital patients, and some types of collective households, such as dormitories. Further, because fiscal limitations restrict the number of repeat visits, those who travel frequently or work shifts are underrepresented. Because of the vast geographic area of Canada and of zones with a very low population density, it was not possible to pick CFS households entirely at random. Eighty clusters of households were selected, representing the square root of population in the various Provinces, with roughly equal sampling of large cities, smaller towns, and rural areas. Fortunately, only minor and statistically insignificant differences of fitness were encountered with area of residence and with the season of testing (February to April versus May to July). A total of 11,884 out of 13,440 households provided information on 30,652 individuals over the age of 7 years; 15,519 (or 50.6%) of the occupants of cooperating households agreed to undertake the required fitness tests, but many of those concerned were children. Only about 1 out of 3 of potential adult participants were tested.

Sampling bias

Incomplete sampling of a population inevitably introduces some bias into data. Because complete sampling is impossible in a free society, it becomes important to ascertain the nature of this bias and the extent of its impact on estimates of population fitness. Where observations are repeated at infrequent intervals, the sampling bias may remain relatively consistent, allowing inferences to be drawn about changes in the condition of that segment of the population that agrees to testing.

If the required procedures involve measurements of fitness, even a small deficit in sampling can substantially augment apparent fitness. In one survey (Shephard, 1978), measurements of vital capacity were made by a team of exercise physiologists who examined some 62% of the younger and 44% of the older adult population. Similar data were collected by a medical team, who saw 61% of the adult population. There was no significant interobserver error on subjects common to the two surveys; but because the exercise physiologists attracted an above-average proportion of fit individuals and the physicians tended to test the sick, the mean values for the men and women of the community reported by the two teams of investigators were 3.76 and 2.74 versus 3.43 and 1.90 l BTPS, respectively.

The proportion of the population that has either acute or chronic disease is relatively low in young adults, but becomes an increasing source of difficulty in surveys of older adults (Brown and Shephard, 1967).

There are two further disease-related issues. First, "healthy" and "diseased" subjects should preferably be distinguished and considered separately; otherwise, it will be unclear whether secular trends of performance for a community reflect a change of fitness or a change in the proportion of people who are healthy. Second, caution may cause paramedical teams to exaggerate the volunteer bias, because those rejected on medical grounds exceed the number of individuals who have clinical disease (Canada Fitness Survey, 1983).

Fitness testing tends further to attract health-conscious individuals of high socioeconomic status. Samples are selectively deficient in such populations as recent immigrants, visible minorities, the unemployed, and single mothers. Many of those who are underrepresented unfortunately have a below average interest in fitness and health. Specifically, few smokers volunteer for fitness testing (Massie and Shephard, 1971; Sidney and Shephard, 1977), yet smokers often have a poor score on measures of aerobic fitness (Glassford and Howell, 1969). Population fitness levels are thus overestimated, and a misleading impression of improving fitness may be formed if the proportion of smokers in a community is decreasing.

The CFS did not attempt to analyse the fitness status of nonrespondents, but it should be possible to persuade at least some of this group to indicate their body mass and activity habits.

Sample size

The sample size needed in a survey depends on the precision of the test procedures, the magnitude of differences between fit and unfit individuals, and the need or the wish to stratify the data. The CFS wished to examine the fitness of children in some detail, and thus deliberately augmented the proportion of those under 20 years of age. It also wished to analyse regional differences, and thus recruited subjects in proportion to the square root of local population in order to ensure the testing of adequate numbers from the less densely populated regions of Canada.

The error of measurement is known fairly precisely for some physiological tests such as vital capacity when determinations are made in a clinical laboratory. If measurements are made in the home using simpler equipment and less experienced personnel, the experimental error is likely to be at least 10–20%. To this must be added true, constitutionally determined variations of score within a population attaining a given fitness category. Again, this variation is unlikely to be less than 20%. Because the two sources of variance are largely independent of one another, the combined variance may amount to $(20^2 + 20^2)^{1/2}$, or about 28%. The accuracy of the population estimate is $N^{1/2}$ of this figure. Thus a sample of 200 yields a precision of about 2%. This is adequate relative to the anticipated gradient of response from fit to unfit (commonly of the order of 20% of average scores).

Although the sample size used in the NHANES is considerably larger than this figure, much of the available sampling margin will be taken up with a necessary and interesting stratification of the data. For example, consideration should be given to effects of age and sex (at least 10 cells) and their interaction with socioeconomic status.

Data transformation

Much of the data relating to fitness is not normally distributed (Jéquier et al., 1977). For example, the usual population sample shows a substantial "tail" of obese individuals, leading to significant skewing of figures for body mass (table 9). The implication is that mean values for the population are biased upwards by very high scores in unfit individuals. Potential options are to normalize the data through some type of logarithmic transformation (a useful approach where statistical techniques are to be applied) or to present alternative estimates of central tendency (the median, mode, or selected percentiles). CAHPER surveys of physical performance (1965, 1980) and physical working capacity (Howell and MacNab, 1968; Métivier and Orban, 1971; Gauthier et al., 1983) provide examples of results that have been presented as age- and sex-specific percentiles.

The choice of option in data reporting is influenced by the purpose of a survey. If the objective is to make a comparison with other large studies, then it may be necessary to present information in the format adopted by earlier investigators. For some health-related interpretations, it may also be important to know the percentage of an age- and sex-specific sample that exceeds an arbitrary limit—for example, that proportion of the population that is more than 10 kg above the "ideal" mass for a person of comparable body build. Although there is a fair amount of information on the health consequences of a specified excess of body mass, much

Table 9. Skewing of body mass (kg)

	Male		Female	
Age	Mean	Median	Mean	Median
6 years	21.1	20.5	19.3	19.9
8 years	25.2	24.6	24.6	24.0
10 years	30.8	29.4	30.4	29.2
12 years	37.4	35.6	39.0	37.8

Source: Based on unpublished data of Shephard and Lavallée for the Trois Rivières region.

less is known about any harmful consequences of a 10% loss of aerobic power or flexibility.

Human subject considerations

Safety

Mass testing of children originally used a 300-yard run (CAHPER, 1965) or a 600-yard walk-run (AAHPER, 1958) because it was thought dangerous to ask children to make an all-out effort over a longer distance. It has now been agreed that such fears were groundless, and the most recent AAHPERD (1980) protocol includes a 1½ mile run for students aged 13 years and over.

Similar but more strongly grounded fears have restricted the fitness testing of the adult population. Although Cooper (1968) originally advocated the testing of aerobic fitness by an all-out 12-minute run, his subsequent writing has become progressively more cautious. Certainly, it is unwise to suggest 12 minutes of all-out running without a warmup to an older person who has been sedentary for many years and who has not had a recent medical examination.

The CHFT (Bailey et al., 1976) was originally assailed on grounds of safety. However, the designers of the CHFT had given considerable thought to this issue. A preliminary screening out of high-risk individuals was assured by means of a simple, self-administered "Physical Activity Readiness" (PAR–Q) questionnaire (Chisholm et al., 1975). A warmup was introduced by a preliminary 3-minute bout of stepping at a rate appropriate for a person 10 years older than the test candidate. Bouts of exercise were brief (3-minute duration), conservative, and submaximal <70% of maximum oxygen intake). Clear pulse ceilings were also set for test abortion. It is difficult even to estimate the precise safety of a procedure where no emergencies have yet occurred. Nevertheless, it can be said that the CHFT has now been performed by over a half million people, without serious problems. There have been a few transient episodes of loss of consciousness immediately following the test, and one subject sustained a mild ankle strain. If care is not taken to place the stepping bench against a smooth wall, injury from stumbling is also conceivable. Theoretical calculations suggest that all 10 million Canadian adults could be tested every year for 33 years before a cardiac emergency would be attributable to the CHFT (Shephard, 1976). Moreover, if only a half of vulnerable individuals had been screened out by the PAR–Q questionnaire, 66 years would pass without a critical incident (Shephard, 1976). There is still scope to improve the design of preliminary screening questionnaires (Shephard et al., 1984), but cardiac problems are not a serious issue when planning the assessment of 20,000 individuals by a test of the CHFT type.

The other tests used in the CFS are also very safe, although repeated pushups could cause an unwelcome rise of systemic blood pressure in an older, hypertensive individual, while straight-leg (as opposed to bent-knee) situps could, at least theoretically, predispose to muscular skeletal problems from hyperextension of the spine.

Informed consent

U.S. investigators usually follow ACSM (1975) guidelines in obtaining a signed and witnessed statement of informed consent before exercise testing, even if the data are being used for clinical rather than experimental purposes. CFS also used a witnessed consent form. However, it is not particularly profitable to make detailed international comparisons of the consent process because of peculiarities of the U.S. legal system.

The United States has an unwelcome record of successful litigation against investigators, apparently resulting in part from the practice of allowing lawyers to claim a percentage of any damages that are awarded. A charge of assault may be successful if it can be demonstrated that all potentially harmful consequences of test participation were not disclosed in a manner that could be understood by the subject. Further, charges of negligence could be brought if, for example, an investigator left a sharp object within stumbling range of a person performing a step test. No statement of waiver of rights would have legal force if it were shown that investigators had failed to take reasonable precautions, consistent with their level of knowledge and experience.

Organization and administration

Management

Many large-scale scientific surveys have failed or at least have achieved less than their potential because of poor management. Project leaders are usually chosen for their scientific reputation, but this is poor preparation for such tasks as efficient budgeting and supervision of a multimillion dollar project. In the context of the NHANES, there is need for a manager who can recruit and train perhaps 80 teams of investigators, inspiring them with enthusiasm to recruit at least 20,000 subjects, to carry out meticulous routine tests on the entire sample and to record the data accurately. Clear plans must be drawn up at an early stage for developing a readily accessible data base, with an early printout of key information and well-defined channels for requests from potential secondary users of the results.

Two patterns of organization have been adopted in Canada. One has brought together ad hoc groups of scientists with supposed common interests who have selected a leader and submitted a multiapplicant grant proposal. Typically, this approach has had only patchy success. Large and important sectors of data have then remained uncollected or unanalyzed, and much of the money allocated has been diverted to side projects of interest only to individual investigators. The alternative has been to allocate a fixed budget to the research committee of a national organization, such as the Canadian Association of Sports Sciences or the Canadian

Association for Health, Physical Education, and Recreation. This has generally led to a better managed but scientifically less brilliant project. Moreover, the national organization exists in perpetuity to account for its stewardship of funding; and, in consequence, dull and repetitive work needed for attainment of survey objectives has been completed on schedule. The Canadian Fitness Survey adopted a third option, being established at an arm's length as "Crown Corporation," to which funds could be awarded. Career civil servants were seconded to the corporation for specific terms of 1–2 years, and consultants interested in specific aspects of the CFS project were also recruited from the academic community on an ad hoc basis. Test administration was organized and supervised by CFS staff, although the procedures to be used had previously been established in detail by a consultant committee of the Canada Fitness Survey; and technical advice was sought from academic exercise physiologists and physical educators at various points in training, testing, and data analysis.

Organizational details

We have noted already that the CFS employed 80 teams of two people to contact a total of some 30,000 people over a 6-month period. Although the Canadian data showed no striking differences between winter and summer seasons or between urban and rural areas, it would be desirable for NHANES to check these points in the United States, and also to examine potential differences of score between the major "visible minorities."

Quality control of data is important, but it does not appear to have been addressed formally in most large fitness surveys, other than through the use of a standard protocol and centralized initial training of observers. The simplest and probably the most effective approach to control interinvestigator measurement bias is a system of biological calibration, where subjects of known characteristics are tested periodically by each of the observer teams (Jones and Kane, 1979). Occasional random errors (for example, misplacement of a decimal or recording of height in inches rather than centimeters) can also create substantial bias, particularly in smaller subsets of the data. It is thus important to establish a window of acceptable upper and lower readings for each variable (for example ±4 SD about the mean value). Results exceeding these limits should be flagged for verification, and unless a reasonable explanation can be found, these figures should not be admitted to the data base.

Secondary users may wish to explore relationships to other large bodies of data (for example, health care costs, Shephard et al., 1982). Individual cases should thus be identified by several pieces of information (for example, year of birth, sex, postal code, and house number) that can be used in building unequivocal cross-linkages. In Canada, medical coverage is provided on a family basis; and without the additional linkage of sex, several male subjects in one such survey would have been attributed medical costs for antenatal and postnatal care (Shephard et al., 1982).

Procedures for data analysis and interpretation

Cross-sectional versus longitudinal analysis

Cross-sectional analysis may make comparisons with other surveys, in which case the similarity of methodology becomes a crucial and often an insoluble dilemma. Alternatively, comparisons may be drawn within the current data set. The latter appears a useful method of analysing results. For example, the linkage of body mass with activity algorithms allows an exploration of the question, What combination of frequency and intensity of activity keeps body mass within 10 kg of the ideal value? The obvious weakness of intrasample analysis is that one cannot infer causality or directionality. Inactive people may remain within the 10–kg limit not because they are inactive but because they are also heavy smokers. Likewise, if a low level of activity is associated with a substantial excess weight, it may be that excess body weight discourages physical activity rather than the converse.

Longitudinal analysis presupposes budget and enthusiasm for a repetition of observations after the lapse of 5–20 years. If there has been a single and substantial change of environment over the intervening period, both directionality and causality may be inferred. Thus in the Canadian arctic, the indigenous population has undergone a rapid acculturation to a sedentary "Western" lifestyle over the past 10 years, with an associated dramatic drop of fitness (Rode and Shephard, 1984). In the NHANES context, one might hope to demonstrate a substantial trend in the reverse direction through some program initiative such as the targeting of fitness propaganda on a particular community or a particular segment of the population. Because changes of health behavior develop relatively slowly, it would be necessary to wait 5–10 years before seeking a change detectable by survey methodology.

Important practical problems in the conduct of longitudinal surveys include a lack of early publications that can be used by principal investigators for purposes of career advancement; the dating of survey equipment and methodology (with pressures for the change of techniques, which must be resisted); and changes of personnel (with a potential for the development of interobserver errors). Finally, and regrettably, it is difficult to sustain the enthusiasm of government for long-term projects when political dividends can be reaped from much smaller short-term expenditures.

Issues of proportionality

Whether cross-sectional or longitudinal methodology is adopted, comparisons often need to be drawn between groups differing in body dimensions. A theo-

Table 10. Exponents of proportionality for fitness measures—a comparison of theoretical and observed values

Variable	Observed height exponent	Theoretical height exponent
Maximum oxygen intake	2.78	2.0
PWC_{170}	3.37	2.0
Vital capacity	2.68	3.0
Handgrip force	3.24	2.0
Back extension force	2.72	2.0
Leg extension force	2.88	2.0
Situps (number)	1.97	?
50-yard run (sec)	−0.87	0
300-yard run (sec)	−0.81	0

Source: (Shephard et al., 1980)

retical basis for the adjustment of morphological variables has been proposed (Von Döbeln, 1966; table 10), but in practice exponents relating fitness to stature often differ substantially from the anthropometric hypothesis (table 10). In some nations such as Japan, secular trends have increased stature by as much as 2% over a decade, and this in itself could give rise to a 6% increase of absolute maximum oxygen intake (although the relative value, measured in $ml \cdot kg^{-1} min^{-1}$, would remain largely unchanged).

Other issues

The problem of constitutional variation has already been noted. This makes it extremely difficult to interpret fitness scores on any one individual. A person may have low aerobic power because of laziness or a poor constitutional endowment. This is a less serious problem in large surveys, because most investigations suggest there is little constitutional difference of fitness *between* populations. Some basis for the interpretation of individual fitness scores becomes possible if results are linked to measurements of habitual activity. This seems a strong argument for including activity data in the NHANES.

Fitness may also reach less than its potential level because of disease, malnutrition, or adverse aspects of lifestyle such as cigarette smoking. For completeness of interpretation, it is thus desirable to collect at least a minimum of information on health, nutrition, and lifestyle.

Recommendations

Survey objectives

a. Objectives of the NHANES should be listed in order of priority.
b. Choice of test procedures should reflect:
   Survey objectives
   Current definitions of fitness
c. A broad, balanced battery of fitness tests, suitable for home administration, should be linked to a simple but reliable and carefully validated index of habitual activity.

Test methodology

a. All equipment must be rugged, portable, and suitable for use by paramedical health professionals.
b. Proposed procedures must be checked for reliability and validity when used under field conditions.
c. A set square, tape measure and bathroom scales can provide adequate population estimates of height and body mass, although fasting mass will be 1–2 kg lower than the survey estimate.
d. Performance tests are very vulnerable to conditions of measurement, and also reflect recent gymnastic experience. Most items are thus not worth including in a fitness survey.
e. Accurate skinfold measurements require interobserver standardization of techniques. If the currently popular quadratic formula based on abdominal, triceps, and subscapular folds is adopted, biceps and suprailiac folds should also be measured to allow comparisons with European and Canadian data. It is useful to calculate percent body fat and lean mass, but skinfold readings and body density estimates should also be reported.
f. Girths do not add materially to survey estimates of body composition, and moderately trained observers make better estimates of body fatness with calipers that with tape measures.
g. Blood pressure readings are a useful component of a survey:
   To avoid excessive exclusion of supposed hypertensives
   To assess community prevalence of hypertension.
   However, conditions of measurement need clear specification.
h. Aerobic power can be assessed in the field using a submaximal cycle ergometer or step test. The latter is more readily portable and requires no elaborate calibration.
i. Measurement of maximum grip force exerted by the nondominant hand probably offers the best simple field test of general muscularity. If a measure of endurance is desired, the timed grip at 50% of maximal force could be considered.
j. Addition of other tests will be limited by the available time per household. Potential items to be considered included:
   Rating of perceived exertion (linked to the Canadian Home Fitness Test)
   Rating of perceived fitness (with careful description of the type of fitness that is sought)
   Staircase or ergometer measures of anaerobic power and capacity
   Standard lung volumes
k. The main basis of habitual activity measurement will need to be a simple, reliable and carefully validated questionnaire. Other devices such as movement sensors, heart rate recorders and oxygen consumption recorders are too expensive and complicated for

effective use on more than a small subsample of the population.

Statistical design

a. With a test battery of the order described, selective recruitment of volunteers is almost inevitable. Careful note should be kept of the proportions sampled at various ages, and an attempt made to determine bias by recollection of body mass and activity habits
b. The sample size of 20,000 is necessary to allow for stratification of data by age, sex, region, ethnic, and socioeconomic grouping
c. The CFS data showed insignificant differences of fitness scores between February–April and May–July. Nevertheless, NHANES should test in both summer and winter seasons.
d. Although the skewing of some fitness data suggests a need for logarithmic transformation or consideration of percentiles, mean values are also needed to compare with several earlier surveys.

Human subject considerations

a. No major health complications may be anticipated from the testing of 20,000 subjects, given a simple screening by a physical activity readiness questionnaire.
b. The American College of Sports Medicine recommendations on informed consent should be followed in recruiting test volunteers.

Organization and administration

a. The key to a successful survey will be expert management by a professional with experience in research administration.
b. Cluster sampling by eighty teams of one man and one woman might provide an effective approach to testing 20,000 people at home over a period of 3 months.
c. Quality control of the data must be exercised both in collection and in entry to the data base.

Procedures for analysis and interpretation

a. Firm commitment to a longitudinal study is important to the attribution of directionality and causality of relationship between fitness and health.
b. Measures of proportionality should be introduced if the average size of the population changes over the course of the survey.
c. Activity measurements should be used to distinguish clearly between limitations of fitness imposed by constitution, poor health and nutrition from those resulting from a lack of physical activity.

References

AAHPER. *Youth Fitness Test Manual*. Washington, D.C. AAHPER, 1958.

AAHPER. *Youth Fitness Test Manual*. Revised 1976 Edition. Washington, D.C. AAHPER, 1976.

AAHPERD. *Lifetime Health Related Physical Fitness Test Manual*. Washington, D.C. AAHPERD, 1980.

American College of Sports Medicine. *Guidelines for Graded Exercise Testing*. Philadelphia: Lea and Febiger, 1975.

Astrand, I. Aerobic work capacity in men and women with special reference to age. *Acta Physiol. Scand.* 49, Suppl. 169, 1–92, 1960.

Bailey, D.A., Shephard, R.J., and Mirwald, R.L. Validation of a self-administered home test of cardio-respiratory fitness. *Can. J. Appl. Spts. Sci.* 1, 67–78, 1976.

Baker, J.A.S., Humphrey, J.E., and Wolff, H.S. Socially acceptable monitoring instrument (SAMI). *J. Physiol. (Lond.)*. 188, 4P–5P, 1967.

Bar-Or, O. *Pediatric Sports Medicine for the Practitioner*. New York: Springer Verlag, 1984.

Bonen, A., Gardner, J., Primrose, J., Quigley, R., and Smith, D. An evaluation of the Canadian Home Fitness Test. *Can. J. Appl. Spts. Sci.* 2, 133–136, 1977.

Borg, G. The perception of physical performance, pp. 280–294. In: *Frontiers of Fitness*. Ed: R.J. Shephard. Springfield, Ill.: C.C. Thomas, 1971.

Brown, J.R., and Shephard, R.J. Some measurements of fitness in older female employees of a Toronto department store. *Canad. Med. Assoc. J.* 97, 1208–1213, 1967.

Burgert, S.L., and Anderson, C.F. A comparison of triceps skinfold values as measured by the plastic McGaw caliper and the Lange caliper. *Amer. J. Clin. Nutr.* 32, 1531–1533, 1979.

CAHPER. *The CAHPER Fitness Performance Test Manual: For Boys and Girls 7 to 17 Years of Age*. Ed: Hayden, F.J., and Yuhasz, M.S. Ottawa, Ont. CAHPER, 1965.

CAHPER. *Fitness Performance II Test Manual*. Ed: Gauthier, R., Quinney, H.A., Massicotte, D., and Conger, P. Ottawa: CAHPER, 1980.

Canada Fitness Survey. *Fitness and Lifestyle in Canada*. Ottawa: Fitness and Amateur Sport, 1983.

Chisholm, D.M., Collis, M.L., Kulak, L.L., Davenport, W., and Gruber, N. Physical activity readiness. *Brit. Col. Med. J.* 17, 375–378, 1975.

Clarke, H.H. *Muscular Strength and Endurance in Man*. Englewood Cliffs, N.J. Prentice Hall, 1966.

Colburn, T.R., Smith, B.M., Guarini, J.J., and Simmons, N.W. An ambulatory activity monitor with solid state memory. ISA BM 76322 pp. 117–122, 1976.

Cooper, K.H. *Aerobics*. New York: Evan, 1968.

Cotes, J.E. Occupational Safety and Health Series, Rept. 6. Geneva: I.L.O., 1966.

Cotes, J.E., and Woolmer, R.F. A comparison between twenty seven laboratories of the results of analysis of an expired gas sample. *J. Physiol.* 163, 36p–37p, 1962.

Cox, M., Shephard, R.J., and Corey, P. Influence of an employee fitness programme upon fitness, productivity and absenteeism. *Ergonomics* 24, 795–806, 1981.

Cumming, G.R., and Keynes, R. A fitness performance test for school children and its correlation with physical working capacity and maximal oxygen uptake. *Canad. Med. Ass. J.* 96, 1262–9, 1967.

Cumming, G.R., and Alexander, W.D. The calibration of bicycle ergometers. *Can. J. Physiol. Pharm.* 46, 917–919, 1968.

Davies, C.T.M. Limitations to the prediction of maximum oxygen intake from cardiac frequency measurements. *J. Appl. Physiol.* 24, 700–706, 1968.

Drake, V., Jones, G., Brown, J.R., and Shephard, R.J. Fitness performance tests and their relationship to the maximal oxygen uptake of adults. *Canad. Med. Assoc. J.* 99, 844–848, 1968.

Drake, V., White, D., and Shephard, R.J. The fitness performance of Canadian working men. With some comments on adaptation of performance tests to a small gymnasium. *J. Spts. Med. Phys. Fitness* 9, 152–161, 1969.

Durnin, J.V.G.A., and Rahaman, M.M. The assessment of the amount of fat in the human body from measurements of skinfold thickness. *Brit. J. Nutr.* 21, 681–689, 1971.

Durnin, J.V.G.A., and Womersley, J. Body fat assessed from total body density and its estimation from skinfold thickness: Measurements on 481 men and women aged from 16 to 72 years. *Br. J. Nutr.* 32, 77–79, 1974.

Frost, J.D. Triaxial vector accelerometry: A method for quantifying tremor and ataxia. *IEEE Trans. BME* 25, 17–27, 1978.

Gauthier, R., Massicotte, D., Hermiston, R., and MacNab, R. The physical work capacity of Canadian children, aged 7 to 17, in 1983. A comparison with 1968. *CAHPER Journal* 50, 4–9, 1983.

Glagov, S., Rowley, D.A., Cramer, D.B., and Page R.G. Heart rates during 24 hours of usual activity for 100 normal men. *J. Appl. Physiol.* 29, 799–805, 1970.

Glassford, R.G., and Howell, M.L. Smoking and physical fitness: A preliminary report. *Can. Fam. Physician* Oct. 1969.

Godin, G. Psychosocial factors influencing intentions to exercise of young students. University of Toronto: Ph.D. Thesis, 1983.

Godin, G., and Shephard, R.J. A simple method to assess exercise behaviour in the community. Paper presented to Society of Prospective Medicine. Quebec City, Oct. 1982.

Holter, N.J. New method for heart studies. *Sciences* 134, 1214–1220, 1961.

Housh, T.J., Thorland, W.G., and Johnson, G.O. An evaluation of intertester variability in anthropometry and body composition assessment. *J. Sports Med.* 23, 311–314, 1983.

Howell, M.L., and MacNab, R.G.J. *The Physical Work Capacity of Canadian Children Aged 7 to 17.* Toronto: CAHPER, 1968.

Humphrey, S.J.E., and Wolff, H.S. The oxylog. *J. Physiol.* 267, 59p–60p, 1976.

Jackson, A.S., Pollock, M.L., and Gettman, L.R. Intertester reliability of selected skinfold and circumference measurements and percent fat estimates *Res. Quart.* 49, 546–551, 1978.

Jamison, P.L., and Zegura, S.L. A univariate and multivariate examination of measurements of the same subjects in two different institutes. *Amer. J. Phys. Anthrop.* 40, 341–343, 1974.

Jéquier, J.C., LaBarre, R., Rajic, M., Beaucage, C., Shephard, R.J., and Lavallée, H. Le postulat de normalité dans les études longitudinales, pp. 55–66. In: *Frontiers of Activity and Child Health.* Ed. H. Lavalleé and R.J. Shephard. Québec City: Editions du Pélican, 1977.

Jetté, M. The energy requirements of the Canadian Home Fitness Test and their application to the evaluation of work performance. *Can. J. Publ. Hlth.* 74, 401–403, 1983.

Jetté, M., Thoden, J.S., and Gauthier, R. Aerobic exercise prescription intensity in terms of maximal working capacity. *Can. J. Publ. Hlth.* 66, 465–467, 1975.

Jetté, M., Mongeon, J., and Shephard, R.J. Demonstration of a training response by the Canadian Home Fitness Test. *Eur. J. Appl. Physiol.* 49, 143–150, 1982.

Jones, N.L., and Kane, M. Inter-laboratory standardization of methodology. *Med. Sci. Spts.* 11, 368–372, 1979.

Katch, F.I., and McArdle, W.D. An anthropometric estimation of body density and lean body weight in young men. *J. Appl. Physiol.* 27, 25–31, 1969.

Kemper, H.C.G., and Pieters, J.J.L. Comparative study of anthropometric measurements of the same subjects in two different institutes. *Amer. J. Phys. Anthrop.* 40, 341–343, 1974.

Kemper, H.C.G., and Verschuur, R. Validity and reliability of pedometers in habitual activity research. *Eur. J. Appl. Physiol.* 37, 71–78, 1977.

Kofranyi, E., and Michaelis, H.F. Ein Tragbarer Apparat zur Bestimmung des Gasstoffwechsels. *Arbeitsphysiologie* 11, 148–150, 1949.

Kreitler, H., and Kreitler, S. Movement and aging. A psychological approach. In: *Medicine and Sport.* Vol. 4. Physical activity and aging. Ed. D. Brunner and E. Joke. pp. 302–6. Basel: Karger, 1970.

LaPorte, R.E., Kullner, L.H., Kupfer, D.J., McPartland, R.J., Matthews, G., and Caspersen, C. An objective measure of physical activity for epidemiologic research. *Am. J. Epidemiol.* 109, 158–167, 1979.

Leger, L. The Canadian Home Fitness Test revisited. *J.C.A.H.P.E.R.* In Press, 1984.

Leger, L.A., Lambert, J., and Martin, P. Validity of plastic skinfold caliper measurements. *Human Biol.* 54, 667–675, 1982.

Lohman, T.G. Skinfolds and body density and their relation to body fatness: A review. *Human Biol.* 53, 181–225, 1981.

Margaria, R. An outline for setting significant tests of muscular performance. In: *Human Adaptability and its Methodology.* Ed. H. Yoshimura and J.S. Weiner. Tokyo: Society for the Promotion of Sciences. 1966.

Massie, R.B. and Shephard, R.J. Physiological and psychological effects of training. *Med. Sci. Spts.* 3, 110–117, 1971.

Mazess, R.B., and Mathisen, R.W. Lack of unusual longevity in Vileabamba, Ecuador. *Human Biol.* 54, 517–524, 1982.

McPartland, R.J., Kupfer, D.J., and Foster, F.G. The movement-activated recording monitor: A third-generation motor-activity monitoring system. *Behav. Res. Meth. Instr.* 8, 357–360, 1976.

Métivier, G., and Orban, W.A.R. *The Physical Fitness Performance and Work Capacity of Canadian Adults Aged 18 to 44 Years.* Ottawa: CAHPER, 1971.

Montoye, H.J., Washburn, R., Sbrvais, S., Ertl, A., Webster, J.G., and Nagle, F.J. Estimation of energy expenditure by a portable accelerometer. *Med. Sci. Sports Ex.* 15, 403–407, 1983.

Murray, S.J., and Shephard, R.J. Relationship between neck and waist girths, skinfold readings and hydrostatic estimates of body fat in young women. *Can. J. Appl. Spts. Sci.* In Press, 1984.

Nottrodt, J.W., and Celentano, E.J. Use of validity measures in the selection of physical screening tests. pp. 433–6. In: *Proceedings of the 1984 International Conference on Occupational Ergonomics.* Ed. R.D.G. Webb. Toronto, Ont. Human Factors Society of Canada, 1984.

Orban, W.A.R. *Proceedings of National Conference on Fitness and Health.* Ottawa: Health and Welfare, Canada, 1974.

Pãrízková, J. Total body fat and skinfold thickness in children. *Metabolism* 10, 794–807, 1961.

Pãrízková, J. *Body Fat and Physical Fitness.* The Hague: Martinus Nijhoff, 1977.

Pascale, L.R., Grossman, M.I., Sloane, H.S., and Frankel, T. Correlations between thickness of skinfolds and body density in 88 soldiers. *Hum. Biol.* 28, 165–176, 1956.

Pollock, M.L., Laughridge, E.E., Coleman, B., Linnerud, A.C., and Jackson, A. Prediction of body density in young and middle-aged women. *J. Appl. Physiol.* 38, 745–749, 1975.

Pollock, M.L., Hickman, T., Kendrick, I., Jackson, A., Linnerud, A.C., and Dawson, G. Prediction of body density in young and middle aged men. *J. Appl. Physiol.* 40, 300–304, 1976.

Quinney, H.A., Cottle, W.H., MacDougall, A., and Baleshta, J. The effect of temperature variation on cardiovascular response to the STF step test. *Can. J. Appl. Spts. Sci.* 8, 205, 1983.

Renfrew, J.W., Moore, A.M., Grady, C., Robertson-Tchabo, E.A., Cutler, N.R., Rapaport, S.I., Colburn, T.R., and Smith, B.M. A method for measuring arm movements in man under ambulatory conditions. *Ergonomics* 27, 651–661, 1984.

Rode, A., and Shephard, R.J. Cardiorespiratory fitness of an arctic community. *J. Appl. Physiol.* 31, 519–526, 1971.

Rode, A., and Shephard, R.J. Ten years of "Civilization." Fitness of Canadian Inuit. *J. Appl. Physiol. Environ. Exercise Physiol.* 56, 1472–1477, 1984.

Rombeau, J., Carson, S.N., Apelgren, K.N., Waters., L., and Frey, C.F. Clinical comparisons of the Lange and McGaw skinfold calipers. *J. Parenteral Enteral Nutr.* 1, 35A, 1977.

Rowell, L.B., Taylor, H.L., and Wang, Y. Limitations to prediction of maximal oxygen intake. *J. Appl. Physiol.* 19, 919–927, 1964.

Saris, W.H.M., and Binkhorst, R.A. The use of pedometer and actometer in studying daily physical activity in man. Part I. Reliability of pedometer and actometer. Part II. Validity of pedometer and actometer measuring the daily physical activity. *Eur. J. Appl. Physiol.* 37, 219–228; 229–235, 1977.

Saris, W.H.M., Snel, P., and Binkhorst, R.H. A portable heart rate distribution recorder for studying daily physical activity. *Eur. J. Appl. Physiol.* 37, 17–27, 1977.

Schulman, J.L., Stevens, T.M., and Kupst, M.J. The Biomotometer: A new device for the measurement and remediation of hyperactivity. *Child Development* 48, 1152–1154, 1977.

Shephard, R.J. Pulse rate and ventilation as indices of metabolic load. *AMA Archiv. Env. Hlth.* 15, 562–567, 1967a.

Shephard, R.J. The prediction of "maximal" oxygen consumption using a new progressive step test. *Ergonomics* 10, 1–15, 1967b.

Shephard, R.J. Normal levels of activity in Canadian city dwellers. *Can. Med. Assoc. J.* 96, 912–914, 1967c.

Shephard, R.J. Computer programmes for solution of the Åstrand nomogram. *J. Spts. Med. Phys. Fitness* 10, 206–210, 1970.

Shephard, R.J. Do risks of exercise justify costly caution? *Phys. Spts. Med.* 5, 58–65, 1976.

Shephard, R.J. *Endurance Fitness* (2nd Ed). Toronto: University of Toronto Press, 1977.

Shephard, R.J. *Human Physiological Work Capacity*. London: Cambridge University Press, 1978.

Shephard, R.J. *Physiology and Biochemistry of Exercise*. New York: Praeger, 1982.

Shephard, R.J., Allen, C., Benade, A.J.S., Davies, C.T.M., diPrampero, P.E., Hedman, R., Merriman, J.E., Myhre, K., and Simmons, R. Standardization of sub-maximal exercise tests. *Bull. W.H.O.* 38, 765–776, 1968.

Shephard, R.J., Allen, C., Bar-Or, O., Davies, C.T.M., Degré, S., Hedman, R., Ishii, K., Kaneko, M., LaCour, J.R., di Prampero, P.E., and Seliger, V. The working capacity of Toronto school children. *Canad. Med. Assoc. J.* 100, 560–566, 705–714, 1968b.

Shephard, R.J., Corey, P., Renzland, P., and Cox, M.H. The influence of an industrial fitness programme upon medical care costs. *Canad. J. Publ. Hlth*, 73, 259–263, 1982.

Shephard, R.J., and Cox, M. Use of the SRL (M–10 Pulmonary Screener) automated spirometer. *Resp. Technol.* 16, 7–18, 1980.

Shephard, R.J., Cox, M., and Simper, K. An analysis of PAR–Q responses in an office population. *Can. J. Publ. Hlth.* 72, 37–40, 1981.

Shephard, R.J., and Cox, M. Step test predictions of maximum oxygen uptake before and after an employee fitness programme. *Can. J. Appl. Spts. Sci.* 7, 197–201, 1982.

Shephard, R.J., Lavalleé, H., Jéquier, J.C., Rajic, M., and Beaucage, C. Un programme complementaire d'education physique. Etude preliminaire de l'expérience pratiquée dans le district de Trois Rivières. pp. 43–54. *Facteurs Limitant l'Endurance Humaine. Les Techniques d'Amélioration de la Performance*. Ed. J.R. LaCour. St. Etienne, France: Université de St. Etienne, 1977.

Shephard, R.J., Lavallée, H., LaBarre, R., Jéquier, J.C., Volle, M., and Rajic, M. The basis of data standardization in prepubescent children. pp. 360–370. In: *Kinanthropometry II*. Ed. M. Ostyn, G. Beunen, and J. Simons. Baltimore. University Park Press, 1980.

Shephard, R.J., Mahoney, A., Flowers, J.F., and Berridge, M.E. On the preliminary screening of elderly exercise volunteers. *Can. J. on Aging*. 2, 1984.

Shephard, R.J., and McClure, R.L. The prediction of cardiorespiratory fitness. *Int. Z. Angew. Physiol.* 21, 212–223, 1965.

Shephard, R.J., Thompson, M.L., Corey, G.C.R., and Phair, J.J. Field testing of pulmonary dynamics. *J. Appl. Physiol.* 13, 189–193, 1958.

Sidney, K.H., and Shephard, R.J. Attitudes towards health and physical activity in the elderly. Effects of a physical training programme. *Med. Sci. Sports* 8, 246–252, 1977a.

Sidney, K.H., and Shephard, R.J. Activity patterns of elderly men and women. *J. Gerontol.* 32, 25–32, 1977b.

Sidney, K.H., and Shephard, R.J. Maximum testing of men and women in the seventh, eighth and ninth decades of life. *J. Appl. Physiol.* 43, 280–287, 1977c.

Sloan, A.W. Estimation of body fat in young men. *J. Appl. Physiol.* 23, 311–315, 1967.

Sloan, A.W., Burt, J.J., and Blyth, C.S. Estimation of body fat in young women. *J. Appl. Physiol.* 17, 967–970, 1982.

Standardized Test of Fitness. *Operations Manual.* (2nd ed.) Ottawa: Fitness and Amateur Sport, 1981.

Stunkard, A.J., and Albaum, J.M. The accuracy of self-reported weights. *Amer. J. Clin. Nutr.* 34, 1593–1599, 1981.

Von Döbeln, W. Kroppstorlek, Energieomsattning och Kondition. In: *Handbok i Erogonomi*. Eds G. Luthman, U. Aberg, and N. Lundgren. Stockholm: Almqvist and Wiksell, 1966.

Weiner, J.S., and Lourie, J.A. *Practical Human Biology*. New York: Academic Press 1981.

Wilmore, J.H., and Behnke, A.R. An anthropometric estimation of body density and lean body weight in young men. *J. Appl. Physiol.* 27, 25–31, 1969.

Wilmore, J.R., and Behnke, A.R. An anthropometric estimation of body density and lean body weight in young women. *Amer. J. Clin. Nutr.* 23, 267–274, 1970.

Wolff, H.S. The integrating pneumotachograph: A new instrument for the measurement of energy expenditures by indirect calorimetry. *Quart. J. Exp. Physiol.* 43, 270–283, 1958.

Wolff, H.S. Physiological measurement on human subjects in the field, with special reference to a new approach to data storage. In: *Human Adaptability and its Methodology*. Ed: Yoshimura, H., and Weiner, J.S. Tokyo: Japan Society for the Promotion of Sciences 1966.

Yamaji, K., and Shephard, R.J. Factors influencing the use of post-exercise heart rates as indices of cardio-respiratory condition. *J. Spts. Cardiol.* 1985. In press.

Young, C.M., Martin, M.E.K., Tensuan, R., and Blondin, J. Predicting specific gravity and body fatness in young women. *J. Amer. Diet. Assoc.* 40, 102–107, 1962.

Part VI

Contexts of Evaluation

Chapter **19**

Genetic Considerations in Physical Fitness

Robert M. Malina, Ph.D.
University of Texas, Austin

Claude Bouchard, Ph.D.
Laval University, Ste-Foy, Quebec

Introduction

Although concern for the physical fitness of the American population is seemingly high at present, relatively little is established on the genetic bases of fitness. Advances in analytic methods have raised questions about interpretations of earlier genetic data and have emphasized the need to specify environmental sources of variation and genotype–environment interactions in estimating the genotypic contribution to variance in such a multifactorial trait as physical fitness (1).

Operational definition of physical fitness

Physical fitness is a general concept that can be viewed in a number of ways. For convenience, a primary twofold division will be used: physiological and motor components of fitness. The physiological or organic component relates to the processes of energy transformation into work output. It relates to the oxygen-transporting system and depends to a large extent on the efficiency of the cardiovascular system. The amount of energy transformed and used in the working muscle to replenish adenosine triphosphate (ATP) is directly related to the physical or mechanical output. Thus, the physiological component of fitness is sometimes referred to as physical working capacity (total amount of work produced) or power (amount of work per unit time). The kind of work, however, varies; and it is commonly viewed in terms of its dependence on aerobic and/or anaerobic processes. The former refers to the capacity to work as long as possible under aerobic conditions. Maximal aerobic power is defined as the highest oxygen uptake that an individual can attain during physical work. Aerobic power must be differentiated from aerobic capacity or endurance, which is defined as the capacity to work under aerobic conditions for a long period of time. The World Health Organization (2, p.5), for example, has extended the maximal aerobic power concept to a maximal aerobic endurance, "the maximum time a certain fraction of the maximum aerobic power can be sustained." An obvious problem is the determination and maintenance of the appropriate fraction of maximal aerobic power for a given work task. In this report, endurance performance is defined as the total work output in a cycle ergometer test lasting 90 minutes during which the subject is asked to work maximally and produce as much work as possible. Anaerobic capacity refers to the capacity to work as long as possible under predominantly anaerobic conditions. Anaerobic power is the potential to produce work per unit time with muscle ATP being replenished predominantly by the anaerobic energy release mechanisms.

Aerobic and anerobic capacities and powers are related. An individual with a greater maximal oxygen uptake will, on the average, be able to work at a given load for a longer period than an individual with a lesser maximal oxygen uptake. In addition, anaerobic capacity is related to the individual's psychological willingness or motivation to tolerate the physiological discomfort of anaerobic work.

The motor component of physical fitness relates to the development and performance of movement skills, and the term *motor fitness* is often used. Motor tasks are frequently categorized as fine and gross. The former refers to movements requiring precision and dexterity as in manipulative tasks. The latter refers to movements of the entire body and/or major segments of the body as in locomotor activities. Many tasks incorporate both fine and gross motor elements, for example, the precision necessary for projecting an object accurately and with sufficient speed. Factor analytic studies of motor performance have consistently identified several more or less similar factors that underlie proficiency in motor tasks. The components of motor performance and, in turn, motor fitness include the following: strength, speed, power, balance, agility, flexibility, and coordination. An endurance factor is also included. However, it is variably defined by distance runs as an indicator of cardiovascular endurance, or by tasks requiring repeated movements with body weight as the resistance in a certain period of time (e.g., speed situps), or by tasks requiring the support of the body weight for a given period of time (e.g., flexed arm hang) as indicators of muscular endurance.

The health status and physical fitness of an individual are, to some extent, related; and concern for specific aspects of fitness that may be indicative of health status has a long tradition in the fitness literature (e.g., 3, 4). The term health-related fitness is used in this context, and most recently the American Alliance for Health, Physical Education, Recreation and Dance (5) has promoted the use of a health-related fitness for children and youth. It is defined in terms of fatness, flexibility, abdominal muscular strength and endurance, and cardiovascular endurance. Excessive fatness and reduced endurance are related, and both are associated with increased risk for certain degenerative diseases. Reduced flexibility may be associated with limited functional or movement capacity. Genotypic contributions to each of these components of health-related fitness are reviewed in the sections dealing with physiological fitness, motor fitness, and fatness.

Physical fitness: Multifactorial phenotypes

The effects of genes on a given phenotype can occur in at least three different ways: first, by their contribution to a trait or to traits correlated with the phenotype; second, by the heritability, primarily the additive genetic effect, of the phenotype for a given population; and third, by genotype dependence of the adaptive response to components of lifestyle, e.g., diet, level of habitual physical activity, and smoking. Most of the available data on the genetics of physical fitness deal with the second, i.e., heritability estimates. More recently, the third has been recognized as important in understanding the genetics of fitness, for example, the response to training may be genotype dependent. In addition to the above, ample evidence supports the contention that genes also affect physical fitness through their action on traits that covary with the physical fitness phenotypes. Covariates of aerobic as well as motor performance include components of physique and body composition. These covariates of performance should, therefore, be given proper consideration in any discussion of physical fitness, although they will not be considered in detail in this review.

The basic model of quantitative genetics applicable to the study of phenotypes influenced by inherited factors, components of lifestyle, and the interaction of the two has been reviewed in more detail elsewhere (1). Such phenotypes, including physical fitness and its components, are appropriately referred to as multifactorial. Briefly, the total phenotypic variance for such traits can be partitioned accordingly:

$$V_P = V_G + V_E + V_{G \times E} + e$$

where V is the variance, P is the phenotype, G is the genotypic component of the variance, E is the environmental component of the variance, G × E is the interaction effects between the genotype and environment, and e is the random error component.

Most studies of the genetics of human biological traits have assumed that $V_{G \times E} = 0$, strictly for convenience and with little evidence to support it. Hence, it is necessary to be aware of the possibility that for some biological traits, and in particular the components of physical fitness, this assumption may not be met in practice. Several additional assumptions are often made to permit an analysis of casual components of variation in human populations.

First, it is generally assumed under most models that several genes are involved in the particular trait under study and that these genes segregate independently (6, 7). However, linkage may result in the recovery of gene combinations in the offspring generation different from those expected in independent segregation. The phenomenon of linkage does not disturb mean measurements but may affect second degree statistics as variances and covariances (8).

Second, it is assumed that loci contribute in an additive way to the observed variance in a particular trait. This assumes absence of interlocus deviation components, and estimates of this source of variation in human quantitative traits are unsatisfactory (7). However, evidence suggest that the interloci interaction effect would inflate heritability only slightly (9).

Third, it is often assumed that there is no significant genotype–environment correlation. This assumption has received little attention in human biological studies, but it has been discussed at length in connection with genetic studies of behavioral traits. Plomin et al. (10), for example, distinguish between four types of genotype–environment covariation—passive, reactive, active, and negative—and emphasize that the genotype–environment correlation is a function of the frequency with which certain genotypes and certain environments occur together. The presence of a genotype–environment correlation will distort the size of V_G and V_E. The direction of the distortion is related to the sign of the covariation (10).

Fourth, it is often assumed that e is a random source of error uncorrelated with any of the other parameters in the full model. In general, e increases the size of V_P, and reduces the size of genetic estimates accordingly. If e is caused only by measurement error, it becomes possible to correct the attenuated correlations for unreliability and render e = 0. This procedure, however, has not been followed in most studies.

The genetic effect for a given multifactorial phenotype can be considered as the contribution of genes in the population irrespective of environmental factors and lifestyle, i.e., the G effect is a population parameter reflecting largely the extent of the transmissible genetic variance from the parental generation to the next generation. It simply expresses the relative contribution of genetic variation to the interindividual differences observed for the phenotype in the population under average environmental conditions. Therefore, the G effect does not imply that, in a given individual, P is inherited to the extent of G. The knowledge of G for a given P has little predictive value as to the genetic endowment for a given individual.

The estimate of the population parameter G for a given P requires a complex data set and is not a simple task. Ideally, data gathered on all types of two and three generation relatives, including adopted individuals, twins reared together and apart, and families of MZ twins, should be used. Moreover, large sample sizes are necessary to derive reliable estimates of G and related parameters. The majority of the data currently available do not meet these stringent criteria, and thus any estimates of G for the components of physical fitness must be viewed with a degree of caution.

Physical fitness in relatives

Studies of biological relatives have been used most often in studies of the genotypic contribution to variance in physical fitness. These include monozygotic (MZ) twins, dizygotic (DZ) twins, full siblings, half siblings, parents and offspring, and cousins. Causal components of variation or covariation for these various pairs of biological relatives are shown in table 1. In addition, cultural relatives, i.e., spouse pairs, foster parents and adopted children, and siblings by adoption, are occasionally included in analyses. Most of the available analyses of physical fitness from a genetic perspective, however, are based on twins, siblings, and parents and offspring. Parent–offspring and sibling similarities are generally analyzed through correlation and regression techniques. Twin data are also analyzed with correlation and regression techniques in addition to analyses of within-pair and between-pair variances. Note, however, that the twin study methodology is more complex than originally thought by early investigators (11). Twin studies have several limitations—for example, inequality of environmental covariance of MZ and DZ twins, small sample sizes, differential effects of age and sex according to twin types, and differences in means and variances between twin samples (11). One can inquire whether MZ twins are more similar because of genetic similarity or because of environmental pressure for similarity? Or, how important is the role of mutual limitation in the activity habits of MZ twins? Hence, estimates of the genetic contribution to physical fitness must also include other genetic relatives.

Genetic contribution to physiological fitness

Most studies of the G effect in physiological fitness are based on the twin model, with less data available for other types of relatives. Characteristics of twin studies of work capacity and aerobic power are summarized in table 2, and MZ and DZ intraclass correlations derived from these studies are summarized in table 3. Studies in this area date back to a 1971 publication in which Klissouras (12) reported on 15 pairs of MZ and 10 pairs of DZ twins for maximal aerobic power and heart rate. Klissouras concluded that most of the observed variation in maximal aerobic power was determined by the genotype (broad heritability of 0.93). In a subsequent study, Klissouras et al. (13) reported comparable results. It is clear, however, that the available twin studies are of unequal quantitative value and yield heterogeneous and often contradictory results. Estimates of the G effect range from as low as zero to more than 90%. Such discrepancies are believed to be associated with limited sample sizes, uncontrolled age or sex effects, laboratory

Table 1. Covariation components under normal cohabitational conditions accounting for resemblance between pairs of relatives

Relatives	Covariation components [a]
First cousins	1/8 V_A
Uncle–nephew	1/4 V_A
Half-siblings	1/4 V_A
Parent–child	1/2 V_A + V_Ec
Full-siblings	1/2 V_A + 1/4 V_D + V_Ec
DZ twins	1/2 V_A + 1/4 V_D + V_Ec
MZ twins	V_A + V_D + V_Ec

[a] V_A = additive genetic component; V_D = dominance genetic component; V_Ec = cohabitational effect which may be different for various pairs of relatives.

Source: Adapted from Bouchard and Malina (1).

Table 2. Characteristics of twin studies dealing with work capacity and aerobic power

Characteristics	Klissouras (12)	Klissouras et al. (13)	Komi et al. (35)	Howald (79)	Engström and Fischbein (15)	Komi and Karlsson (104)	Bouchard et al. (16)	Bouchard et al. (17)
Sample (*n* pairs)	15 MZ, 10 DZ	23 MZ, 16 DZ	15 MZ, 14 DZ	11 MZ, 6 DZ	39 MZ, 55 DZ	15 MZ, 16 DZ	54 MZ, 56 DZ	53 MZ, 33 DZ
Variable	$\dot{V}O_2$max/kg	$\dot{V}O_2$max/kg	PWC_{205}/kg[b]	$\dot{V}O_2$max/kg	PWC_{max}/kg	$\dot{V}O_2$max/kg	PWC_{150}/kg	$\dot{V}O_2$max/kg
Age (years)	7–13	9–52	10–14	$\bar{X} \cong 19.3$	18–19	11–24	9–24	16–34
Effects of age [a]	1.5% in MZ 1.8% in DZ	20% in MZ 28.4% in DZ	12% in MZ 29.7% in DZ	not available	probably not	not available	statistically controlled	statistically controlled
Sex	Male	Male and female	Male and female	Male and female	Male	Male and female	Male and female	Male and female
Effects of sex [a]	none	35% in MZ 53% in DZ	6.8% in MZ 1.2% in DZ	not available	none	not available	statistically controlled	statistically controlled
Means and twin types [a]	no difference	significant difference	no difference	not available	no difference	significant difference	no difference	no difference
Variances and twin types [a]	significant difference	no difference	significant difference	not available	no difference	no difference	no difference	no difference

[a] Computations based on the data available in the original paper.
[b] Computation based on data reported for PWC_{205} and body weight.
Source: Adapted from Bouchard and Malina (20), with several additions.

methods, and differences in mean or variance between twin types (14). The three studies with the largest sample sizes with control over age and sex effects and with no mean and variance differences within twin types give reasonably similar results (15–17). In these studies, MZ intraclass correlations reach about 0.6–0.7, and DZ intraclass correlations reach only 0.3–0.5. The latter correlations are compatible with those generally reported for regular biological siblings (see below).

The study by Engström and Fischbein (15) is significant among the twin studies summarized in tables 2 and 3. It is based on a large sample of twins of the same age and sex—i.e., no age and sex effects in the data—and it had some control over the amount of habitual activity of the subjects. Using the maximum amount of work performed during 6 minutes on a bicycle ergometer (PWC_{max}/kg), the authors observed approximately twice as much within-pair variance in DZ pairs as in MZ pairs. When including only those pairs of twins concordant for the amount of habitual physical activity, MZ pairs were as variable as DZ pairs. The results reported by Weber et al. (18), who studied the effects of training on one member of a twin pair while the other served as a control, are generally consistent with those of Engström and Fischbein (15), i.e., the relative contribution of G is reduced significantly with the influence of training.

Relatively few studies of aerobic work capacity and power in various family members are available. Montoye et al. (19) tested the heart rate response to exercise in about 4,500 participants in the Tecumseh community health study. The subjects stepped up and down a 20.3 cm bench at a rate of 24 steps per minute for 3 minutes, an energy expenditure of about 5 METS. There was no significant relationship between the heart rate responses of husbands and wives to the exercise test. However, siblings and parents and their offspring exhibited statistically significant associations in their respective heart rate response to the step test. Using measured or estimated maximal aerobic power adjusted for age, body weight, and body fatness, Montoye and Gayle (20) reported a significant father–son correlation ($r = 0.34$, 93 pairs). The relationship between fathers and sons was stronger in those pairs in which the father was less than

Table 3. Intraclass correlations from twin studies of work capacity and aerobic power

Study	*n* pairs MZ	*n* pairs DZ	Variable	MZ	DZ
Klissouras (12)	15	10	Max ml O_2/kg·min^{-1}	.91	.44
Klissouras et al. (13)	23	16	Max ml O_2/kg·min^{-1}	.95	.36
Komi et al. (35)	15	14	PWC_{205}/kg	[a].93	[a].43
Engström and Fischbein (15)	39	55	PWC_{max}/kg	.68 [b].70	.28 [b].50
Bouchard et al. (16)	54	56	PWC_{150}/kg	.60	.41
Bouchard et al. (17)	53	33	Max ml O_2/kg·min^{-1}	.70	.51

[a] Calculated from data in original paper. Data were missing for one pair of MZ twins.
[b] For twins concordant for physical activity patterns.
Source: Adapted from Bouchard and Malina (20), with several additions.

40 years of age, thus implying an age factor in such correlational studies. The brother–brother (r = 0.19, 70 pairs) and spouse (r = 0.18, 27 pairs) correlations, however, were not statistically significant.

More recently, Bouchard (21) reported the degree of resemblance in estimated and directly measured maximal oxygen uptake per unit body weight between several kinds of relatives in a French Canadian sample (table 4). Correlations for measured $\dot{V}O_2$max/kg are generally lower than those for estimated $\dot{V}O_2$max/kg, with the exception of the mother–child correlation. The parent–child and twin correlations appear to suggest a relatively small genetic effect for $\dot{V}O_2$max/kg, and the sibling data suggest a larger genetic effect. The latter, however, tends to be inflated. The difference between the mother–child (r = 0.28) and father–child (r = −0.01) in measured $\dot{V}O_2$max/kg is possibly of interest in that it suggests a maternal effect for this phenotype. One can build a theoretical case to account for this effect by mitochondrial DNA, which is known to be maternally inherited (22). Because this 16-kb base pairs DNA is known to code for 2 rRNAs, 22 tRNAs, and 13 mRNAs apparently associated with the maintenance and function of the mitochondrion, this observation perhaps merits more attention.

Another relevant finding in this analysis of French Canadian familial data is the persistence of familial resemblance in estimated $\dot{V}O_2$max/kg even when fatness, socioeconomic status, habitual physical activity, and smoking habits are statistically controlled (23). Thus, although gene sharing seems to be associated with covariation between biological relatives in maximal aerobic power, so does cohabitation as evident in the data for spouses.

More recently, Bouchard et al. (17) considered $\dot{V}O_2$max and endurance performance in a large sample of brothers (27 pairs), DZ twins (33 pairs), and MZ twins (53 pairs). The G effect reached about 40% for $\dot{V}O_2$max/kg and 70% for 90-minute work output/kg. When expressed per unit of fat-free mass, estimates of the G effect decreased to about 10% for $\dot{V}O_2$max and to about 60% for 90-minute work output. These results thus suggest that endurance performance capacity in this sample is more affected by inherited factors than is maximal aerobic power.

Table 4. Familial resemblance in estimated and measured maximal aerobic power per unit body weight ($\dot{V}O_2$max/kg) based on residuals of age and sex

	Estimated		Measured	
Relatives	n pairs	r	n pairs	r
Spouses	119	.33	[a]20	.20
Parent–child	564	.17	[a]109	.03
Father–child	307	.17	[a]60	−.01
Mother–child	257	.17	[a]49	.28
Biological siblings	223	.33	[a]47	.19
DZ twins			20	.73
MZ twins			37	.76

[a] Data obtained in a treadmill test; all others in a cycle ergometer test.

Source: Adapted from Bouchard (21).

Resemblance between French Canadian relatives in submaximal power output (PWC_{150}/kg body weight) is shown in table 5. Spouses significantly resemble each other in this submaximal test, and such similarity is probably associated with positive assortative mating for the trait (or relevant covariate) and/or a shared lifestyle, i.e., habitual energy expenditure or exercise habits. In contrast, sibships of adoptees and of first cousins do not show significantly more variance between sibships than within sibships. However, sibships of biological brothers and sisters and DZ twins and MZ twins characteristically have significant levels of resemblance for PWC_{150}/kg. The DZ and MZ data also suggest that the submaximal power output measurements are affected by a relatively low genetic effect, around 25–30%.

In summary, results concerning the heritability of physiological fitness as measured by maximal oxygen uptake and submaximal power output must be viewed with caution. The claims of earlier studies based on rather small samples of twins appear to be unrealistic when compared with the more recent and more comprehensively analyzed twin and familial data.

Data for long-distance runs, which are used as indicators of aerobic capacity or sometimes endurance, are rather limited. Klissouras's (24) correlations for preadolescent twins (MZ = 12 pairs, DZ = 5 pairs, sex not specified) give an estimated heritability of 0.94 for the 1,000-meter run, while Weiss's (25) observations on 480 pairs of 10-year-old twins (male and female) give an estimated heritability of 0.93 for the distance run in 7 minutes. The distance run in the latter study was given as a group test, so that the motivation factor may not have been controlled and the zygosity of the twins was not tested. Although these results suggest a significant genetic component in the two distance-running tests, the estimated heritabilities must be accepted with caution.

Level of habitual physical activity and/or training is an environmental variable that probably interacts with the genotype. Further, the evidence indicates considerable individual variation in response to aerobic training. For example, Lortie et al. (26) submitted 24 sedentary subjects (as rated on the basis of questionnaire and interview) to a standardized, 20-week aerobic-training

Table 5. Familial resemblance in submaximal power output (PWC_{150}/kg)

Relatives	n cells	Subjects	Intraclass correlation
Spouses	276	552	.19
Adoptive sibships	46	107	.00
Cousin sibships	33	87	.14
Biological sibships	225	531	.25
DZ sibships	56	117	.46
MZ sibships	54	110	.60

Source: Adapted from Bouchard et al. (16); based on residuals of age and sex.

program. Average improvement in maximal aerobic power was 33%, but individual variation in gains varied from 5% to 88%. The individual scores thus suggested high responders and low responders to the endurance-training program that was designed to improve $\dot{V}O_2max$ and endurance. Given this range of response to a standardized training program, it is imperative that factors underlying such differences be more closely examined.

Figure 1 presents a model of factors potentially related to human variation in sensitivity to aerobic training. These factors include age and sex, prior training experience, current phenotypic level, and genotype. Although age can be readily controlled to some extent, the available evidence indicates that the response to aerobic training is quite variable even within relatively narrow age ranges. The issue of prior experience or training history has generally not been considered. However, by focusing on sedentary subjects with no history of regular participation in sports and/or other physical activities, the issue of prior experience can be controlled to some degree. The remaining three factors are related in part to the individual's genotype. Sex differences in trainability may be a factor. Lortie et al. (26), for example, reported that maximal aerobic power had the same trainability in males and females. However, endurance performance was, on the average, more trainable than maximal aerobic power, and males had a greater trainability than females under the similar laboratory conditions. Hence, the sex of subjects probably accounts for some of the variation observed in the response to training.

One of the most difficult factors to interpret in the model is the current phenotype, i.e., the pretraining level in the present context, and its relationship with the observed response to training. Evidence suggests that the trainability of maximal aerobic power and other aerobic output criteria is negatively related to the pretraining level (27–30). In a review of data from 50 published training experiments, the correlation between training-induced changes in $\dot{V}O_2max/kg$ and the pretraining level of this attribute was negative and significant ($r = -0.5$) (27). Further, when the frequency, intensity and duration of training sessions as well as the number of weeks of training were statistically controlled, the correlation was reduced negligibly ($r = -0.4$). In the study of Lortie et al. (26), cited above, the correlation between training gains in $\dot{V}O_2max/kg$ and pretraining values in the sedentary subjects was about -0.6, and that between training gains in endurance and pretraining values was about -0.5. These observations thus imply that the initial phenotypic level accounts for a significant proportion of the variance observed in response to aerobic training. But, when dealing with subjects confidently judged as sedentary, the current phenotypic level is partly a reflection of the genotype. The variance in the training response associated with the pretraining phenotypic level can under these conditions be viewed as a component of the G × E effect.

The fifth factor in the model proposed in figure 1 is the genotype-training interaction component associated with genetic variation, which is presently unknown but which is independent of pretraining differences in phenotypic level (P). This interaction component is represented by the term G × E and implies that sensitivity to aerobic training depends to some extent on an individual's genotype. Although doubts were initially expressed on the presence of an interaction effect (18, 31), the currently available evidence supports the presence of a G × E interaction component in aerobic work performance. More importantly, G × E appears to be a major component of the variance in response to training. For example, Prud'homme et al. (32) subjected 10 pairs of MZ twins to a 20-week endurance-training program, initially meeting four and then increasing to five times

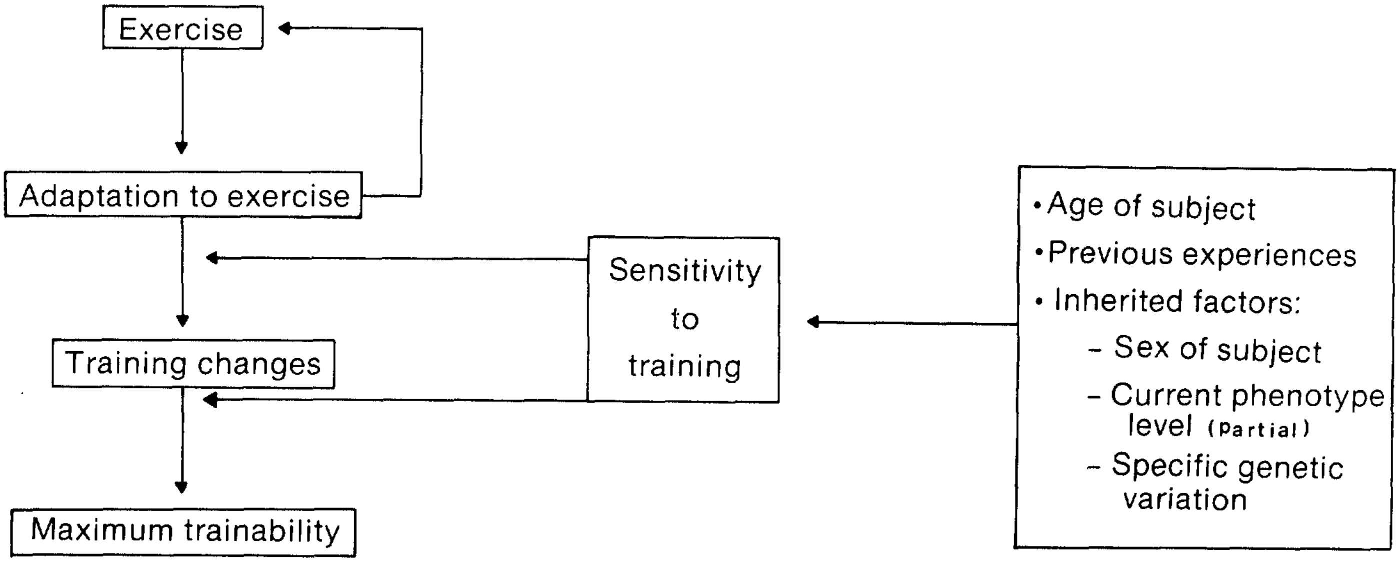


Figure 1. A model of factors associated with human variation in sensitivity to exercise training. Reprinted from Bouchard (24) as modified after Bouchard and Malina (1).

per week. Training sessions initially lasted for 40 minutes and then increased to 45 minutes. Training intensity averaged 80% of the maximal heart rate reserve (beginning at 60% and increasing to 85%). Under this program, maximal aerobic power improved by 14%. There was, however, considerable interindividual variation in response to the training program. Gains in $\dot{V}O_2$max/kg ranged from 0% to 41%. The differences in response to training were not, however, randomly distributed among the twin pairs. Thus, intraclass correlations computed with the gain expressed as a percentage of pretraining $\dot{V}O_2$max/kg reached 0.82, indicating that members of the same pair yielded a reasonably similar response to the training program, i.e., 82% of the variance in the training response appeared to be genotype dependent. Hence, these observations suggest that the sensitivity of maximal aerobic power to training is largely genotype dependent and that the presence of a meaningful G × E effect implies that there are high responders and low responders to aerobic training (see 21, 33).

Tentative estimates of sources of variation in maximal aerobic power and endurance performance in sedentary individuals submitted to training are summarized in table 6. Most of the population variance in $\dot{V}O_2$max/kg is genotype dependent, providing that age variation does not contribute significantly to the data. Presumably prior training experience contributes negligibly and is set at zero. Under these conditions, the average environmental effect reaches about 20%, the genetic effect reaches about 30%, and the total G × E effect reaches about 50% of the variance. The pretraining effect, which is largely genotype dependent in sedentary subjects, accounts, on the average, for about 10% of the variance or 20% of the G × E effect. These parameter estimates are only slightly different for endurance performance.

A major problem that has not yet been adequately investigated is the identification of individuals who are high or low responders to aerobic training. There are at present no genetic markers that one can use to type an individual for sensitivity to training, i.e., given the present state of knowledge, sensitivity of maximal aerobic power and maximal aerobic capacity to aerobic training is unpredictable. It can be estimated only from past training experiences or as training progresses (see 21).

Table 6. Tentative parameter estimates of causal sources of variation in maximal aerobic power ($\dot{V}O_2$max/kg) and 90-minute endurance performance (J/kg in 90 min) in sedentary individuals submitted to training

Parameter	Maximal aerobic power	90-minute endurance performance
G effect	30%	20%
E effect	20%	40%
G × E effect:		
Main G × E effect	40%	25%
Pre-training effect	10%	5%
Sex effect	0%	10%

Note: Most probable values in a population of sedentary individuals submitted to training, but ranges of values remain possible at this stage of research, particularly, when dealing with small samples and/or individuals of mixed training status.
Source: Adapted from Bouchard (21).

Genetic contribution to motor fitness

The domain of motor fitness includes a variety of components, including strength, speed, power, agility, coordination, and flexibility, among others. A variety of tests have been used to evaluate performance on these components. Further, many of the tests are not suitable for adult samples, e.g., dashs and jumps. Hence, studies of the genetic contribution to motor fitness are limited, to a large extent, to samples of schoolage children and youth. This is in contrast to the data for maximal aerobic power, which are derived largely from young adult samples.

Estimates of the genetic contribution to the various components of motor fitness are offered subsequently. They are based largely on the traditional twin model with less information available for siblings and parents and their offspring. Further, estimates of heritability are derived by different methods (see 1).

Estimated heritabilities for measures of muscular strength are given in table 7. Strength is an expression of muscular force, or the capacity of an individual to exert force against some external resistance. Three kinds of strength measurements are commonly utilized: (1) static or isometric in which force is exerted without any change in muscle length; (2) dynamic in which force is generated by shortening (concentric contraction) or lengthening (eccentric contraction) the muscles; (3) explosive in which maximal force is released in the shortest possible time. The latter is often referred to as muscular power, and it is measured quite frequently by tests of jumping and throwing. The data summarized in table 7 are based primarily on static and dynamic strength tasks. Estimated heritabilities vary among studies, are generally higher for boys than for girls, and are generally higher in studies of twins than of siblings. Age variation within most samples is not ordinarily controlled. The low estimated heritabilities for females suggest environmental sources of variation in such tasks—e.g., social and motivational factors—or, perhaps, result from the random inactivation of one of the X-chromosomes (Lyonization) in the various cell lines, causing phenotypic heterogeneity not observed with genes on the autosomes.

Heritability estimates from four Eastern European studies that were not readily available to the authors but were reported by Kovar (34) range from 0.24 to 0.71 for grip strength and 0.11 to 0.71 for trunk extension strength. Further, heritability estimates for strength expressed per unit body weight or for strength scores summed over several muscle groups tend to be similar to those for specific muscle groups.

In a related study, Komi et al. (35) used several measures of muscular force (isometric, concentric, eccentric) in 15 pairs of male and 14 pairs of female twins 10–14 years of age. Sample sizes, however, were small so that the genetic and/or environmental significance of the observations is not clear. Although differences in muscular forces within eight male MZ pairs were, on the average, less than those within seven male DZ pairs, the intrapair variances did not differ. Among female twins, on the other hand, seven pairs of MZ twins showed as much variability in muscular forces as seven pairs of DZ twins.

Using a composite strength score per unit height (based on handgrip, knee stretch, and arm bent tests) in male twins 18–19 years of age, Engström and Fischbein (15) noted more similar performances in 39 MZ pairs than in 55 DZ pairs. The intrapair correlations were 0.83 and 0.47 for MZ and DZ twins, respectively. When the amount of habitual leisure time physical activity in a subset of the twins was controlled, the intraclass correlation for MZ twins changed only slightly from 0.83 to 0.80, and that for DZ twins decreased from 0.47 to 0.33. Thus, as in the studies of aerobic capacity discussed earlier, there is a need to control for the amount of habitual activity in attempting to quantify the genotypic contribution to variance in strength.

In addition to the sibling data presented in table 7, other sibling data are analyzed in a different manner. Montoye et al. (19) compared the aggregation of siblings in the Tecumseh community health study for several measurements of strength: arm strength using both arms simultaneously, the sum of right and left grip strength, and a relative strength index based on the two measurements corrected for body size and fatness. Siblings (314–416 pairs) were classified into tertiles by age and sex and similarities then analyzed, i.e., whether both siblings were in the same tertile. The results indicated a significant degree of sibling similarity in the strength

Table 7. Estimated heritabilities for measures of muscular strength

Source or sample	Test	Heritabilities		
		Males	Females	Sexes combined
Venerando and Milani-Comparetti (105), twins, 9–11 and 15–17 years, males: 12 MZ, 12 DZ pairs in each age group	right grip	.32		
	left grip	.45		
	back extension	.49		
Sklad (43), twins, 8–15 years, males: 23 MZ, 19 DZ pairs, females: 22 MZ, 36 DZ pairs	elbow flexion	.61	.71	
	elbow extension	.64	.46	
	knee flexion	.69	.44	
	knee extension	.82	.68	
	back lift	.80	.73	
Kovar (41), twins, 11–25 years, males: 17 MZ, 13 DZ pairs	push-ups	.22		
	sit-ups	.45		
	bent arm hang	.62		
	grip	.45		
	arm	.75		
	trunk extension	.57		
	elbow flexion ergometry	.83		
Weiss (25), twins, 10 years, 327 same sex, 153 opposite sex pairs, zygosity not tested	pull-ups			.85
Malina and Mueller (38), siblings, 6–12 years, males and females, 198 pairs, all estimates are body weight and test reliability corrected	right grip			.55
	left grip			.45
	push			.41
	pull			.34
Szopa (37), siblings, 3–42 years, males and females, approximately 728 pairs, all estimates adjusted for assortative mating	Siblings 3–10 years			
	right grip	.37	.20	
	left grip	.33	.00	
	grip/weight	.27	.00	
	arm	.37	.38	
	arm/weight	.38	.32	
	Siblings 11–17 years			
	right grip	.21	.09	
	left grip	.32	.00	
	grip/weight	.30	.23	
	arm	.48	.27	
	arm/weight	.46	.32	
	Siblings 18–42 years			
	right grip	.40	.29	
	left grip	.32	.15	
	grip/weight	.37	.03	
	arm	.45	.48	
	arm/weight	.31	.50	

tests, but the strength of the resemblance was not quantified. Montoye et al. (19) also noted significant parent–child resemblances (475 pairs) in the three strength indicators, and there did not appear to be an age effect, i.e., resemblances between parents and older children (16–39 years) were similar to those between parents and younger children (10–15 years). However, female offspring appeared to resemble their parents more than male children.

Wolański and Kasprzak (36) compared several strength tests in parents and their offspring, 7–39 years, from 570 three-generation families in rural Poland. Numbers in specific age groups were small and ages of offspring and of parents were apparently not controlled in the analysis. Given these limitations, grip strength showed a reasonable degree of parent–offspring similarity across the various pairings, and father–offspring correlations tended to be slightly higher than mother–child correlations. In contrast, parent–offspring correlations for back and shoulder strength were uniformly low.

In an urban Polish sample, Szopa (37) considered age variation in parent–child resemblances in strength in 1,420 individuals from 347 primarily one-generation families. Children ranged in age from 3 to 42 years, and parents ranged from 22 to 78 years. The correlations were generally low for grip strength and higher for arm strength. Parent–son correlations were higher than parent–daughter values for grip strength, but father–son and mother–son correlations were similar. In arm strength, mother–daughter correlations were generally similar to those for father–son pairs, and mother–son correlations were consistently higher than those for father–daughter pairs. Hence, male offspring were more similar to both parents in arm strength (to fathers more so than mothers), but females were only similar to their mothers.

Comparing the two Polish studies (36, 37), parent–offspring correlations for grip strength in the urban sample were consistently lower than those observed in the rural sample. However, the correlations for shoulder strength in the latter were less than those in the urban sample. The differences may reflect variation in socioeconomic status and life style (including physical activity that was probably greater in the rural sample), as well as sample size and test procedures. On the other hand, the pattern of correlations for arm and shoulder strength was similar in both studies. Girls were more similar to their mothers and sons were more similar to their fathers, but father–daughter correlations were lower than mother–son correlations in both studies.

Tests of running, jumping, and throwing are among the more commonly used tasks in motor fitness assessments of children and youth. Most such tests are indicators of explosive strength. Dashes require the body to be moved as rapidly as possible over a short distance. Jumps require the body to be moved either vertically or horizontally through space usually from a standing position. Throws, on the other hand, require an object to be projected forcefully, usually as far as possible. It should be noted that each of these tasks requires a considerable degree of coordination, although the explosive nature of the tasks is most emphasized. A variant of the runs, the shuttle run, is often used as an indicator of agility, the ability to change direction of movement rapidly.

Estimated heritabilities for performance on several tests of running, jumping, and throwing are shown in table 8. Although the heritabilities are estimated by different methods, the data suggest a greater genetic influence in the runs and jumps than in the throws. Sex differences in heritabilities are not as evident as in the strength tests (table 7), but the estimated heritabilities for the vertical jump in the data of Szopa (37) show a sex difference at the younger ages and an increase in esti-

Table 8. Estimated heritabilities for performance on tests of running, jumping, and throwing

Source or sample	Test	Heritabilities		
		Males	Females	Sexes combined
Sklad (43), twins, 8–15 years, males: 15–23 MZ, 12–19 DZ pairs, females: 12–22 MZ, 14–36 DZ pairs	60 meter dash	.72	.80	
	vertical jump	.75	.63	
Kovar (41), twins, 11–25 years, males: 17 MZ, 13 DZ pairs	shuttle run	.90		
	vertical jump	.86		
	medicine ball throw	.60		
Weiss (25), twins, 10 years, 327 same sex, 153 opposite sex pairs, zygosity not tested	60 meter dash			.85
	long jump			.74
	triple jump			.66
	cricket ball throw			.54
	shot put			.71
Malina and Mueller (38, siblings, 6–12 years, males and females, 198 pairs, all estimates corrected for body weight and test reliability	35-yard dash			.81
	standing long jump			.60
	softball throw			.67
Szopa (37), siblings, 3–42 years, males and females, approximately 728 pairs, all estimates adjusted for assortative mating	vertical jump by sibling age:			
	3–10 years	.44	.33	
	11–17 years	.68	.47	
	18–42 years	.59	.63	

Table 9. Estimated heritabilities for various tests of balance in twins

Source or sample	Test	Heritabilities: Males	Females	Sexes combined
Vandenberg (85), high school age, males and females, 41 MZ, 32 DZ pairs	Seashore beam balance			.48
Sklad (43), 8–15 years, male: 23 MZ, 19 DZ pairs, female: 22 MZ, 36 DZ pairs	One foot lengthwise on rail, eyes open: right foot	.86	.63	
	left foot	.81	.76	
Williams and Hearfield (106), mean age 14.8 years, males and females: 15 MZ, 8 DZ pairs	Bachman ladder climb			.46
Williams and Gross (107), 11–18 years, males and females: 22 MZ, 41 DZ pairs	Stabilometer			.27

mated heritability with age. Although the heritability estimates reported by Malina and Mueller (38) are based on a combined sample of siblings, brother–brother correlations were higher than sister–sister correlations for the dash and the throw but were negligible for the long jump. Heritability estimates from Eastern European studies of twins between 5 and 17 years, which were not readily available to the authors but were reported by Kovař (34), are 0.45–0.74 for the long jump, 0.82 for the vertical jump, and 0.14 for the shotput. Estimates for the dash vary with distance but not in a consistent pattern, i.e., 0.83 for a 20–meter dash, 0.62 for a 30–meter dash, and 0.45 and 0.91 for a 60–meter dash. These estimates are generally consistent with those summarized in table 8.

Parent–offspring data for such explosive strength tasks are not extensive, probably reflecting the task demands and age-associated decrements in performance on such tests. Szopa (37), however, did include parent–child comparisons for the vertical jump (see table 8 for an indication of sample sizes). The correlations ranged from 0.17 to 0.54 and midparent correlations tended to be higher than father–child or mother–child correlations. However, in the oldest group of siblings (18–42 years), sons were more similar to their fathers and daughters were more similar to their mothers.

In an interesting study of a small sample, Cratty (39) compared the performances of 24 college-age males with those of their fathers who attended the same college 34 years earlier, i.e., when their fathers were of college age. Father–son correlations were 0.86 for the running long jump, 0.59 for the 100–yard dash, and 0.04 for the fence vault. Hence, fathers and sons attained reasonably similar performances in the two explosive strength tasks.

The preceding data concerns only the outcomes of performance on the explosive strength tasks. Some limited data also indicate a significant genotypic contribution to variation in the kinematic structure of the 60-meter dash in twins 11–15 years of age (40). Further, the heritability estimates for stride length, tempo, and limb and trunk angles at different phases of the run were greater in males than in females. This would suggest that running performance in girls is more amenable to environmental influences, including of course social and motivational factors for an all-out performance.

Balance is a skill that requires a combination of gross and fine motor control in the maintenance of equilibrium. It is obviously an integral component of skillful performance in many motor tasks. Heritability estimates for a variety of balance tasks derived from studies of twins are summarized in table 9. Estimates range from 0.27 to 0.86 and vary with the test used. Some limited data suggest a slight sex difference.

Wolański and Kasprzak (36) reported parent–child similarities in three balance tasks in a rural Polish sample. Ages of the offspring varied between 6 and 19 years, but ages of the parents were not reported nor apparently controlled in the analysis. Nevertheless, correlations for beam walking and a timed turning balance test were consistently low (-0.1–0.23), with no clear pattern of father–child or mother–child correlation. In contrast, with the time factor removed from the turning balance test, parent–child correlations were higher (0.12–0.41).

Estimated heritabilities for speed of limb movement, manual dexterity, and manual coordination are summarized in table 10. Speed of limb movement tends to exhibit higher heritabilities than manual dexterity and manual coordination. Right–left differences in estimated heritabilities for speed of limb movement are negligible. The estimates for both speed and manipulative tasks are higher in males, but the opposite is true for arm movement speed and accuracy. Evidence for relatively fine motor skills, i.e., mirror drawing, tweezer dexterity, etc. (table 11), also indicates a significant genetic effect. However, the estimated heritabilities vary with the task and indicate generally higher genetic contributions for performances with the right hand than with the left. Parent–offspring correlations for eye–hand coordination in a tracing task in a rural Polish sample (36) also show a familial pattern, but there is considerable variation in correlations by parent–child pairing.

Data for flexibility are not extensive. In his summary of largely Eastern European data, Kovař (39) reported reasonably high heritabilities for three flexibility tests in a combined sample of male and female twins (50 pairs, 12–17 years of age): 0.84, 0.70, and 0.91 for trunk, hip,

Table 10. Estimated heritabilities for speed of limb movement and manual coordination in twins

Source or sample	Test	Heritabilities Males	Heritabilities Females
Sklad (43), 8–15 years, males: 15–23 MZ, 12–19 DZ pairs, females: 11–22 MZ, 13–36 DZ pairs	Upper limb plate tapping:		
	right	.86	.81
	left	.86	.62
	One foot tapping:		
	right	.90	.65
	left	.79	.72
	Arm movement speed and accuracy	.13	.57
	Manipulative dexterity	.51	.38
Kovar (41), 11–25 years, males: 17 MZ, 13 DZ pairs	Labyrinth (maze) task:		
	time	.48	
	errors	.64	

and shoulder flexibility, respectively. In a sample of male twins 11–25 years of age, the estimated heritability for trunk flexibility was 0.69 (41). Devor and Crawford (42) reported correlations of 0.43 and 0.29, respectively, for trunk flexibility in siblings (n = 53 pairs) and parents and offspring (n = 182 pairs) from a Mennonite community. A subsequent analysis using the Tau path model indicated a high degree of transmissibility for this flexibility measurement. However, the Tau model does not differentiate between the percentage of the transmissible component of phenotypic variance that is genetic and that which is cultural.

The number of possible motor tasks available for testing is almost limitless. Some tests include the performance of a series of stunts that is subsequently combined into a single composite score. Sklad (43), for example, reported moderately high heritabilities of 0.80 and 0.74 for male and female twins, respectively, on the Johnson Test, a test suggested as a measure of inherent capacity for neuromuscular skill (4). The test consists of 10 stunts, e.g., stagger skips, half turns right and left, and each is scored on a 10-point basis.

In summary, estimates of the genotypic component of variance in a number of motor fitness test items vary among studies, tasks and types of genetically related individuals considered. Most estimates are derived from younger subjects, and the majority of the data are derived from twins. The results of the studies, however, are of unequal quantitative value. Many studies are limited to males, and others combine the sexes. Some data suggest higher heritabilities for males than females, and the more limited sibling data suggest a similar trend. Thus, brothers tend to resemble each other more than sisters in most strength and motor tasks. This would suggest, in turn, sex influences in estimated heritabilities of motor fitness items. Perhaps the simplest explanation would appear to be environmental covariation that differs with respect to sex, but Lyonization of one of the X-chromosomes may also be a factor.

Table 11. Estimated heritabilities for fine motor skills in same-sex twins of high school age

Test	Heritabilities Right hand	Heritabilities Left hand
Mirror drawing	.70	.24
Tweezer dexterity	.71	.63
Santa Ana dexterity	.58	.05
Hand steadiness	.37	.17
Rotary pursuit	.52	.32
Card sorting	.61	.71

Note: 40–45 MZ, 30–35 DZ pairs.
Source: Adapted from Vandenberg (85).

The tests used to assess motor fitness are generally done on a single occasion and do not consider change in performance over time. Improvement in performance may include a learning component (analogous to a training effect in physiological fitness), a genetic component, and a genotype–environment interaction. The model proposed in figure 1 for aerobic training is also applicable to motor learning. The sensitivity to specific training or practice can be viewed as potentially related to age, sex, prior experience, current phenotypic level, and genotype.

Studies of motor learning in twins have a reasonably long history, though the evidence is not very extensive and the traditional twin model is used (44). Results of five studies reported between 1933 and 1980 are summarized in table 12. Estimates of the genetic contribution to the learning of motor skills vary from task to task, emphasizing the specificity of motor learning, and one study suggests possible sex differences. Heritability estimates also vary over a series of practice trials or training sessions. It should be noted, however, that most of the tasks used, with the exception of the stabilometer, tend to be fine motor skills that generally stress manual dexterity and precision of movement and more gross movements of isolated body segments. The factors potentially capable of influencing sensitivity to learning are not considered in any of the studies. Hence, more comprehensive analyses of learning effects in tasks more commonly used in motor fitness assessment are obviously needed.

Fatness and fat distribution

The influence of fatness on physiological and motor fitness has been known for some time. Excessive fatness, either absolute or relative, exerts a negative effect on physiological fitness, i.e, reduced aerobic power and capacity can be viewed as a functional consequence of excessive fatness. Similarly, excessive fatness has a negative effect on motor fitness, especially in tasks that require the body to be projected, e.g., jumps and runs, or lifted, e.g., pullups. Excessive fatness is also associated with morbidity and mortality from certain diseases, e.g., heart disease, hypertension, adult onset diabetes.

Fatness, however, is a complex entity. Considerations of fatness must include the absolute and relative amounts of fat, the distribution of adipose tissue in the body, as well as regional variation in adipose tissue metabolism. Most emphasis, however, has been placed on the quantification of absolute or relative fatness, with less attention to variation in fat distribution and metabolism.

Estimates of fatness are derived primarily from densitometry, hydrometry, and gamma ray spectrometry (45–47). These procedures provide estimates only of the absolute or relative fatness of the body as a whole and not of variation in the distribution of fat. Measurements of skinfold thicknesses at various sites on the body are the most commonly used indicators of subcutaneous fat and its distribution and are also frequently used to predict absolute and relative fatness.

Fat distribution refers to the relative anatomical distribution of fat in the body. The literature on fat distribution, however, is confounded by the variety of methods used to define it. Methods include standard score profiles and principal components and discriminant function analysis of skinfold thicknesses taken at a variety of body sites and ratios of specific skinfolds (quite often only the subscapular and triceps skinfolds) and circumferences (most often waist and hip circumferences). One can thus inquire whether comparable indexes of fat distribution are being identified. Skinfolds measure only subcutaneous or external fat, but waist and hip circumferences include subcutaneous and deep fat, in addition to muscle, bone, and viscera. Subcutaneous fat is quite labile, and the ratio of subcutaneous fat to total fat mass may be a meaningful indicator of fat distribution. An estimate of the latter may be derived from the ratio of the sum of several skinfolds taken on the upper and lower extremities and on the upper and lower aspects of the trunk to total fat mass as derived from densitometry (48). Computed tomography provides a more complete description of fat distribution, but the method has limited applicability to epidemiological studies (49).

It is suggested that the distribution of subcutaneous fat in the body is an individual characteristic and that the major component of variation in fat distribution is biological rather than environmental (50, 51). Variation in subcutaneous fat distribution is also related to ethnicity and race (52–56). Environmental factors such as dietary restriction (50) or physical training (57) apparently have more of an influence on fatness per se than on the relative distribution of subcutaneous fat in adults. On the other hand, subcutaneous fat distribution is related to biological maturity status during growth (58–60) and perhaps to environmental changes during growth (61, 62). The latter, if substantiated, suggests the possibility of a G × E effect during the period of growth. Fat distribution is also an important distinguishing characteristic among individuals at risk for several clinical conditions, e.g., diabetes, cardiovascular disease, and aberrant metabolic profiles (63–65).

Genetics of fatness and fat distribution

Information on the degree of resemblance among relatives for measures of fatness and fat distribution is not extensive. Comparisons of skinfold thicknesses in twins indicate moderate heritabilities. The estimated mean genetic effects in five studies of twins that included the triceps and subscapular skinfolds were 0.56 and 0.64, respectively; the mean genetic effect in four studies that included an abdominal skinfold was 0.30; and the mean genetic effect in two studies that included

Table 12. Summary of studies of the heritability of motor learning in twins

Source or sample	Task	Learning or training effect
McNemar (108), junior high school age, males: 47 MZ, 48 DZ pairs	Pursuit rotor Spool packing Card sorting	7 practice segments did not alter similarity between MZ twins, but increased similarity between DZ twins in 2 tasks; changes in heritabilities: 0.78 to 0.67 on pursuit rotor, 0.30 to 0.00 on spool packing, 0.44 to 0.43 on card sorting.
Brody (109), 8–14 years, males: 29 MZ, 33 DZ pairs, reanalyzed by Wilde (110)	Mechanical ability	6 practice trials, greater increase in MZ than DZ similarity; heritability increased from 0.35 on 1st trial to 0.59 on 6th trial.
Sklad (111), 9–13 years, males and females in like-sex pairs: 24 MZ, 22 DZ pairs	Plate tapping with hand One foot tapping Mirror tracing Ball toss for accuracy	Performance curves over 10–14 days were more similar in MZ than DZ, more similar in MZ females than MZ males, especially for mirror tracing and ball toss; 2 parameters of curves: (1) level of learning, intraclass correlations generally higher in male MZ for all tasks except mirror tracing, (2) rate of learning, intraclass correlations generally higher in male MZ, but variable by task.
Marisi (112), 11–18 years, both sexes: 35 MZ, 35 DZ pairs	Pursuit rotor	Heritability decreased from 0.96 on trial 1 to 0.45 after 30 trials; after rest period, increased to 0.85 and decreased to 0.58 after 20 additional trials.
Williams and Gross (107), 11–18 years, both sexes: 22 MZ, 41 DZ pairs	Stabilometer	Heritability low on 1st day of practice, 0.27; increased to 0.69 on 2nd day and remained close to this level over next 4 days.

Table 13. Correlations for the sum of four skinfolds in various sibling pairs in childhood and adulthood

Sibling pairs	*n* pairs	Childhood	Adulthood
MZ twins, males	29	.81	.74
MZ twins, females	20	.81	.81
DZ twins, males	46	.48	.68
DZ twins, females	45	.51	.16
Brothers	[a] 15/25	.20	.30
Sisters	13/20	.34	.05

[a] The first number refers to child and the second to adult pairs.
Source: Adapted from Hawk and Brook (67).

the medial calf skinfold was 0.77 (66). Some evidence indicates lower heritabilities for fatness in adult twins compared with younger twins, implying a significant role for environmental factors in adult fatness (51).

In contrast, comparisons of skinfold thicknesses in parents and their biological offspring and in biological siblings are lower. A comprehensive review of correlations based on approximately 8,800 parent–child pairs and 3,300 full sibling pairs yielded low order correlations that ranged primarily between 0.2 and 0.3 (66). Most data, however, were limited to the triceps and subscapular skinfolds, with less data for other sites. On the average, sibling correlations were slightly higher than parent–child correlations, and sister–sister correlations were consistently higher than brother–brother values. Correlations tended to be lower for parents and adopted offspring and for biologically unrelated siblings (66). Nevertheless, some of the correlations were statistically significant, and thus suggest a role for common environment or social inheritance.

Hawk and Brook (67) did a follow-up study of a series of siblings and twins who were initially measured between 2 and 15 years of age (table 13). Correlations for the sum of four skinfolds in MZ twins changed only slightly from childhood to adulthood, but those for DZ twins and nontwin siblings were more variable. They increased from childhood to adulthood in males but decreased in females. Small numbers and variable ages during childhood are factors that must be considered in evaluating these correlations. Serial observations for skinfold thicknesses and fat widths measured on standardized radiographs from several longitudinal studies of children indicate significant tendencies for individuals to retain quartile ranks for subcutaneous fatness across age if the early measurements were made after 4 years of age (68). There was less stability for measurements made before 4 years of age. Other data for twins, biological siblings, and unrelated individuals indicate higher within sibship similarity for subcutaneous fat and relative fatness (densitometry) in MZ twins, followed in order by DZ twins, biological siblings, and unrelated individuals (66).

Subcutaneous fat is influenced by sex and age. Hence, studies of familial resemblance in fatness should control for these factors. Results of an analysis of resemblance in six skinfold thicknesses in 1,698 individuals from 409 families of French Canadian ancestry, encompassing 9 different kinds of relatives and controlling for age and sex (69), are summarized in table 14. The foster parent–adopted child correlations are comparable to the biological parent–natural child correlations. On the other hand, correlations for DZ twins are consistently higher than those for full siblings, even though DZ twins and full siblings share about 50% of their genes by descent. This observation indicates that common environmental conditions are contributing to the resemblance in these relatives. Comparison of the correlations for foster midparent–adopted child and midparent–natural child pairs indicates little difference between relatives by adoption and relatives by biological inheritance for the six skinfold thicknesses. This suggests that heredity may have little to do with the amount of subcutaneous fat.

Subcutaneous fat is also influenced by diet, regular physical activity, and socioeconomic status (note the well-established inverse relationship between socioeconomic status and obesity). However, differences between age and sex standardized correlations for skinfold thicknesses and those controlling for energy intake, energy expenditure, and socioeconomic status in a subsample of French Canadian relatives are slight (70). Thus, current daily energy intake and expenditure and socioeconomic status apparently have only a small effect

Table 14. Interclass correlations for individual skinfold thicknesses in various pairs of relatives

Relatives	*n* pairs	Triceps	Biceps	Medial Calf	Subscapular	Suprailiac	Abdominal
Foster parent-adopted child	322	.24	.20	.19	.27	.28	.23
Sibs by adoption	120	.06	.12	.17	.17	.18	.12
Cousins (first degree)	95	.26	.26	.07	.30	.25	.27
Uncle/aunt-nephew/niece	88	.06	.00	.11	.14	.11	.00
Parent-natural child	1239	.21	.21	.23	.25	.34	.18
Full sibs	370	.18	.26	.26	.25	.42	.29
DZ twins	69	.30	.39	.48	.31	.49	.32
MZ twins	87	.77	.78	.76	.78	.87	.85
Foster midparent-adopted child	154	.32	.31	.24	.38	.36	.33
Midparent-natural child	622	.29	.30	.31	.35	.47	.25

Note: Correlations based on scores adjusted for the effects of age and sex.
Source: Adapted from Bouchard (69).

in covariation between biological relatives in this sample.

The preceding would seem to suggest that the environmental effect on covariation between relatives for subcutaneous fatness may be less significant than is commonly believed, i.e., genetic variation contributes significantly to interindividual variation in fatness. Results from a comprehensive familial study of fatness based on six skinfolds and densitometry and fat distribution (48, 71, 72) suggest little or no genetic effect in the sum of six skinfolds, i.e., subcutaneous fatness. In contrast, there is a significant genetic effect in densitometrically estimated total fat mass, i.e., a combination of subcutaneous and internal fat (about 22%), in the ratio of extremity-to-trunk skinfold thicknesses, i.e., the distribution of subcutaneous fat (about 31%), and in the ratio of subcutaneous fat-to-fat mass (about 23%). Thus, a significant proportion of interindividual differences in fat mass and fat distribution is associated with genetic variation.

Factors associated with physical fitness

Many factors, biological and cultural, influence an individual's physiological and motor fitness. These include body size, physique, body composition, rate of biological maturation, socioeconomic status, lifestyle, child-rearing atmosphere, level of habitual physical activity, and so on (14, 45). These factors are not necessarily mutually exclusive. The association between physiological and motor fitness and size, physique and body composition is reasonably well established (73, 74), and the genotypic contribution to variation in morphological characteristics is rather well documented (14, 66). Three factors presumably more directly associated with physiological and motor fitness are considered subsequently: characteristics of muscle tissue, characteristics of the heart and circulatory system, and perceptual characteristics. The contribution of genetic effects to measurements of the blood and other body fluids and respiratory functions, which are presumably related to physical fitness, is beyond the scope of this overview (see 1). Similarly, a large number of studies have been published concerning familial resemblance and genetic variation in serum triglycerides, total cholesterol, and lipoprotein fractions (see 75).

Characteristics of muscle tissue per se, relationships with performance, and responses to training have been previously reviewed (14, 44, 66, 76). Muscle tissue is obviously a substrate of physiological and motor fitness, and can be influenced by training. Training effects, however, are specific to the type of program, e.g., endurance training, and strength training. Hence, genetic considerations in muscle tissue include at least two foci: first, genetic effects in muscle tissue characteristics, and second, genetic effects in response to specific training programs, i.e., the G × E effect.

Initial estimates of the genetic effect in muscle fiber distribution based on the traditional twin model indicated a high degree of heritability for Type I fibers, e.g., coefficients of 0.93 or higher for the percent of Type I fibers in the vastus lateralis (77). More recent estimates based on biological siblings and DZ and MZ twins, on the other hand, indicate almost total independence of the relative fiber distribution of the vastus lateralis from the genotype (78). Intraclass correlations for the percent of Type I fibers were 0.33 for siblings (32 pairs), 0.52 for DZ twins (26 pairs), and 0.55 for MZ twins (35 pairs). The intrapair resemblance in brothers and DZ twins suggests both an influence of heredity and of common environment because both kinds of relatives share 50% of their genes by descent, and the resemblance is higher in DZ twins. In contrast, the similar degree of resemblance in DZ and MZ twins suggests that common environmental conditions contribute more to the percent of Type I fibers in skeletal muscle than genetic factors. Similar trends are apparent for the percent of Type IIa and IIb fibers. Thus, if there is an effect of heredity in muscle fiber type distribution in humans, at least for the vastus lateralis muscle, it appears to be quite low.

A variety of muscle enzyme activities were also considered in the study of brothers, DZ twins, and MZ twins (78). The results suggest that environmental and nongenetic factors influence most enzyme activities with several exceptions. However, variation of regulatory enzymes of the glycolytic (phosphofructokinase) and citric acid cycle (oxoglutarate dehydrogenase) pathways and perhaps the ratio of oxidative to glycolytic activity show significant genetic effects that range between 25% and 50% of the total phenotypic variation (78). Small samples of MZ (11 pairs) and DZ (6 pairs) twins showed no gene-associated variation in the mitochondrial density, the ratio of mitochondrial to myofibril volumes, and the internal and external surface densities of mitochondria (79).

The issue of genetic effects in responses of muscle tissue to regular training, i.e., a genotype-training interaction (G × E effect), is not yet clear. Results of a 15-week, high-intensity training program of continuous and interval work patterns indicated small but significant changes in the proportions of Type I and IIb fibers, i.e., fiber type transformation (80). These observations are consistent with experimental data for animals that show changes in fiber type distribution with prolonged intensive training (76) and with clinical observations of athletes that show changes in fiber type distribution following prolonged immobilization (81).

The subjects in the study of responses to the high-intensity intermittent training program were 14 pairs of MZ twins (80). Variation in fiber type after training were randomly distributed among twins, thus indicating a negligible genetic effect. On the other hand, analysis of enzyme activities suggested a significant genetic effect for the response of most enzyme activities to high-intensity training.

In a related study of six pairs of MZ twins (82), no change in fiber type distribution was observed after 15 weeks of endurance training. However, responses of skeletal muscle enzyme activities varied with the duration of the training program. After 7 weeks of training, within pair similarities in responses of most enzyme activities were low, but after 15 weeks, the moderate-to-high intraclass correlations indicated significant genotype dependence of the responses of enzymes to the endurance-training program. Thus, as individuals approach maximal trainability, adaptation of skeletal muscle enzymes to the training stimulus is genotype dependent (82).

A 10-week isokinetic strength training program in five pairs of MZ twins resulted in significant increases in peak muscular torque output and several muscle tissue enzymes (83). With the exception of a significant genetic effect in the response of oxoglutarate dehydrogenase to training, there was no significant genotype-training interaction for peak torque output and other muscle tissue enzymes (creatine kinase, hexokinase, malate dehydrogenase).

Data for biological siblings and twins thus suggest that the genetic effect in muscle tissue fiber composition is quite low, but it is higher in several muscle tissue enzyme activities. There is a significant role for regular training, i.e., an environmental component, in modifying fiber distribution, size and metabolic capacity, and evidence for a G × E interaction in the adaptation to training. The G × E interaction effect apparently becomes more evident as maximal training responses are approached. Responses and estimated genetic effects vary with the type and duration of the training stimulus.

The role of genetic factors in heart size and function as well as circulatory function has been previously reviewed (14, 21, 66). Studies in humans indicate that biological inheritance plays a major role in the vascular wall thickness of the left coronary arteries in infants, in the anatomical pattern of the coronary arteries, and in the branching pattern of the pulmonary arteries. Echocardiographic data suggest the presence of significant familial resemblance in several left ventricular dimensions and an estimated genetic effect of about 50% of the total variance. Heart volume per unit body weight, on the other hand, has a rather low heritability, 25- 30%. Twin studies indicate a significant contribution of genotype to heart rate variability and the duration of the ECG P–R interval. On the other hand, heritability is virtually zero for the QRS and the Q–T intervals. For blood pressure, it is estimated that about 50–60% of the variation in systolic blood pressure in normotensive individuals can be attributed to a genetic effect, and 45% of the variation in diastolic blood pressure can be attributed to a genetic effect. Present estimates indicate no ethnic variation in the size of the total genetic effect in blood pressure.

Significant changes in echocardiographic cardiac dimensions occur in response to endurance training (84). The subjects of this study included 10 pairs of MZ twins, in addition to unrelated subjects. Posttraining measurements of cardiac dimensions indicated an increase in variance between twin pairs but a decrease in variance within twin pairs, which suggests genetic dependency in the adaptive responses to the endurance-training program.

Individual differences in perceptual characteristics such as spatial abilities, perceptual speed, perception of direction, and intersensory integration have a significant genetic component (85–88). The role of such perceptual characteristics in skillful motor performance and presumably in motor fitness, though often suggested (89), has not been fully investigated (14). The same is true for reaction time, another characteristic often indicated as important in motor performance. Results of twin studies, however, yield variable results for the genetic component of reaction time. For example, three estimates of heritability for reaction time to a light stimulus in school-age twins vary considerably, 0.22, 0.55, and 0.86; and one study of parent–offspring similarities in three reaction time measures indicates moderate correlations, with slightly higher values for parent–daughter pairs compared with parent–son pairs (44).

Assortative mating for physical fitness and fatness

Studies of biological relatives should allow for assortative mating, the tendency of individuals to select mates with similar characteristics. Positive assortative mating for ability, education, religion, and so on is reasonably well established (90–92), as is that for physical characteristics (93–95). Husband–wife correlations for body size and related measurements tend to be low in populations of primarily European ancestry. Nevertheless, assortative mating, and especially its genetic component, is of considerable significance because it will not only lead to an increase in the frequency of homozygotes for the relevant genetic loci but will also inflate the additive genetic variance for a given characteristic.

Husband–wife correlations for several measures of physiological and motor fitness are shown in table 15. Spouse correlations tend to be low to moderate; and many are in the same general range as those for physical characteristics, although those reported by Wolański (96) for strength are rather high. Spouse correlations for tasks requiring accuracy, speed, and precision of movement tend to be in the same range as those for aerobic performance and strength (44). The data thus indicate some degree of positive assortative mating for components of both physiological and motor fitness. Deviations from random mating or the effects of a common lifestyle, i.e., cohabitation, should, in turn, be considered in estimating the genetic effect in physical fitness.

Spouse similarities for subcutaneous fatness as indicated by the triceps and subscapular skinfolds, the two more commonly measured sites, are shown in table 16. Correlations are variable, and several are similar in magnitude to those for other anthropometric dimensions (93–95). Spouse correlations for other skinfolds are not extensive, but the study of Savard et al. (70)

Table 15. Spouse correlations on tests of physiological and motor fitness

Source or sample	Test	Correlation
Montoye and Gayle (20), 27 pairs	$\dot{V}O_2$max, measured or estimated, adjusted for age, weight and fatness	.18
Lortie et al. (23), 119 pairs	$\dot{V}O_2$max./kg, estimated	.33
Lesage et al. (113), 20 pairs	$\dot{V}O_2$max/kg, measured	.21
Bouchard (21), 276 pairs	PWC_{150}/kg	.19[a]
Wolański (96), number of pairs vary between 36 and 72	Strength:	
	right grip	.74
	left grip	.77
	shoulder	.53
	back	.44
	Balance:	
	beam walk	−.05
	turning, timed	.25
	turning, number of turns	.04
Kovar (34), 60 pairs	Strength:	
	grip	.26
	back	.26
Szopa (37), 347 families	Strength:	
	right grip	.15
	left grip	.26
	grip/weight	.23
	arm	.17
	arm/weight	.17
	Motor: vertical jump	.35
Devor and Crawford (42), 53 pairs	Strength: dominant grip	.01[b]
	Trunk flexibility	−.09[b]

[a] Intraclass correlation based on analysis of variance.
[b] Correlations based on standardized residuals.

reported low correlations for the following skinfolds: −0.03 and 0.08 for the biceps, 0.00 and −0.01 for the suprailiac, −0.04 and 0.01 for the abdominal, 0.17 and 0.24 for the calf, and 0.10 and −0.01 for the sum of six skinfolds. In each instance, the first correlation given is age and sex standardized; and the second correlation includes statistical control for age, sex, energy intake, energy expenditure, and socioeconomic status.

Activity level and level of habitual physical activity

Studies of temperament and personality often indicate an activity factor that has a significant genetic component. For example, Frischeisen-Kohler (97, p. 302) reported a significant genetic/familial contribution to variation in an individual's "personal tempo, which is

Table 16. Spouse correlations for the triceps and subscapular skinfold thicknesses

	Triceps		Subscapular	
Source or sample	*n* pairs	*r*	*n* pairs	*r*
Garn et al. (114), Ten State Nutrition Survey,[a] age:				
3rd decade	425	.28	283	.24
4th decade	451	.27	320	.25
Tecumseh Project,[a] age:				
4th decade	276	.16	276	.24
Garn et al. (115), Ten State Nutrition Survey [b]	1803	.25		
Tecumseh Project [b]	1946	.07		
Savard et al. (70), French Canadians	228	.12[c]	228	−.05[c]
		.14[d]		.22[d]
Bouchard (69), French Canadians	348	.08[c]	348	.06[c]
Malina et al. (103), San Antonio Heart Study,[e]				
Anglos	297	.15	297	.24
Mexican Americans	378	.16	378	.16

[a] Correlations are based on normalized age-specific T scores.
[b] Mean correlations from mean Z-transforms of correlations.
[c] Correlations standardized for age and sex.
[d] Correlations standardized for age and sex, and after controlling for energy intake, energy expenditure, and socioeconomic status.
[e] Correlations controlled for age and reported level of physical activity.

expressed more or less markedly in all of our doings." This conclusion was based on an early observational study of same-sex twins, siblings, parents, and offspring using a series of tapping and metronome experiments. Among 6- to 10-year-old MZ and DZ twins, Scarr (98) noted a moderate genetic influence (estimated heritability of 0.4) on activity motivation that includes such measures as reaction time, preference for physical activity, anxiety, and impatience. Inouye (99) noted a similar trend in a large series of Japanese twins, i.e., greater concordance for bodily motion and activity among MZ than among DZ twins. Similarity in bodily movements and activity level also persisted in MZ twins who were reared apart. More recently, Buss and Plomin (100) defined an activity component in their temperament theory of personality development. Activity level refers to total energy output, i.e.,

> . . . an active person moves around more, tends to be in motion, hurries more than others, and keeps busier than those around him. . . . His preferences in sports reflect his need for pulsating activity and a driving tempo. He likes tennis, handball, and squash best; these are followed by football, basketball, and volleyball; lower on the list are baseball, skating, swimming, and canoeing. Note that the best-liked sports involve not only a huge energy output but also bursts of vigorous exertion. Note also the sheer range of energy-depleting activity. (100, pp. 30–31)

Activity level is viewed in terms of vigor and tempo, which are perhaps alternative ways of expending energy. Twin studies indicate a definite genetic influence on activity level and moderate evidence for stability through childhood into adulthood (see 100, 101).

Activity level as defined in studies of temperament and personality is not the same as the pattern of habitual physical activity, as commonly understood in studies of physical fitness. Nevertheless, an individual's activity temperament may be a significant determinant of physical activity interests and pursuits, which in turn influence energy expenditure. Specific evidence to this effect, however, is not presently available. Similarly, evidence for a significant genotypic contribution to variance in habitual physical activity, i.e., energy expenditure, is limited. Evidence from a Beta path analysis of resemblance between various pairs of biological and nonbiological relatives in overall energy expenditure based on a 3-day activity record indicates a transmissibility effect of about 30% between generations, and less than one-half of the transmissible effect is genetic (102). The pattern of high-intensity activity shows little transmissibility between generations but a fairly high level of resemblance within generation. The latter thus suggests a significant role for common environment and cohabitation effects.

Spouse correlations for estimates of overall energy expenditure, 0.24 and 0.26 in the French Canadian familial study, are reasonably similar to those for maximal aerobic power and strength (table 15), but the correlation for the pattern of intense physical activity is lower, 0.13 (102). Husband–wife correlations for reported levels of physical activity among Anglos and Mexican Americans in the San Antonio Heart Study are 0.13 and 0.20, respectively, and show a tendency to decrease as socioeconomic status increases (103).

Design implications and recommendations

It is assumed at this point that the assessment of physical fitness, fatness, and activity patterns as well as other relevant lifestyle parameters can be undertaken within the scope of a health examination survey of the U.S. population. In turn, it is recommended that several well-defined objectives relevant to the genetic component of physical fitness and activity patterns be included in such a survey. Particular attention should be given to the following objectives:

A. Determination of familial resemblance (family line effect) in physical fitness and activity patterns
B. Determination of the degree of assortative mating in physical fitness and activity patterns
C. Determination of whether there is a maternal and/or a paternal effect in physical fitness and activity patterns
D. Determination of whether there is a parental effect that is associated with children of one sex only in physical fitness and activity patterns
E. Determination of the level of resemblance in physical fitness and activity patterns among relatives sharing genes by descent and living together
F. Determination of the level of resemblance in physical fitness and activity patterns among relatives by adoption and living together
G. Determination of the level of resemblance in physical fitness and activity patterns among various types of relatives as a function of the duration of cohabitation
H. Determination of the level of resemblance in physical fitness among relatives with similar and dissimilar patterns of physical activity
I. Determination of the transmissibility effect in physical fitness and activity patterns between generations
J. Determination of the genetic and cultural (i.e., nongenetic) components of transmissibility for physical fitness and activity patterns
K. Determination of whether a major gene effect exists in physical fitness.

It should be noted that in a survey situation, it is almost impossible to obtain a direct assessment of the G × E component as it requires an experimental study with rather stringent conditions.

The attainment of the above-stated objectives within the context of a health examination survey has significant design implications. It is assumed that the various physical fitness components can be measured with acceptable validity and reliability in the survey situation. It is also assumed that activity patterns, perhaps habitual energy expenditure, and other relevant lifestyle indicators can be quantified in a valid and reliable manner. An additional assumption is that appropriate concomitant variables, e.g., chronological age, body size, fatness, and socioeconomic status, will be measured and be available for the proper statistical adjustment of the dependent variables, physical fitness, and activity patterns.

The main consequence of an attempt to assess the genetic effects in physical fitness and activity patterns for sampling procedures is that the household must be the sampling unit rather than the individual. The household sampling strategy would permit the recovery of various pairs of biological and cultural relatives required to meet the above-stated objectives. Ideally, the household units would be selected at random in order to adequately represent the various geographic areas of the country, socioeconomic strata, and ethnic groups.

If such a sampling strategy is used, the following types of relatives will be found in a large series of households:

A. Spouses
B. Brothers and sisters
C. Natural parents and their offspring
D. DZ and MZ twins (an estimated frequency is one pair of twins in approximately 100 households)
E. Foster parents and adopted children (approximately 1% of the child population of the United States in nonrelative adopted children)
F. Siblings by adoption
G. Occasionally, grandparent, parent, and child combinations
H. Occasionally, uncle/aunt, nephew/niece, and possibly other relatives (cousins, half-sibs, etc.).

The minimum set of data required to develop a meaningful estimate of the genetic component of physical fitness and activity patterns must include the types of relatives in A through F above in sufficient numbers. The sample size of 20,000 planned for the national survey would translate into a household sample size of approximately 7,000. This should be adequate for relatives indicated in A, B, and C above, but probably not sufficient for relatives indicated in D and definitely not sufficient to provide sample sizes large enough for the other types of relatives (E through H). A sample size between 35,000 and 50,000 households would have to be considered in order to meet all the objectives described earlier. These are, however, only approximate estimates.

Genetic epidemiologists have developed the analytical tools required to deal with a data base such as that suggested in this report. These procedures imply the study of score distributions, adjustment for concomitant variables, correlation and regression analysis between pairs of relatives, analysis of variance for the between and within mean squares in various sets of relatives, path analysis for biological and cultural transmissibility effects in the presence of other effects, commingling analysis for the presence of more than one distribution, segregation analysis, and others. In most instances, statistical packages are available for these analyses. The interpretation of the results from such a comprehensive analysis will undoubtedly require the close collaboration of geneticists and human biologists with an interest in physical fitness, among others, as the issues are complex.

References

1. Bouchard, C., and Malina, R.M.: Genetics for the sport scientist: Selected methodological considerations. *Exercise and Sport Sciences Reviews* 11:275–305, 1983.
2. World Health Organization: Optimum physical performance capacity in adults. World Health Organization, Technical Report Series, No. 436, 1969.
3. Cureton, T.K.: *Physical Fitness Appraisal and Guidance* St. Louis: Mosby, 1947.
4. Larson, L.A., and Yocom, R.D. 1951: *Measurement and Evaluation in Physical, Health and Recreation Education.* St. Louis: Mosby.
5. AAHPERD: *Health Related Physical Fitness Test Manual.* American Alliance for Health, Physical Education, Recreation and Dance, Reston, Va., 1980.
6. Falconer, D.S.: *Introduction to Quantitative Genetics.* New York: Ronald Press, 1960.
7. Cavalli-Sforza, L.L., and Bodmer, W.F.: *The Genetics of Human Populations.* San Francisco: Freeman, 1971.
8. Mather, K., and Jinks, J.L.: *Biometrical Genetics.* Ithaca, N.Y.: Cornell University Press, 1971.
9. Meredith, W.: A model for analyzing heritability in the presence of correlated genetic and environmental effects. *Behavioral Genetics* 3:271–277, 1973.
10. Plomin, R., DeFries, J.C., and Loehlin, J.C.: Genotype–environment interaction in the analysis of human behavior. *Psychological Bulletin* 84:309–322, 1977.
11. Christian, J.C.: Testing twin means and estimating genetic variance. Basic methodology for the analysis of quantitative twin data. *Acta Geneticae Medicae et Gemellologiae* 28:35–40, 1979.
12. Klissouras, V.: Heritability of adaptive variation. *Journal of Applied Physiology* 31:338–344, 1971.
13. Klissouras, V., Pirnay, F., and Petit, J.M.: Adaptation to maximal effort: Genetics and age. *Journal of Applied Physiology* 35:288–293, 1973.
14. Bouchard, C., and Malina, R.M.: Genetics of physiological fitness and motor performance. *Exercise and Sport Sciences Reviews* 11:306–339, 1983.
15. Engström, L.M., and Fischbein, S.: Physical capacity in twins. *Acta Geneticae Medicae et Gemellologiae* 26:159–165, 1977.
16. Bouchard, C., Lortie, G., Simoneau, J–A., Leblanc, C., Thériault, G., and Tremblay, A.: Submaximal power output in adopted and biological siblings. *Annals of Human Biology* 11:303–309, 1984.
17. Bouchard, C., Lesage, R., Lortie, G., Simoneau, J–A., Hamel, P., Boulay, M.R., Pérusse, L., Thériault, G., and Leblanc, C.: Aerobic performance in brothers, dizygotic and monozygotic twins. *Medicine and Science in Sports and Exercise* 18:639–646, 1986.

18. Weber, G., Kartodihardjo, W., and Klissouras, V.: Growth and physical training with reference to heredity. *Journal of Applied Physiology* 40:211–215, 1976.
19. Montoye, H.J., Metzner, H.L., and Keller, J.K.: Familial aggregation of strength and heart rate response to exercise. *Human Biology* 47:17–36, 1975.
20. Montoye, H. J., and Gayle, R.: Familial relationships in maximal oxygen uptake. *Human Biology* 50:241–249, 1978.
21. Bouchard, C.: Genetics of aerobic power and capacity. In R.M. Malina and C. Bouchard eds., *Sport and Human Genetics* Champaign, Ill.: Human Kinetics Publishers, pp. 59–88, 1986.
22. Giles, P.E., Blanc, H., Cann, H.M., and Wallace, D.C.: Maternal inheritance of human mitochondrial DNA. Proceedings of the National Academy of Sciences, U.S.A. 77:6715–6719, 1980.
23. Lortie, G., Bouchard, C., Leblanc, C., Tremblay, C., Simoneau, J–A., Thériault, G., and Savoie, J.P.: Familial similarity in aerobic power. *Human Biology* 54:801–812, 1982.
24. Klissouras, V.: Prediction of potential performance with special reference to heredity. *Journal of Sports Medicine and Physical Fitness* 13:100–107, 1973.
25. Weiss, V.: Die Heritabilitaten sportlicher Tests, berechnet aus den Leistungen zehnjahriger Zwillingspaare. *Arztliche Jugendkunde* 68:167–172, 1977.
26. Lortie, G., Simoneau, J–A., Hamel, P., Boulay, M.R., Landry, F., and Bouchard, C.: Responses of maximal aerobic power and capacity to aerobic training. *International Journal of Sports Medicine* 5:232–236, 1984.
27. Bouchard, C., Carrier, R., Boulay, M.R., Thibault-Poirier, M–C., and Dulac, S.: Le Développement du Système de Transport de l'Oxygène chez les Jeunes Adultes. *Éditions du Pélican* Québec, 1975.
28. Bouchard, C., Boulay, M.R., Thibault, M–C., Carrier, R., and Dulac, S.: Training of submaximal working capacity: frequency, intensity, duration and their interactions. *Journal of Sports Medicine and Physical Fitness* 20:29–40, 1980.
29. Pollock, M.L.: The quantification of endurance training programs. *Exercise and Sport Sciences Reviews* 1:155–188, 1973.
30. Saltin, B.: The effect of physical training on the oxygen transporting system in man. In G. R. Cumming, D. Snidal, and A. W. Taylor eds., *Environmental Effects on Work Performance.* Canadian Association of Sport Sciences, Edmonton, 1972. pp. 151–162.
31. Klissouras, V.: Prediction of athletic performance: Genetic considerations. *Canadian Journal of Applied Sport Sciences* 1:195–200, 1976.
32. Prud'homme, D., Bouchard, C., Leblanc, C., Landry, F., and Fontaine, E.: Sensitivity of maximal aerobic power to training is genotype dependent. *Medicine and Science in Sports and Exercise* 16:489–493, 1984.
33. Bouchard, C. Human adaptability may have a genetic basis. In F. Landry ed., *Health Risk Estimation, Risk Reduction and Health Promotion.* Canadian Public Health Association, Ottawa, 1983. pp. 463–476.
34. Kovař, R.: *Human Variation in Motor Abilities and Its Genetic Analysis.* Charles University, Prague, 1981.
35. Komi, P. V., Klissouras, V., and Karvinen, E.: Genetic variation in neuromuscular performance. *Internationale Zeitschrift fur angewandte Physiologie* 31:289–304, 1973.
36. Wolański, N., and Kasprzak, E.: Similarity in some physiological, biochemical and psychomotor traits between parents and 2–45 year old offspring. *Studies in Human Ecology* 3:85–131, 1979.
37. Szopa, J.: Familial studies on genetic determination of some manifestations of muscular strength in man. *Genetica Polonica* 23:65–79, 1982.
38. Malina, R.M., and Mueller, W.H.: Genetic and environmental influences on the strength and motor performance of Philadelphia school children. *Human Biology* 53:163–179, 1981.
39. Cratty, B.J.: A comparison of fathers and sons in physical ability. *Research Quarterly* 31:12–15, 1960.
40. Sklad, M.: Similarity of movement in twins. *Wychowanie Fizycznie i Sport* 16:119–141, 1972.
41. Kovař, R.: Prispevek ke studiu geneticke podminenosti lidske motoriky. Doctoral dissertation, Charles University, Prague, 1974.
42. Devor, E.J., and Crawford, M.H.: Family resemblance for neuromuscular performance in a Kansas Mennonite community. *American Journal of Physical Anthropology* 64:289–296, 1984.
43. Sklad, M.: Rozwój fizyczny i motorycznośc bliźniąt. *Materialy i Prace Antropologiczne* 85:3–102, 1973.
44. Malina, R.M.: Genetics of motor development and performance. In R.M. Malina and C. Bouchard eds., *Sport and Human Genetics.* Champaign, Ill.: Human Kinetics Publishers, pp. 23–58, 1986.
45. Malina, R.M.: Quantification of fat, muscle and bone in man. *Clinical Orthopaedics and Related Research* 65:9–38, 1969.
46. Malina, R.M.: The measurement of body composition. In F.E. Johnston, A.F. Roche, and C. Susanne eds., *Human Physical Growth and Maturation: Methodologies and Factors.* New York: Plenum, 1980. pp. 35–59.
47. Roche, A.F. ed.: Body-Composition Assessments in Youth and Adults. Ross Laboratories, Columbus, Ohio, 1985.
48. Bouchard, C., Savard, R., Després, J–P., Tremblay, A., and Leblanc, C.: Body composition in adopted and biological siblings. *Human Biology* 57:61–75, 1985.
49. Sjöström, L., Kvist, H., and Tylen, U.: Methodological aspects of measurements of adipose tissue distribution. In J. Vague, P. Bjorntorp, B. Guy-Grand, M. Rebuffe-Scrive, and P. Vague eds., *Metabolic Complications of Human Obesities.* Excerpta Medica, Amsterdam, 1985. pp. 13–19.
50. Garn, S.M.: Relative fat patterning: An individual characteristic. *Human Biology* 27:75–98, 1955.
51. Mueller, W.H.: The genetics of human fatness. *Yearbook of Physical Anthropology* 26:215–230, 1983.
52. Malina, R.M., Mueller, W.H., Bouchard, C., Shoup, R.F., and Lariviere, G.: Fatness and fat patterning among athletes at the Montreal Olympic Games, 1976. *Medicine and Science in Sports and Exercise* 14:445–452, 1982.
53. Malina, R.M., Little, B.B., Stern, M.P., Gaskill, S.P., and Hazuda, H.P.: Ethnic and social class differences in selected anthropometric characteristics of Mexican American and Anglo adults: The San Antonio Heart Study. *Human Biology* 55:867–883, 1983.
54. Mueller, W.H., Shoup, R.F., and Malina, R.M.: Fat patterning in athletes in relation to ethnic origin and sport. *Annals of Human Biology* 9:371–376, 1982.
55. Mueller, W.H., Joos, S.K., Hanis, C.L., Zavaleta, A.N., Eichner, J., and Schull, W.J.: The Diabetes Alert Study: Growth, fatness and fat patterning, adolescence through adulthood in Mexican Americans. *American Journal of Physical Anthropology* 64:389–399, 1984.
56. Robson, J.R.K., Bazin, M., and Soderstrom, R.: Ethnic differences in skinfold thickness. *American Journal of Clinical Nutrition* 29:864–868, 1971.
57. Després, J–P., Bouchard, C., Tremblay, A., Savard, R., and Marcotte, M.: Effects of aerobic training on fat distribution in male subjects. *Medicine and Science in Sports and Exercise* 17:133–118, 1985.
58. Frisancho, A.R., and Flegel, P.N.: Advanced maturation associated with centripetal fat pattern. *Human Biology* 54:717–727, 1982.
59. Deutsch, M.I., Mueller, W.H., and Malina, R.M.: Androgyny in fat patterning is associated with obesity in adolescents and young adults. *Annals of Human Biology* 12:275–286, 1985.
60. Baumgartner, R.N., Roche, A.F., Guo, S., Lohman, T., Boileau, R. A., and Slaughter, M.A.: Adipose tissue distribution: The stability of principal components by sex, ethnicity and maturation stage. *Human Biology* 58:719–735, 1987.
61. Bogin, B., and MacVean, R.B.: Nutritional and biological determinants of body fat patterning in urban Guatemalan children. *Human Biology* 53:259–268, 1981.

62. Ramirez, M.E., and Mueller, W.H.: The development of obesity and fat patterning in Tokelau children. *Human Biology* 52:675–687, 1980.
63. Vague, J.: The degree of masculine differentiation of obesities: a factor determining predisposition to diabetes, atherosclerosis, gout, and uric calculous disease. *American Journal of Clinical Nutrition* 4:20–34, 1956.
64. Kissebah, A.H., Evans, D.J., Peiris, A., and Wilson, C.R.: Endocrine characteristics in regional obesities: Role of sex steroids. In J. Vague, P. Bjorntorp, B. Guy-Grand, M. Rebuffe-Scrive, and P. Vague eds., *Metabolic Complications of Human Obesities.* Excerpta Medica, Amsterdam, 1985. pp. 115–130.
65. Björntorp, P.: Fat patterning and disease: A review. In N.G. Norgan ed., *Human Body Composition and Fat Distribution* (Euro-Nut Report 8). Stichting Nederlands Instituut voor de Voeding, Wageningen, 1986. pp. 201–209.
66. Bouchard, C., and Lortie, G.: Heredity and endurance training. *Sports Medicine* 1:38–64, 1984.
67. Hawk, L.J., and Brook, C.G.D.: Family resemblances of height, weight, and body fatness. *Archives of Disease of Childhood* 54:877–879, 1979.
68. Roche, A.F., Siervogel, R.M., Chumlea, W.C., Reed, R.B., Valadian, I., Eichorn, D., and McCammon, R.W.: *Serial Changes in Subcutaneous Fat Thicknesses of Children and Adults.* Karger, Basel, 1982.
69. Bouchard, C.: Genetics of body fat, energy expenditure and adipose tissue metabolism. In E.M. Berry, S.H. Blondheim, H.E. Eliahou, and E. Shafrir eds., *Recent Advances in Obesity Research:* V. Libbey, London, 1987. In press.
70. Savard, R., Bouchard, C., Leblanc, C., and Tremblay, A.: Familial resemblance in fatness indicators. *Annals of Human Biology* 10:111–118, 1983.
71. Bouchard, C.: Inheritance of fat distribution and adipose tissue metabolism. In J. Vague, P. Bjorntorp, B. Guy-Grand, M. Rebuffe-Scrive, and P. Vague eds., *Metabolic Complications of Human Obesities.* Excerpta Medica, Amsterdam, 1985. pp. 87–96.
72. Bouchard, C., and Tremblay, A.: Genetics of body composition and fat distribution. In N.G. Norgan ed., *Human Body Composition and Fat Distribution* (Euro-Nut Report 8). Stichting Nederlands Instituut voor de Voeding, Wageningen, 1986. pp. 175–188.
73. Malina, R.M.: Anthropometric correlates of strength and motor performance. *Exercise and Sport Sciences Reviews* 3:249–274, 1975.
74. Boileau, R.A., and Lohman, T.G.: The measurement of human physique and its effect on physical performance. *Orthopedic Clinics of North America* 8:563–581, 1977.
75. Sing, C.F., and Skolnick, M. eds.: Genetic analysis of common diseases: Applications to predictive factors in coronary disease. *Progress in Clinical and Biological Research,* Vol. 32, 1979.
76. Malina, R.M.: Growth of muscle tissue and muscle mass. In F. Falkner and J.M. Tanner eds., *Human Growth.* Vol. 2. *Postnatal Growth and Neurobiology.* Plenum, New York, 1986. pp. 77–99.
77. Komi, P.V., Viitasalo, J.H.T., Havu, M., Thorstensson, A., Sjödin, B., and Karlsson, J.: Skeletal muscle fibres and muscle enzyme activities in monozygous and dizygous twins of both sexes. *Acta Physiologica Scandinavica* 100:385–392, 1977.
78. Bouchard, C., Simoneau, J–A., Lortie, G., Boulay, M.R., Marcotte, M., and Thibault, M–C.: Genetic effects in human skeletal muscle fiber type distribution and enzyme activities. *Canadian Journal of Physiology and Pharmacology* 64:1245–1251, 1986.
79. Howald, H.: Ultrastructure and biochemical function of skeletal muscle in twins. *Annals of Human Biology* 3:455–462, 1976.
80. Simoneau, J–A., Lortie, G., Boulay, M.R., Marcotte, M., Thibault, M–C., and Bouchard, C.: Inheritance of human skeletal muscle and anaerobic capacity adaptation to high-intensity intermittent training. *International Journal of Sports Medicine* 7:167–171, 1986.
81. Howald, H.: Training induced morphological and functional changes in skeletal muscle. *International Journal of Sports Medicine* 3:1–12, 1982.
82. Hamel, P., Simoneau, J–A., Lortie, G., Boulay, M.R., and Bouchard, C.: Heredity and muscle adaptation to endurance training. *Medicine and Science in Sports and Exercise* 18:690–696, 1986.
83. Thibault, M–C., Simoneau, J–A., Côte, C., Boulay, M.R., Lagassé, P., Marcotte, M., and Bouchard, C.: Inheritance of human muscle adaptation to isokinetic strength training. *Human Heredity* 36:341–347, 1986.
84. Landry, F., Bouchard, C., and Dumesnil, J.: Cardiac dimension changes with endurance training. *Journal of the American Medical Association* 254:77–80, 1985.
85. Vandenberg, S.G.: The hereditary abilities study: hereditary components in a psychological test battery. *American Journal of Human Genetics* 14:220–237, 1962.
86. Vandenberg, S.G.: Contributions of twin research to psychology. *Psychological Bulletin* 66:327–352, 1966.
87. McGee, M.G.: Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin* 86:889–918, 1979.
88. Rose, R.J., Miller, J.Z., Dumon-Driscoll, M., and Evans, M.M.: Twin-family studies of perceptual speed ability. *Behavioral Genetics* 9:71–86, 1979.
89. Williams, H.G.: *Perceptual and Motor Development.* Englewood Cliffs, N.J.: Prentice-Hall, 1983.
90. Garrison, R.J., Anderson, V.E., and Reed, S.C.: Assortative marriage. *Eugenics Quarterly* 15:113–127, 1968.
91. Harrison, G.A., Givson, J.B., and Hiorns, R.W.: Assortative marriage for psychometric, personality, and anthropometric variation in a group of Oxfordshire villages. *Journal of Biosocial Science* 8:145–153, 1976.
92. Johnson, R.C., DeFries, J.C., Wilson, L.R., McClearn, G.E., Vandenberg, S.G., Ashton, G.C., Mi, M.P., and Rashad, M.N.: Assortative marriage for specific cognitive abilities in two ethnic groups. *Human Biology* 48:343–352, 1976.
93. Susanne, C.: Contribution à l'étude de l'assortiment matrimonial dans un échantillon de la population Belge. *Bulletin de la Societe Royale Belge d'Anthropologie et de Préhistoire* 78:147–196, 1967.
94. Spuhler, J.N.: Assortative mating with respect to physical characteristics. *Eugenics Quarterly* 15:128–140, 1968.
95. Roberts, D.F.: Assortative mating in man: husband/wife correlations in physical characteristics. *Bulletin of the Eugenics Society,* Suppl. 2, 1977.
96. Wolański, N.: Assortative mating in the Polish rural populations. *Studies in Human Ecology* 1:182–188, 1973.
97. Frischeisen-Kohler, I.: The personal tempo and its inheritance. *Character and Personality* 1:301–313, 1933.
98. Scarr, S.: Genetic factors in activity motivation. *Child Development* 37:663–673, 1966.
99. Inouye, E.: Twin studies and human behavioral genetics. *Jinrui Idengaku Zasshi* 15:1–25, 1970.
100. Buss, A.H., and Plomin, R.: *A Temperament Theory of Personality Development.* New York: Wiley, 1975.
101. Buss, A.H., and Plomin, R.: Temperament: Early Developing Personality Traits. Erlbaum Associates, Hillsdale, New Jersey, 1984.
102. Bouchard, C., unpublished data.
103. Malina, R.M., Little, B.B., Stern, M.P., Gaskill, S.P., and Hazuda, H.P., unpublished data.
104. Komi, K.V., and Karlsson, J.: Physical performance, skeletal muscle enzyme activities, and fibre types in monozygous and dizygous twins of both sexes. *Acta Physiologica Scandinavica,* Supplement 462, 1–28, 1979.
105. Venerando, A., and Milani-Comparetti, M.: Twin studies in sport and physical performance. *Acta Geneticae Medicae et Gemellologiae* 19:80–82, 1972.

106. Williams, L.R.T., and Hearfield, V.: Heritability of a gross motor balance task. *Research Quarterly* 44:109–112, 1973.
107. Williams, L.R.T., and Gross, J.B.: Heritability of motor skills. *Acta Geneticae Medicae et Gemellologiae* 29:127–136, 1980.
108. McNemar, Q.: Twin resemblances in motor skills, and the effect of practice thereon. *Pedagogical Seminary and Journal of Genetic Psychology* 42:70–99, 1933.
109. Brody, D.: Twin resemblances in mechanical ability, with reference to the effects of practice on performance. *Child Development* 8:207–216, 1937.
110. Wilde, G.J.S.: An experimental study of mutual behaviour imitation and person perception in MZ and DZ twins. *Acta Geneticae Medicae et Gemellologiae* 19:273–279, 1970.
111. Sklad, M.: The genetic determination of the rate of learning motor skills. *Studies in Physical Anthropology* 1:3–19, 1975.
112. Marisi, D.Q.: Genetic and extragenetic variance in motor performance. *Acta Geneticae Medicae et Gemellologiae* 26:197–204, 1977.
113. Lesage, R., Simoneau, J-A., Jobin, J., Leblanc, J., and Bouchard, C.: Familial resemblance in maximal heart rate, blood lactate and aerobic power. *Human Heredity* 35:182–189, 1985.
114. Garn, S.M., Clark, D.C., and Ullman, B.M.: Does obesity have a genetic basis in man? *Ecology of Food and Nutrition* 4:57–60, 1975.
115. Garn, S.M., Cole, P.E., and Bailey, S.M.: Living together as a factor in family-line resemblances. *Human Biology* 51:565–587, 1979.

Chapter **20**

Biochemical Correlates of Fitness and Exercise

William L. Haskell, Ph.D.
Stanford University

Introduction

Physical activity and physical fitness are complex entities, difficult to accurately and reliably measure in population-based surveys. Physical activity has nearly unlimited permutations of type, intensity, duration and frequency; whereas physical fitness has multiple dimensions consisting of such attributes as endurance, flexibility, strength, power, speed, and agility. In large-scale surveys physical activity has usually been evaluated by self- or interviewer-administered questionnaires or diaries; and in smaller scale projects, motion sensors and continuous recording of heart rate have been used. Fitness has usually been defined in health-oriented surveys as endurance or aerobic capacity, and it has been estimated by recording heart-rate response to submaximal exercise or exercise intensity during maximal exercise. Many of the advantages and limitations of these approaches to measuring physical activity and physical fitness have been reviewed in other chapters of these proceedings (see chapters 4–6 and 24).

During vigorous physical activity profound physiologic and biochemical changes occur and may last for some hours or even days. Repeated bouts of activity, or exercise training, lead to marked changes in both function and structure of various tissue and organs. It would seem reasonable then that the recent activity status of a person or the impact of that activity on the individual's physical fitness could be determined with some high degree of accuracy and reliability by measuring a "biochemical marker" of exercise or fitness. If such a marker could be identified that was valid, reliable, and easily obtainable at a reasonable cost, it would be extremely valuable either as a standard for validating other measures of physical activity or fitness or as a primary measure in population surveys.

Search for a biochemical marker

Many acute responses and chronic adaptations to various exercise regimens have systematically been investigated and delineated. Although a great deal is yet to be determined regarding the breadth and magnitude of these responses and the biologic mechanisms controlling them, enough is known to consider the potential for selected measures to be used as biochemical markers for physical activity or physical fitness status. Very few biochemical responses to exercise have been measured in population-based samples. Most data on biochemical adaptations to exercise have been derived from cross-sectional studies, usually comparing highly trained athletes and secondary controls, or from relatively short-term longitudinal studies, evaluating subjects before and after exercise training. Measurements have been made with subjects at rest, during submaximal exercise, maximal exercise, or a recovery period following exercise. Most data on acute responses to exercise have been collected either during or soon after the completion of exercise (less than 6 hours) with relatively few measurements made over days or weeks.

Because of the diverse nature of biochemical responses that occur during exercise or of alterations that result from vigorous exercise training, a potential biochemical marker for exercise status or fitness might be obtained using any one of various bodily fluids or tissues. From the standpoint of convenience, a marker substance available in saliva or urine would be preferable to one in blood or feces; and if a tissue sample, other than skin, was needed, the marker would probably have limited usefulness in population-based surveys. Also, in the future, it may be possible to measure biochemical status noninvasively without actually removing a tissue sample from the body, by using such techniques as positron emission tomography (PET) or magnetic resonance imaging (MRI). However, both of these techniques are still in the developmental stage and their use for assessing activity status or fitness is probably at least a decade away.

No biochemical marker has generally been accepted as an index of habitual activity status or physical fitness in humans. The major deterrents to the designation of such a marker have been the limited research specifically implemented for this purpose, the diverse inherited inter-individual variability in biochemical responses to exercise, and the confounding influences of other environmental factors or personal traits on an individual's acute or chronic responses to exercise.

The generally accepted indexes of various components of physical fitness include measures of performance (e.g., exercise duration, weight lifted, power applied) or physiologic responses to exercise (e.g., heart rate at a specified submaximal workload or measured/ estimated maximal oxygen uptake). For many of these measures data are available that characterize the response in selected samples and are potentially feasible for use in general population surveys designed to ascertain the relationship of selected fitness components and health status. Measures of exercise status in population-based surveys have been limited to self-reporting by questionnaire or diary, with data on body movement determined by motion sensors, or heart rate recorded in relatively small samples in observational or experimental studies.

Biochemical correlates of exercise and physical fitness

Data have been published demonstrating significant relationships between selected biochemical measures and either endurance exercise performance or maximal oxygen uptake as measures of physical fitness or physical activity status as determined by questionnaire or training logs. Biochemical measures shown to have such relationships vary from muscle fiber composition (1) and metabolic enzyme activity of skeletal muscle (2) to

Table 1. Biochemical measures that are potential markers for physical activity or endurance fitness status

Variable	Association with physical activity or endurance fitness
Urine	
CO_2 production by use of doubly labeled water isotopes ($^2H_2{}^{18}O$)	Provides estimate of energy expenditure
Catecholamine concentration	Increases following vigorous exercise
Blood	
Triglyceride	Lower with endurance exercise training
High-density lipoprotein cholesterol	Higher with endurance exercise training
Apolipoprotein A–I	Higher with endurance exercise training
Lactic acid concentration following exercise	Lower with increased endurance fitness
Catecholamine concentration following exercise	Lower with increased endurance fitness
Glucogon	Higher after vigorous exercise
Cortisol	Higher after vigorous exercise
Growth hormone	Higher after vigorous exercise
Testosterone	Higher after vigorous exercise
Insulin	Decrease response to and glucose load with exercise training
Skeletal muscle	
Myoglobin content	Higher with endurance exercise training
Mitochrondial mass	Higher with endurance exercise training
Succinate dehydrogenase activity	Higher with endurance exercise training
Citrase synthase activity	Higher with endurance exercise training
Citrochrome oxidase activity	Higher with endurance exercise training
Alanine transaminase activity	Higher with endurance exercise training
Aspartate transaminase activity	Higher with endurance exercise training
Hexokinase activity	Higher with endurance exercise training
Lipoprotein lipase activity	Higher with endurance exercise training
Glycogen concentration	Higher with endurance exercise training
Capillary density	Higher with endurance exercise training

high-density lipoprotein cholesterol (3) and sex hormone concentrations (4) in plasma. Table 1 contains a listing of selected biochemical measures that have been shown to be related to endurance exercise or fitness status. This list is not meant to be comprehensive. The correlations for these relationships generally tend to be higher for fitness than exercise measures, probably due, at least in part, to an attenuation of the correlation for exercise by the imprecision of its measurement and the greater influence of heredity on some fitness relationships. That is, genetic factors that partly determine the value of the biochemical measure also influence the fitness measure in either a causal or noncausal manner.

An example of the biochemical changes that occur in skeletal muscle in response to endurance exercise is the enzyme succinate dehydrogenase (SDH). In cross-sectional studies it has been demonstrated that relatively sedentary adults have SDH activity in leg muscles of approximately 6–7 mM/kg·min (5), and highly trained endurance athletes have SDH activity of 20–25 mM/kg·min (6). Also, immobilization of the leg in a cast for 4 weeks decreases the SDH activity by half, and vigorous retraining for 4–6 weeks returns the exzyme activity to normal values (7). Endurance exercise training for 8 weeks by previously sedentary men increases SDH activity by approximately 30% (8). Thus, SDH activity may be a reasonably good biochemical marker of recent exercise stress. Also, endurance type exercise primarily alters SDH activity in slow-twitch muscle fibers and more intensive activity increases SDH activity mostly in fast-twitch fibers, possibly providing a tool for documenting the relative intensity of activity performed by the individual (1). A major limitation for using SDH activity or other metabolic enzymes in skeletal muscle as an exercise marker in population-based surveys is the invasive nature of the biopsy technique now required to obtain the muscle sample. However, SDH activity could be considered for validating activity questionnaires, especially if the questionnaires were used to measure a change in exercise status over weeks or months. Even though these procedures are quite difficult, their possibilities need to be investigated.

When group means are used, many of the biochemical measures correlate quite strongly with measures of endurance exercise or fitness status. For example, figure 1 contains group mean data for plasma high-density lipoprotein cholesterol (HDL–C) concentration versus either maximal oxygen uptake or endurance exercise status. A highly positive relationship ($r = 0.97$) exists between mean HDL–C and level of fitness or exercise for a diverse group of individuals, ranging from patients who were on bed rest due to spinal cord injury all the way to world class endurance athletes who ran more than 100 miles (166 km) per week. Data presented here come from more than 20 studies using similar laboratory procedures and have been summarized by Wood and colleagues (3).

However, such group data hide substantial interindividual variations in HDL–C at similar levels of fitness or activity. Individual HDL–C concentrations are correlated on the order of 0.15 for treadmill test performance in the general population; the coefficient increases to 0.41 when change in HDL–C and treadmill performance over one year are correlated for men undergoing endurance exercise training (9). This variation appears to result from genetic differences in plasma HDL–C concentration, interindividual differences in the response of HDL–C to a similar amount of exercise, and the effect of other personal characteristics (e.g., diet composition and body composition) on HDL–C. For example, the magnitude of the negative correlation between percent body fat and HDL–C is similar to that for physical activity status and HDL–C. In the Stanford Exercise Training Study, the correlation for the change in HDL–C and change in percent body fat was −0.47, which is very similar in magnitude to the relationship between the mean change in HDL–C and the change in maximal oxygen uptake or miles per week (9).

An even more problematic situation exists for the relationship for individual measures of plasma HDL–C and the measurement of physical activity status. Although this relationship has been reported to be significantly positive (3), the amount of interindividual variation in HDL–C concentration accounted for by the measurement of physical activity is quite small. For men randomly selected from communities in northern California, the R^2 for calories expended during heavy and very heavy activity (determined by 7-day physical activity recall questionnaire) versus HDL–C ranged from 0.09 to 0.13. The R^2 was even less (0.06 to 0.) when estimated total caloric expenditure was used. Values of a similar magnitude were obtained for women. Thus, plasma HDL–C concentration can characterize *groups* of people that have very different exercise and fitness levels, but it is not very useful when trying to classify the exercise or fitness status of *individuals* in the general population.

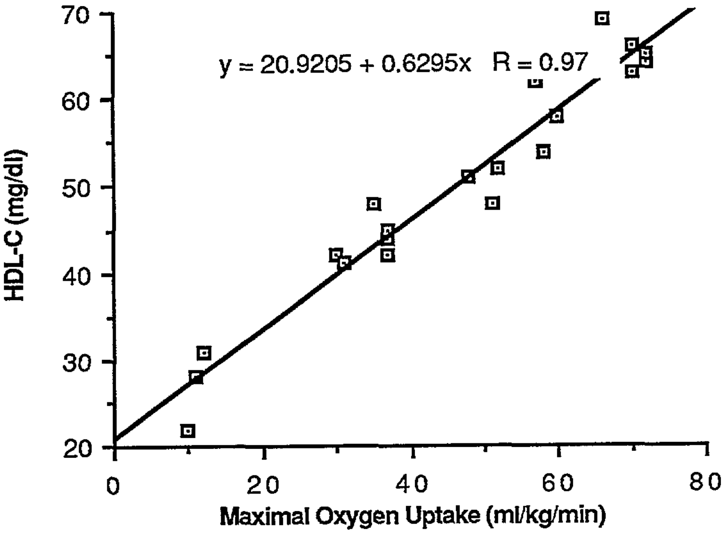


Figure 1. Relationship of plasma high-density lipoprotein cholesterol (HDL–C) concentration to maximal oxygen uptake or endurance exercise status in men. Group means from 20 cross-sectional studies. From Wood et al. (3)

This situation appears to be true for a number of the biochemical changes that have been reported to vary with endurance exercise status or fitness.

Exercise has been demonstrated to transiently change the concentration of a variety of blood constituents; but none of these constituents appears to be effective biochemical markers for exercise or fitness. Included in these measures are the serum enzymes creatine phosphokinase (CPK), serum glutamic oxaloacetic transaminase (SGOT) and lactic dehydrogenase (LDH), and alkaline phosphatase and bilirubin (10). These substances are released into the blood in response to a variety of stimuli including exercise but also includes psychological stress, alcohol and other dietary factors, and changes in body temperature due to illness. Their normal ranges are large, and test-retest reliability generally is quite poor. A variety of hormones also fall into a similar classification; their blood concentration is modified by exercise but the response is too variable and they are influenced by too many other factors to be useful biochemical markers. Included here are epinephrine and norepinephrine, growth hormone, insulin, cortisol, progesterone, and testosterone. Because these hormones are involved in various bodily functions, their blood concentrations are under the influence of numerous regulatory processes and thus are unlikely to be good indicators of just one stress, such as exercise, placed on the body.

Some biochemical measures do not differentiate between physically fit and unfit persons if measured while the person is at rest, but they do demonstrate marked differences when measured during or immediately following a standardized bout of exercise. Blood lactate concentration is a good example of such a measure because it is well documented that endurance type exercise training significantly reduces the blood lactate concentration following a submaximal exercise test (11). Following a set amount of exercise, the blood lactate concentration decreases as aerobic capacity increases and as a person's ability to work at a higher percentage of his or her capacity increases. Population-based standards for blood lactate concentration following a standardized bout of exercise have not been developed, but recent technologic advances in lactate analysis now make such developmental research possible. Lactate samples can now be rapidly and relatively inexpensively measured in a venous blood sample obtained using the finger-stick technique. In conjunction with recording heart rate during or immediately following a standard amount of exercise, the post-exercise blood lactate measure probably is one of the most effective ways of determining endurance fitness status short of performing a maximal exercise test.

Doubly labeled water as a marker for energy expenditure

One biochemical marker of exercise that has shown some promise in small animals (12) and humans (13) is the use of the doubly labeled water isotope ($^2H_2{}^{18}O$). This method was developed after it was observed that the oxygen atoms in exhaled carbon dioxide and body water were in isotopic equilibrium (14). The change in the concentration of these isotopes in urine over several days or weeks can be used to estimate the amount of CO_2 produced by the body. Assuming a representative respiratory quotient (usually 0.85), the total oxygen uptake and thus energy expenditure for this period of time can be calculated. The procedure is as follows: (a) The subject drinks a specified amount of doubly labeled water ($^2H_2{}^{18}O$); (b) after a set period of time to allow the isotope to become distributed in all fluid compartments of the body (usually 12–24 hours), a urine sample is obtained and the concentration of the isotope is determined to provide starting values for $^2H_2$ and ^{18}O in body fluids and an estimate of total body fluids; and (c) some days or weeks later a second urine sample is obtained and the concentration of $^2H_2$ and ^{18}O are once again determined by mass spectrometry. The decrease in ^{18}O concentration from the first to the second urine sample will reflect the total H_2O and CO_2 produced during this time; and the decrease in $^2H_2$ is a measure of the H_2O produced. The total CO_2 production then is obtained by subtraction. Energy expenditure and material balance are calculated from the CO_2 production using standard direct calorimetric relationships (15).

This procedure provides an objective measure of total energy expenditure over days or weeks without encumbering the subject with heart rate or motion sensor monitoring equipment. In preliminary studies with humans a difference in estimated energy expenditure using doubly labeled water and measured caloric intake was 2% with a coefficient of variation of 6% (16). When compared to energy expenditure estimates based on respiratory gas exchange over 5 days, the mean difference between the two techniques was 5.9% (SD ± 7.6%), with an 8% coefficient of variation (13). Thus, the accuracy of the measurement over days or weeks looks quite good when the procedure is performed by experienced investigators, but more studies are needed to demonstrate its applicability as a means to validate questionnaires designed to assess total energy expenditure. It appears from theoretical calculations and preliminary studies that both the precision and reliability of the procedure would exceed that of other procedures for estimating or measuring energy expenditure for days or weeks in free living subjects. As more research is performed in humans using this isotope (data on only nine human subjects have been published), the differences in assumptions and procedures that currently exist among scientists working in the area should be resolved and a highly standardized procedure developed (15). The relatively high cost of the isotopes has inhibited the rapid development of this procedure in humans. Although the dosage of the isotopes can be reduced to 0.3 g $^2H_2{}^{18}O$/kg total body water and 0.12 g ^2H_2O/kg total body water by increasing the sensitivity of the detection procedure using a differential isotope ratio mass spectrometer, it still costs approximately $300 per human subject just for the isotopes. This cost currently makes the use of doubly labeled water as a measure of physical activity in large-scale population surveys unlikely.

Measurement of exercise versus physical fitness

Given the transient nature of most biochemical responses to a bout of exercise, especially if the exercise is not very vigorous or sustained, it is more likely that an intrinsic biochemical marker can be identified for endurance fitness than for habitual exercise status. Generally it is considered that "fitness" is maintained for some days or weeks after the most recent bout of exercise. During this time if a biochemical marker is obtained from physically trained and sedentary persons, it should be able to accurately differentiate them. If one uses as an index of fitness, endurance performance capacity (the capacity to exercise for extended periods of time) or maximal oxygen uptake, then biochemical measures exist that correlate quite well with fitness status. Most notable of these measures are selected characteristics of skeletal muscle usually determined by analysis of tissue obtained using the needle biopsy procedure (see table 1). Individuals that have a high endurance capacity demonstrate increased oxidative capacity of skeletal muscle; for example, they have a higher mitochondrial mass and oxidative enzyme activity (1), lipoprotein lipase activity (17) and glycogen content (2). These biochemical indexes are closely linked to a greater capillary density in muscle and a greater proportion of skeletal muscle fitness being of the slow-twitch or red variety (1). These relationships are partly caused by the effects of exercise training on endurance capacity and on the biochemical characteristics of skeletal muscle, but a substantial portion of this relationship results from genetic influences. Skeletal muscle fiber composition is basically an inherited trait, and many of the biochemical features of muscles are determined largely by the muscle fiber composition (1, 2).

We run into the same problem with biochemical markers for fitness as encountered using physiologic or performance measures: Only part of the interindividual variation in fitness results from activity. Thus, even if we can accurately assess fitness status using a biochemical marker, this does not mean that we can accurately classify individuals according to their physical activity habits. It is important to realize that measures of physical fitness cannot be used as surrogate measures of physical activity in cross-sectional studies except for the general classification of people (very sedentary versus highly active). However, in longitudinal training studies, a major change in exercise status usually is highly correlated with changes in performance and physiologic and biochemical indexes of fitness.

Cross-sectional versus longitudinal measurements

Selected variables may not be good biochemical markers of exercise status or fitness in cross-sectional studies, but they do reflect quite accurately changes in fitness or activity by the individual. Variables most likely to fall in this category are those that have a major genetic component determining their value when the person is sedentary but relatively small heredity influences on interindividual responses to exercise training. Aerobic capacity (VO_2max) is a reasonably good physiologic example of this feature. It appears that the genetic influence on the interindividual variation in VO_2 in a population is substantial (18) and that there is a stronger correlation between change in endurance exercise and change in VO_2 max in endurance training studies than in the cross-sectional correlation for the two variables. It is likely that a similar situation exists for some biochemical markers of exercise (e.g., SDH activity of skeletal muscle), but adequate data are not available to clearly identify them.

Homogeneity of exercise habits

In surveys of the general population, a major difficulty encountered in attempting to identify a biochemical marker of exercise or fitness is the homogeneity of exercise among a majority of the population. For example, most recent surveys of exercise habits of American adults indicate that less then 15% of men and women over age 50 years of age participate in any vigorous exercise on a regular basis and the other 85% are quite homogenous in their activity pattern (19). Such homogeneity of exercise will inhibit any attempt to identify a truly significant relationship between exercise status and a biochemical marker when its inter-individual variation also is influenced by genetic and other environmental factors. This difficulty is made worse when an activity questionnaire with less than excellent validity and/or reliability is used.

Potential value of an index

Some biochemical measures that are influenced by exercise status but also by other personal characteristics might become a better marker for exercise if the biochemical measure and the other characteristics were combined in some type of index. For example, if body composition is the personal characteristic, then a ratio of the biochemical measure to body weight, body mass index, or percent body fat might significantly improve the relationship to exercise or fitness in the same way maximal oxygen uptake per kilogram body weight (or lean mass) is better correlated with exercise performance or exercise status than maximal oxygen uptake expressed as liters per minute.

Summary and recommendations

Except for use of the doubly labeled water isotopes $^2H_2{}^{18}O$ to determine total energy expenditure over a period of days or weeks, no biochemical marker of exercise or fitness that is feasible for use in large-scale, health-oriented, population-based studies has been identified. Selected blood chemistry or skeletal muscle metabolic features can be used to place individuals into broad categories of endurance fitness, but these lack accuracy because of the confounding influence of genetic and environmental factors on all of the measures identified. At best, these biochemical measures provide only an integrated measure of exercise or fitness and do not provide any data on the nature of the activity performed, particularly valuable would be the amount of time spent at various intensities. An understanding of such an exercise profile may be critical in relating exercise status to various health-related outcomes.

Very little research has focused directly on the issue of attempting to identify and validate a specific biochemical marker for exercise. Even promising measures like the doubly labeled water technique have been investigated only in several laboratories and on a very limited number of humans. It seems worth the effort to concentrate research on considering how to control or adjust for some of the genetic and environmental influences that confound the use of specific measures as biochemical markers. Looking at the ratio between selected measures might control for some nonexercise sources of variation, or expressing the values per unit of muscle or total body mass might increase both validity and reliability. Also, performing the analysis at the end or immediately following a standard submaximal exercise test might improve the effectiveness of the biochemical measure as an exercise or fitness marker.

Additional research should be supported to further investigate the potential use of newly developed techniques as possible biochemical markers of exercise or fitness. Such research should include the further evaluation of the use of isotopes like $^2H_2{}^{18}O$ to determine metabolic rate, the identification of other blood or urinary constituents that might reflect recent energy expenditure or fitness status, and the evaluation of advanced technologies like magnetic resonance imaging or positron emission tomography for use in assessing biochemical markers of exercise. Although many of these procedures may not be feasible for use in large studies of the general population, they may prove to be very useful in determining the validity and reliability of various physical activity questionnaires and diaries that can then be used more effectively in population studies.

References

1. Saltin B, Henriksson J, Nygaard E, Andersen. Fiber type and metabolic potentials of skeletal muscles in sedentary man and endurance runners. *Annals New York Academy of Sciences* 1977; 301:3–29.
2. Holloszy J. Adaptations of muscular tissue to training. In *Exercise and Heart Disease,* Editors E.H. Sonnenblick and M. Lesch. Grune and Stratton, New York, 1977; 25–38.
3. Wood PD, Williams PT, Haskell WL. Physical activity and high-density lipoproteins. In *Clinical and Metabolic Aspects of High-Density Lipoproteins,* Editors N.E. Miller and G.J. Miller. Elsevier Science Publishers. London, 1984; 133–165.
4. Bonen A, Ling W, MacIntyre, Neil R, McGrail J, Belcastro A. Effects of exercise on the serum concentrations of FSH, HL, Progesterone and Estradiol. *European J Appl Physiology* 1979; 42:15–23.
5. Costill D, Daniels J, Evans W, Fink W, Krahenbuhl G, Saltin B. Skeletal muscle enzymes and fiber composition in male and female and track athletes. *J Appl Physiol* 1986; 90:149–154.
6. Nygaard E. Adaptional changes in human skeletal muscle with different levels of physical activity. *Acta Physiol Scand* 1976; Suppl 440:291.
7. Henriksson J, Reitman J. Time course of activity changes in human skeletal muscle succinate dehydrogenase and cytochrome oxidase activities and maximal oxygen uptake with physical activity and inactivity. *Acta Physiol Scand* 1977; 99:91–97.
8. Henriksson J, Reitman J. Quantitative measures of enzyme activities in type I and type II muscle fibers in man after training. *Acta Physiol Scand* 1986; 97:392–397.
9. Wood P, Haskell W, Blair S, Williams P, Krauss R, Lindgren F, Albers J, Ho P, Farquhar J. Increased exercise level and plasma lipoprotein concentrations: A one-year, randomized, controlled study in sedentary, middle-aged men. *Metabolism* 1983; 32:31–39.
10. Martin R, Haskell W, Wood P. Blood chemistry and lipid profiles of elite distance runners. *Annals of the New York Academy of Sciences* 1977; 201:346–360.
11. Ekblom B, Astrand P-O, Saltin B. Effect of training on circulating responses to exercise. *J Appl Physiology* 1968; 24:518–528.
12. Lifson M, McClintock R. Theory of use of the turnover rates of body water for measuring energy and material balance. *J Theor Biol* 1966; 12:46–74.
13. Schoeller D, Webb P. Five-day comparison of the doubly labeled water method with respiratory gas exchange. *Am J Clin Nutrition* 1984; 40:153–158.
14. Lifson N, Gordon G, Visscher M, Nier A. The fate of utilized molecular oxygen and the source of heavy oxygen of respiratory carbon dioxide, studies with the aid of heavy oxygen. *J Biol Chem* 1949; 180:803–811.
15. Schoeller D. Energy expenditure from doubly labeled water: some fundamental considerations in humans. *Am J Clin Nutrition* 1983; 38:999–1005.
16. Schoeller D, van Saten E. Measurement of energy expenditure in humans by doubly labeled water method. *J Appl Physiol* 1982; 53:955–959.
17. Lithell H, Cadermark M, Froberg J, Tesch P, Karlsson J. Increase of lipoprotein-lipase activity in skeletal muscle during heavy exercise. *Metabolism* 1981; 30:1130–1134.
18. Bouchard C, Malina R. Genetics of physiological fitness and motor performance. In *Exercise and Sport Science Reviews* (volume 11). Editor, R.L. Terjung, Franklin Institute Press, Philadelphia, 1983; 306–339.
19. Sallis J, Haskell W, Wood P, Fortmann S, Rogers T, Blair S, Paffenbarger R. Physical activity assessment methodology in the Five-City Project. *Am J Epidemiology* 1985; 121:91–106.

Chapter 21

Evaluating the Health Effects of Demanding Work on and off the Job

James S. House, Ph.D.
David A. Smith, Ph.D.
University of Michigan

This paper addresses measurement and analysis issues in the assessment of the physical and psychological demands of work in general population surveys, especially those of the National Center for Health Statistics (NCHS). The paper has three thrusts that differentiate it from both the general literature on physical fitness and activity and other chapters prepared for this book. First, the paper is concerned with activity associated with paid and unpaid *work.* Second, the paper views work as having potentially *deleterious as well as beneficial* consequences for health; that is, work activity can be excessively demanding as well as energizing and strengthening. Third, the paper focuses on the *psychosocial* as well as physical activities associated with work.

The focus on work is important because paid employment and household work occupy close to half of the waking hours of most adults under age 65 in our society, as shown in table 1. This table presents data from the University of Michigan Institute for Social Research's 1975–76 Time-Use Survey, a comprehensive and nationally representative study of time use in the full adult population of the United States (1). We suggested that the growing concern with nonwork-related activity and exercise as determinants of health stems in large part from the changing nature of work in modern society. Prior to the 19th century, most adults were engaged 60 or more hours per week in agricultural or household work, which provided them with considerable physical activity as part of their daily work routines. With the advent of the Industrial Revolution, the locale of work moved increasingly from farm and home to factory, mine, or transport vehicle. Much of this work also provided ample physical activity. More recently the locale of work has moved increasingly to offices and service businesses, and concomitantly the activities of work have become more mental and less physical. Thus, as physical activity has become an intrinsic part of work for fewer and fewer people, increasing attention is being paid to assessing and enhancing physical activity levels outside of work as a means of promoting and maintaining health.

This new focus on leisure physical activity must, however, continue to recognize that work still occupies the majority of adults' waking time. On the basis of sheer exposure, work must be considered a major potential influence on health, for better or worse. To the extent that individuals engage in ample physical activity as part of their work, less needs to be done to assess or enhance their physical activity levels outside of work. However, in work as opposed to leisure, individuals do not as often choose the nature and level of the activities in which they engage. The nature and level of work activity, both physical and mental, are often dictated by external forces and may easily become potentially noxious or excessively demanding. Thus, in studying activity at work, one must be equally concerned with its deleterious as with its beneficial effects for health. Any comprehensive effort to assess physical activity and fitness must consider work as well as nonwork activity and view such activity as potentially health damaging as well as health promoting.

Finally, the increasingly mental and interpersonal nature of work makes it imperative that we attend to the psychosocial as well as physical nature of work activity and demands. Social and psychological factors are increasingly recognized as major determinants of health, both at work and more generally (2–4). Indeed any activity, whether at work or outside of work, is simultaneously physical or biological, social, and psychological. To assess and evaluate only the physical component of any activity could be to seriously misspecify the dimensions of the activity that are important for health.

Occupational demands and conditions related to health

Given that the physical and psychosocial demands and conditions of work must be a component of general population epidemiologic studies of fitness, activity, and health, what are the specific demands and conditions that should be measured? Priority must be given to those factors that extant theory and research suggest are most clearly and consequentially linked to health. This paper cannot exhaustively review previous work in these areas but will draw on previous reviews to briefly summarize the evidence.

House and Cottington's (5) selective review of literature on health in the workplace and the effort of Karasek et al. (6) to identify and evaluate measures of working conditions related to cardiovascular disease (and health more generally) have arrived quite independently at very similar conclusions regarding the critical aspects or conditions of work that are most consequential for health. Their conclusions are consonant with a broader literature (4). These conditions can be summarized as:

- Physical-chemical-biological factors
  - Physical-chemical-biological hazards
  - Physical activity at work

- Psychosocial demands and conditions of work
  - Job insecurity and loss
  - Psychosocial job demands
  - Worker control, autonomy, and participation
  - Work-related social support

Let us comment briefly on each of these as they relate to the assessment of health-related demands and conditions of work.

Table 1. Hours per week devoted to various activities by adults, by sex and age: United States, 1975–76

	Male		Female			
			18–64 years			
Activity	18–64 years (*n* = 348)	65 years and over (*n* = 63)	Working [1] (*n* = 260)	Not working [1] (*n* = 206)	65 years and over (*n* = 95)	Total (*n* = 976)
Total	168.00	168.00	168.00	168.00	168.00	168.00
Paid employment	42.14	5.53	34.46	4.01	2.95	25.77
Housework and yardwork	7.42	14.33	15.88	24.86	22.18	15.26
Child care	1.40	0.75	3.14	6.93	0.47	2.89
Services, shopping	3.82	5.01	6.14	7.25	6.23	5.47
Personal care	20.58	28.03	20.33	21.89	28.32	22.05
Education	2.41	0.26	1.00	1.84	0.09	1.54
Nonreligious organizations	1.11	1.88	0.94	1.92	2.27	1.40
Religious organizations	1.20	1.53	1.15	2.24	2.51	1.56
Entertainment (movies, plays, etc.)	3.75	1.98	3.88	3.06	1.41	3.29
Visiting friends, relatives	3.86	5.29	4.80	6.42	5.60	4.91
Active leisure outdoors	2.98	2.71	1.30	1.30	0.80	1.94
Active leisure indoors	2.54	4.38	2.53	4.30	6.65	3.42
Television	13.19	21.78	12.02	14.91	16.30	14.14
Conversations	1.62	1.83	1.78	3.03	1.51	1.96
Other passive leisure	5.47	12.03	4.39	5.95	11.20	6.30
Night sleep	54.60	60.75	54.35	58.19	59.59	56.20

[1] As of fall 1975.

Note: The hours per week spent in each activity are based, of course, on measures obtained at three or four points in time from fall 1975 through spring/summer 1976. Thus, persons not working in fall 1975 may have subsequently obtained paid employment.

Source: Computed from data provided by 1975–76 University of Michigan Institute for Social Research Study of Time Use. See Juster and Stafford (1).

Physical-chemical-biological factors

Physical-chemical-biological hazards

The topic of physical, chemical, and biological hazards at work is generally beyond the scope of this chapter and the NCHS data collection vehicles. A voluminous literature on occupational safety and health has documented the adverse effects of a wide range of such hazards (7–9). Assessment of exposure to specific hazards, however, must vary by occupation and industry and even by production methods within a given industry. Such assessments will generally be best accomplished at the worksite. Hence, general population surveys are not the appropriate vehicle for assessing very specific hazards.

On a number of grounds, however, it would be useful and appropriate to include in such surveys some brief and general measures of exposure to physical-chemical-biological hazards. First, such general measures have been, and are likely to continue to be, associated with a wide range of health outcomes from injuries to cancers, to respiratory diseases, to cardiovascular diseases (10, 11). Thus, NCHS can provide useful, if crude, national data on the prevalence of such hazards and their relation to health outcomes. Second, some control for such factors is essential in analyses of the health effects of work-related physical activity or psychosocial conditions because these variables may be confounded with physical-chemical-biological exposures. Finally, there may be synergistic relationships among physical activity, psychosocial conditions, and physical-chemical-biological hazards in predicting health (11, 12).

Physical activity at work

There is by no means perfect consensus regarding the existence of a predictive association between levels of work activity and health, much less the magnitude or interpretation of such an association (13, 14). Balanced and careful reviews, however, increasingly support the view that at least moderate levels of physical activity, both on the job and at leisure, reduce the risk of coronary heart disease and can promote more general physical and mental health. Such reviews also recognize the need for more careful analysis examining the health effects of work-related and leisure activity in relation to each other and to other major disease risk factors (15, 16). Included here would be analysis of the degree to which physical activity may be confounded with psychosocial factors such as job stress and social support. Thus, assessment of physical activity on the job will be considered here, although recommendations for measurement made here should be coordinated with more general recommendations, often encompassing physical activity on and off the job, made in other chapters (chapter 24).

Psychosocial demands and conditions of work

Psychosocial demands and conditions of work, often discussed under the general rubric of "job stress," have been increasingly implicated in the etiology of a wide range of physical and mental health problems (4, 17–24). As with the data on physical activity, consensus is not complete on the meaning of existing data (25, 26), but the weight of evidence is increasingly compelling regarding at least selected psychosocial job demands and conditions (5, 6).

Job insecurity and loss

Lack or loss of a job, or the threat thereof, has been repeatedly found to increase the risk of a wide range of adverse health outcomes (4, 27, 28). Job insecurity and loss have reemerged in the past decade as increasingly prevalent job conditions because of growing international economic competition, accelerating organizational and technological change, and volatility in the international and domestic economies and labor markets. Thus, unemployment and job insecurity must be attended to in any serious assessment of health-related job demands and conditions.

Psychosocial job demands

An increasing array of cross-sectional, retrospective, and prospective studies document the importance for health of a broad category of what may best be termed "psychosocial job demands," including high levels of workloads, responsibility, and interpersonal or role conflicts (5, 6). Related to this work is the voluminous literature linking the Type A behavior pattern to coronary heart disease because job involvement and pressures have been a major component of the Type A construct from its inception (21, 29, 30). Thus job demands and Type A will be considered in our discussion of assessment procedures.

In recent years, the literature on stress and health has focused increasingly on variables that may not only affect health in their own right but also may ameliorate (or exacerbate) the impact of stress on health. In this search for so-called buffers of the deleterious health effects of stress at work or more generally, two factors have emerged as clearly preeminent—social support and personal control.

Social support

House (31) has reviewed a wide range of research on animals and humans, in the laboratory and in the field, showing that supportive social relationships on and off the job can reduce stress, improve health, and buffer the impact of stress on health. Subsequent major prospective studies have confirmed this conclusion (32, 33). Both cross-sectional (34–37) and prospective (38–41) studies have found that support from work supervisors and coworkers reduces occupational stress and/or its deleterious health effects. Thus, social support at work is a major determinant of health that must be included in any comprehensive assessment of health-related demands and conditions of work.

Worker control

Sutton and Kahn (42) have similarly reviewed a wide variety of laboratory and field studies suggesting that if individuals have greater ability to "predict, control, or understand" events in their organizational environment, they will experience less organizational stress or be less adversely affected by it. This conclusion is supported by research of Karasek, Theorell, and associates on worker control or decision latitude in the United States and Sweden (25, 43, 44) and by a larger literature showing the stress-buffering and health-protective effects of a sense of control or mastery over one's environment (45, 46).

Strategies for assessment of health-related demands or conditions of work in epidemiologic surveys

Given that we know what psychological and physical demands of work we want to assess, how should we go about assessing them? This question devolves into still others. First, what should be the *object* or target of our assessment? Second, what should be the *source* or method of our assessments? The answers to these questions can clearly be neither simple nor definitive, despite substantial work having already been done on these problems. (This discussion will deal with the assessment of a particular object at a particular time. Obviously, the nature of a job or occupation may change over time. Thus, there is a need for repeated measures, with that need varying directly with the rate of change versus stability in the object of measurement.)

Objects of assessments

Jobs versus occupations

The first issue to be confronted is whether we wish to assess the demands and conditions of work associated with a particular job and a particular worker (e.g., Secretary to the President of the United States) or the prototypical demands and conditions of work associated with a group of similar jobs (e.g., executive secretary or secretaries) or the members of a particular organization or industry (e.g., U.S. Government or government in general) or combinations thereof (governmental secretaries). A group of similar jobs, often defined within a particular industry, is usually termed an occupation.

For purposes of predicting the health of individuals, measures at the level of individual jobs and workers are likely to prove most useful because demands and conditions of work are likely to vary across jobs within occupations. However, for other reasons, ranging from ease of assessment to the need to understand and influence the demands and conditions confronting broad classes of workers, measures at the level of occupations are often the best or most feasible option. Substantial data bases and data-gathering mechanisms organized by occupations and industries already exist and will be discussed later. Nevertheless, we will point to the importance of complementary efforts to make assessments at the level of individual jobs or workers.

External conditions versus overt behavior versus internal reactions

Whether we consider an individual job or worker or consider a group of jobs or workers in an occupation, we can ask several questions about the demands and conditions of their work. First, what are the *external* demands or conditions of work that confront anyone in the job or occupation? Second, how does a worker or group of workers *overtly behave* in response to these demands or conditions? Finally, what are the *internal psychological or physiological reactions* of a worker or group of workers to these external demands or work behavior?

This trichotomy of external demands and conditions versus overt behavior versus internal reactions corresponds to a trichotomy embodied in more general theories of social roles, of which jobs and occupations are one type. Levinson (47), for example, distinguishes among "structurally given demands," the "actions" of individual role occupants, and the role occupant's "orientation-conception" or "definition" of the role. Similarly, Katz and Kahn (48, Chapter 7) distinguish among "role expectations," "the received (or perceived) role," and "role behavior." Levinson, Katz and Kahn, and we agree essentially on the meaning of external role demands or expectations and role behavior. However, our definitions of individuals' perceptions and reactions to roles differ. Katz and Kahn's concept of "received role" focuses on the perception of external role demands or expectations; Levinson's concept of "role-definition" focuses on the role occupant's interpretations and redefinition of the role and allows for a transformation of both sent and received roles in response to the individual's own personality. Our conception of internal reactions denotes the psychological and physiological responses produced by, or associated with, given external demands, role behaviors, or personal perceptions and definitions of the role. Causally, these concepts may be ordered as follows. (Lagged effects are not shown but are important possibilities.)

(Sent) role demands or expectations → (Received) perception of role demands or expectations → Personal role conception or definition → Role behavior → Psychological or physiological reactions

These distinctions are, however, more easily made conceptually than operationally. To operationalize these distinctions clearly, we would need to measure external demands by assessing how people other than the focal person define them; role definition by self-reports; role behavior by observation; and psychological and physical reactions by a combination of self-reports and, ideally, physiological tests. Whether we use the reports of role occupants or of observers or role senders, however, all of these sources are likely to have knowledge simultaneously of role demands or expectations, role behavior, and role perceptions, definitions, and reactions. Thus, when a role occupant or an observer reports on role demands, he or she will generally make use of observed behavior and even reactions in defining what the role demands. Similarly, in observing, interpreting, and reporting about role behavior, we generally take account of the context, including role demands and attributes of the person and also including aspects of role perceptions and definitions. Thus, for example, we may infer the existence of both role demands and role conceptions or reactions from behavior.

We neither have, nor are likely to see in the near future, methods of measuring very separately these various aspects of roles, occupational or otherwise. We can, however, aspire toward distinguishing these different aspects of job and role operationally as well as theoretically. Thus our discussion of measures will focus primarily on the distinction between jobs and occupations as targets of assessment, and a second distinction between self-reports versus observations as sources of assessments. At various points, however, we will note possibilities of distinguishing external demands, behavior, and reactions.

Sources of assessments: Self-reports versus observations

We can generate assessments of jobs or occupations by asking persons in particular jobs or occupations to report about the externally given demands or conditions of work and their behaviors and reactions. Such data are commonly termed *self-reports.* Alternatively, we can attempt to derive assessments of these objects by asking persons external to the job or occupation to describe the demands or conditions of the job or occupation and how workers behave and react in response to these demands or occupations. Or we can use an external mechanical instrument; for example, having a machine record how often a telephone rings, how often a secretary answers it, and how the secretary's heart rate varies over the workday (or in relation to the ringing or answering of the telephone). Such data derived from human or mechanical sources external to the worker, job, or occupation in question can be termed *observations.*

A typology of assessment modes

Combining the distinction between jobs and occupations as potential targets of assessment with the distinction between observations and self-reports as sources of assessments yields the typology of modes for assessing health-related demands and conditions of work, as shown in figure 1. Each of these types can potentially be used to assess external demands, worker behavior, worker reactions, or amalgams thereof.

From the perspective of epidemiologic surveys in the general population, the assessment modes in the lower left-hand cell (C) are generally not feasible. These require direct worksite observation or monitoring of an individual worker in a particular job. Hence, these procedures require access to the worksite and are very expensive in terms of time and resources. For these reasons, they are not widely used even in more focused research studies. Perhaps the most extensive effort of this type to date has been the University of Michigan Survey Research Center's Effectiveness in Work Roles Project (49), which focused on external demands and conditions and on overt behavior. The primary relevance of such individual job or worker rating procedures for purposes of epidemiologic surveys would be as a step toward generating or validating assessments in the other cells that are potentially very useful in epidemiologic surveys. Thus, such procedures will be considered here only as they pertain to the development or validation of the other types of measures.

More relevant for epidemiologic surveys are the procedures in cell D of figure 1, which are termed "aggregated ratings of job demands/conditions." The *Dictionary of Occupational Titles* (DOT) of the U.S. Department of Labor (50) provides the most useful example of this type of assessment and will be discussed in detail. The DOT is derived from the aggregation of observer ratings of individual jobs. These ratings consti-

Source of assessment	Target of assessment	
	(Individual) job	Occupation
Self-reports	A. Job description (e.g., Karasek's Job Content Survey)	B. Aggregated job description (e.g., Karasek's Aggregated Job Content Survey Scores for Census Occupations)
(External) observations	C. Rating of demands and conditions for (individual) job (e.g., Survey Research Center Effectiveness in Work Roles Project)	D. Aggregated rating of job demands/conditions (e.g., *Dictionary of Occupational Titles)*

Figure 1. Typology of modes of assessing health-related demands and conditions of work.

tute an amalgam of the external demands and conditions of a job and the way workers behave and react in the face of these demands. Both types of self-report methods shown in cells A and B of figure 1 are potentially very useful in epidemiologic surveys of the general population and will be discussed fully, with special attention to the work of Karasek and colleagues (51).

Aggregated ratings of job demands and conditions by occupation: Existing procedures and future directions

Because the observation of the particular jobs of individual workers is simply not feasible in NCHS general epidemiologic surveys, and indeed most research, efforts to observe or rate demands and conditions of work must be focused at the level of occupations. The general strategy is somehow to generate a set of observations or ratings of the demands and conditions associated with particular *sets* of jobs comprising an occupation. Once this set of ratings of an aggregate of jobs or an occupation is generated, it can be applied in further research to anyone working in that occupation.

The generation of the set of aggregate job (or occupational) ratings is a highly complex and expensive process, especially on a national scale. Once these ratings exist, however, they can be used relatively easily and inexpensively in other research. Other researchers must simply be able to classify or code individuals into the set of occupations for which ratings are available. They can then assign the occupational ratings to those individuals and use the ratings to predict individual level outcomes, such as health.

Obviously, the validity and utility of this strategy depends on the quality (i.e., validity and reliability) of the occupational ratings and their substantive appropriateness to specific research uses (e.g., whether the occupational ratings cover those demands and conditions of work most relevant to health). These are the considerations we use in evaluating existing procedures and future directions in the observation of external demands and conditions of work at the level of occupations.

The *Dictionary of Occupational Titles*

The *Dictionary of Occupational Titles,* created and published by the U.S. Department of Labor (50, 52–56) represents the only comprehensive national effort to observe and rate the demands and conditions of work for the full range of occupations in the United States. This section relies heavily on the review and critique of the DOT by the Committee on Occupational Classification and Analysis of the Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, as reported by Miller et al. (57) and Cain and Treiman (58). Although originally created to match job applicants with jobs, the DOT contains a great deal of information that has been increasingly utilized in social science research (59–66) but only rarely in analyses related to health (67, 68).

The fourth edition of the DOT, published in 1977, contains written descriptions of 12,099 "base" occupations, which are considered to include 28,801 specific jobs or job titles (the remainder of which are close variants or synonyms of the base occupations). Also available, on computer tape, are ratings of each occupation on 44 characteristics or traits, including the type and complexity of work, the education and training required, aptitudes, temperaments and interests appropriate for the occupation, and the physical demands and physical working conditions of the occupation. These characteristics and their coding are listed in table 2 (58). It is also possible to match standard U.S. Bureau of the Census occupation and industry codes to the DOT codes (and vice versa) and thus to link selected ratings of job characteristics to the Census codes (57, 58, 65).

Until the recent review of the DOT by the Committee of the National Research Council, National Academy of Sciences, or NRC/NAS (57), little was understood about the way in which the list of occupations in the DOT and the descriptions and ratings associated with them had been generated and hence about their validity, reliability, and utility for general research purposes. This NRC/NAS review both explicated and critiqued these procedures and hence assessed the strengths and weaknesses of the resulting data base. The DOT clearly has a number of limitations both for general research purposes and for the research purpose at issue in this paper. Nevertheless, the NRC/NAS committee concluded: "Despite the deficiencies in the fourth edition worker trait and worker function ratings identified here, they remain the most comprehensive set of occupational characteristics currently available. As such, their use should be encouraged" (57, p. 195). In the words of Cain and Treiman (58, p. 273):

> Notwithstanding . . . the . . . problems and peculiarities we have explored in this paper, the DOT is indisputably a rich and unique source of occupational data. We therefore encourage the increased use of these data . . . , both because we believe the potential of the data has not yet been fully realized and because increased use will afford still better understanding of the data's properties. We thus strongly urge that occupational responses in sample surveys be coded with the nine-digit DOT codes as well as with the commonly-used three-digit Census codes.

We will shortly come to a similar recommendation for NCHS, coupled with a recommendation to make major efforts to influence the future nature of the DOT in ways that will make it more relevant for health research.

The strengths of the DOT are almost self-evident. It provides a wealth of external observer ratings for a full range of rather finely detailed occupations covering the

Table 2. ***Dictionary of Occupational Titles*, occupational characteristics, fourth edition**

Variable label	Description [1]	Scoring
Worker functions		
DATA	Complexity of function in relation to data	[2] 0 to 6
PEOPLE	Complexity of function in relation to people	[2] 0 to 8
THINGS	Complexity of function in relation to things	[2] 0 to 7
Training times		
GED	General educational development	1 to 6
SVP	Specific vocational preparation	1 to 9
Aptitudes		
INTELL	Intelligence	[2, 3] 1 to 4
VERBAL	Verbal aptitude	[2] 1 to 5
NUMER	Numerical aptitude	[2] 1 to 5
SPATIAL	Spatial perception	[2] 1 to 5
FORM	Form perception	[2] 1 to 5
CLERICAL	Clerical perception	[2] 1 to 5
MOTOR	Motor coordination	[2] 1 to 5
FINGDEX	Finger dexterity	[2] 1 to 5
MANDEX	Manual dexterity	[2] 1 to 5
EYEHAND	Eye-hand-foot coordination	[2] 1 to 5
COLORDIS	Color discrimination	[2] 1 to 5
Temperaments		
DCP	Direction, control, and planning	0/1
FIF	Feelings, ideas, or facts	0/1
INFLU	Influencing people	0/1
SJC	Sensory or judgmental criteria	0/1
MVC	Measurable or verifiable criteria	0/1
DEPL	Dealing with people	0/1
REPCON	Repetitive or continuous processes	0/1
PUS	Performing under stress	0/1
STS	Set limits, tolerances, or standards	0/1
VARCH	Variety and change	0/1
Interests		
DATACOM	Communication of data vs. activities with things	[4] −1 to 1
SCIENCE	Scientific and technical activities vs. business contact	[4] −1 to 1
ABSTRACT	Abstract and creative vs. routine, concrete activities	[4] −1 to 1
MACHINE	Activities involving processes, machines, or techniques vs. social welfare	[4] −1 to 1
TANGIBLE	Activities resulting in tangible, productive satisfaction vs. prestige, esteem	[4] −1 to 1
Physical demands		
STRENGTH	Lifting, carrying, pulling, pushing	1 to 5
CLIMB	Climbing, balancing	0/1
STOOP	Stooping, kneeling, crouching, crawling	0/1
REACH	Reaching, handling, fingering, feeling	0/1
TALK	Talking, hearing	0/1
SEE	Seeing	0/1
Working conditions		
LOCATION	Outside working conditions	1 to 3
COLD	Extreme cold	0/1
HEAT	Extreme heat	0/1
WET	Wet, humid	0/1
NOISE	Noise, vibration	0/1
HAZARDS	Hazardous conditions	0/1
ATMOSPHR	Fumes, odors, dust, gases, poor ventilation	0/1

[1] Descriptions are from U.S. Department of Labor (1972). See Appendix B for additional details on variable scoring.

[2] High scores correspond to low values.

[3] Level 5 is not assigned on this aptitude because it is assumed that every job requires at least a 4.

[4] Interest variables are sets of bipolar contrasts: 0 corresponds to the presence of neither interest in the pair; −1 corresponds to the presence of the second interest in the pair; 1 corresponds to the presence of the first interest in the pair

Source: Adapted from P.S. Cain and D.J. Treiman: The Dictionary of Occupational Titles as a source of occupational data. *American Sociological Review,* 46, 253–278, 1981. (Copyright; used by permission.)

entire U.S. labor force. These are available for use at minimal cost in any research survey that codes the occupations of respondents into the DOT system. The weaknesses of the DOT fall into three main areas: (a) The sampling of occupations (or lack thereof), (b) the uniformity and quality of the observational and rating procedures, and (c) the domains of content included in the ratings (57, 58). The first two derive from the procedures by which the DOT is generated and are relevant to almost any research use of the DOT. The third arises because the DOT was originally developed for vocational counseling purposes, but it is increasingly used for a variety of research purposes.

Neither the selection of occupational titles for inclusion in the DOT nor the generation of descriptions and ratings for these titles has been guided by systematic and uniform procedures. The basic approach of the DOT is to use onsite observations of workers performing jobs to generate descriptions and ratings of specific jobs (job analysis schedules), which are then aggregated into occupational descriptions and ratings. These observations and ratings are carried out by job analysts at 10 field centers. Each center is assigned a group of *industries,* and each analyst is generally charged with creating a comprehensive set of job analyses for each industry by visiting several establishments in the industry. There is no representative sampling of industries, establishments, or jobs and occupations within establishments and industries. Further there is no systematic procedure for deciding what constitutes a unique or "base occupational title" and what jobs fit within a given title. Finally, the quality and quantity of the observations entering into each occupational description and rating are highly variable. For some jobs, onsite observation is difficult and data are obtained indirectly from professional or trade associations, unions, or employers. For the vast majority of occupations, descriptions and ratings carry over from one edition of the DOT to the next with little or no new onsite observation.

The lack of systematic occupational sampling, coupled with the strong carryover of titles, descriptions, and ratings between DOT editions, results in substantial underrepresentation of the industries and occupations (i.e., professional, technical, managerial, sales, clerical, and services) that increasingly predominate in our economy. As Cain and Treiman (58, p. 260) put it:

> Two-thirds of the establishments in which job analyses were conducted were in manufacturing industries despite the fact that these industries made up only 8% of the total U.S. establishments and employed one-third of U.S. employees. Retail trades and services, on the other hand, which accounted for approximately 56% of employees, constituted only 11% of the establishments selected for DOT job analysis.

Similarly, although 66% of the labor force are in professional, managerial, clerical, sales, and service occupations, these occupations constitute only 24% of the titles in the DOT.

The variable and often unknown validity and reliability of the rating procedures is another cause for concern about the DOT. Cain and Treiman (58, p. 260) note that "21% of the occupational descriptions in the fourth edition were *not* derived from on-site observation (but rather from past observations or other information), while 43% of the occupational descriptions and ratings were based on only one or two job analyses." A 1-in-40 random sampling of occupational titles in the fourth edition of the DOT found 81% of them identical to those in the third edition. Thus, the procedures for observing, rating, and describing jobs are by no means uniform or equally current across occupations.

The NRC/NAS panel commissioned Cain and Green (69, 70) to conduct a specific study of the reliability and validity of the DOT ratings. They had experienced analysts at seven of the regional centers rate a common set of 24 occupational descriptions on 20 of the 44 standard DOT dimensions. Cain and Green concluded (70, pp. 163–164):

> Our results indicate the need for some caution in the use of ratings available from the DOT. Since reliabilities differed considerably from characteristic to characteristic [ranging from about .4 to .8, with most toward the higher end], selective use of these indicators is advised. The ratings of work function in relation to things and the strength variables, for example, should not be relied upon heavily. Fortunately, the ratings of some of the most commonly used DOT variables [i.e., DATA, PEOPLE, GED, and SVP in table 2] appear to be highly reliable. The low reliability of certain DOT ratings can be partially offset by using them in multiple-item scales . . . . The findings of our study lead us to recommend informed and selective use of the ratings available from the DOT . . . but we nonetheless endorse their continued use. Such use will yield additional information about other aspects of the measurements properties of these data, in particular their construct and predictive validity in problem-specific applications.

In sum, the DOT data suffer from nontrivial sampling and measurement problems, and the NRC/NAS Committee report recommends specific efforts to reduce these problems in later editions. The uniqueness of the data and the degree to which the DOT accomplishes an almost impossible task commends the continued use of the DOT. The issue for researchers in the area of health is what to use and how to use them.

The ratings in table 2 have been factor analyzed (57, 58), with results shown in table 3. These results can be compared with the six dimensions of demands and conditions of work delineated earlier as priorities for NCHS assessment. Factor 6 in table 3 covers key aspects of physical-chemical-biological hazards, with additional ratings relevant to this domain shown in table 2. Factor 3 provides a measure of physical demands, although the degree to which it represents potentially beneficial aerobic activity as opposed to potentially deleterious constraining and anaerobic demands is uncertain. Both of these sets of DOT ratings are clearly relevant for NCHS and other health-related research.

The DOT has little to say about job insecurity, although a DOT-type measure of job insecurity might be derived from unemployment rates for these occupations. Similarly, the ratings have little to say about social support and cannot say much, although they might include ratings of the extent of contact with other people. The ratings of complexity of work with people and interpersonal skills should be explored for their relevance as proxies for social support (or at least social isolation).

Aspects of job demands and worker control are interspersed throughout factors 1 and 4 in table 3. Demands are reflected in one item (PUS) in table 2, which does not load on any of the factors in table 3. In the work of Kohn and Schooler (60, 61), substantive complexity is related to a more general concept of autonomy or self-direction in work. Thus, the DOT may contain reasonable proxies for control. Improved assessment of psychosocial and, to a lesser extent, physical work demands would be necessary, however, before the DOT becomes fully appropriate for purposes of assessing job demands and conditions relevant to health.

Recommendations

The DOT represents a useful existing resource for NCHS studies and could be an even greater resource if future editions were modified to reflect health concerns. Thus, we recommend that:

Table 3. Factor Analysis of *Dictionary of Occupational Titles* occupational characteristics: Items and lodings for 6 factors

	Factor 1: Substantive complexity,[1] 49.3% [2]	
GED	General educational development	.86
SVP	Specific vocational preparation	.86
INTELL	Intelligence [3]	.83
DATA	Complexity of functioning with data [3]	.81
REPCON	Repetitive or continuous processes	−.81
NUMER	Numerical aptitude [3]	.78
VERBAL	Verbal aptitude [3]	.76
ABSTRACT	Abstract and creative vs. routine, concrete activities	.68
MVC	Measurable or verifiable criteria	.64
CLERICAL	Clerical perception [3]	.64
SPATIAL	Spatial perception [3]	.55
PEOPLE	Complexity of functioning with people [3]	.47
FORM	*Form perception* [3]	.46
TALK	Talking	.44
DCP	Direction, control, and planning	.43
VARCH	Variety and change	.42
DATACOM	Communication of data vs. activities with things	.41
	Factor 2: Motor skills, 22.6%	
FINGDEX	Finger dexterity [3]	.69
MOTOR	Motor coordination [3]	.68
MANDEX	Manual dexterity [3]	.67
THINGS	Complexity of functioning with things [3]	.66
FORM	Form perception [3]	.52
SPATIAL	Spatial perception [3]	.47
SEE	Seeing	.43
REACH	Reaching	.42
STS	Set limits, tolerances, or standards	.37
MACHINE	Activities involving processes, machines vs. social welfare	.33
	Factor 3: Physical demands, 9.9%	
LOCATION	Outside working conditions	.67
STOOP	Stooping, kneeling, crouching, crawling	.53
EYEHAND	Eye-hand-foot coordination [3]	.52
CLIMB	*Climbing, balancing*	.49
STRENGTH	Lifting, carrying, pulling, pushing	.48
	Factor 4: Management, 5.4%	
DEPL	Dealing with people	.78
DCP	Direction, control, planning	.74
PEOPLE	Complexity of functioning with people [3]	.70
TALK	Talking	.64
TANGIBLE	Activities resulting in tangible satisfaction vs. prestige	−.63
SCIENCE	Scientific, technical activities vs. business contact	−.57
DATACOM	Communication of data vs. activities with things	.49
DATA	Complexity of functioning with data [3]	.44
	Factor 5: Interpersonal skills, 4.9%	
SJC	Sensory or judgmental criteria	.51
FIF	Feelings, ideas, facts	.41
INFLU	Influencing people	.41
MACHINE	Activities involving processes, *machines vs. social welfare*	−.37
	Factor 6: Undesirable working conditions, 2.9%	
HAZARDS	Hazardous conditions	.52
ATMOSPHR	Fumes, odors, dust, poor ventilation	.42
HEAT	Extreme heat	.37

[1] This title is borrowed from Kohn (1969), who uses it to describe a conceptually similar but not operationally identical occupational dimension.
[2] Percentage of common variance explained.
[3] Sign reflected on this variable.
Source: Adapted from P.S. Cain and D.J. Treiman: The Dictionary of Occupational Titles as a source of occupational data. *American Sociological Review,* 46, 253–278, 1981. (Copyright; used by permission.)

1. Detailed occupational information should be collected in all NCHS surveys to allow all occupations to be coded into the DOT as well as Census codes.
2. Analyses of NCHS data should explore the predictive and construct validity of existing DOT characteristics in relation to health.
3. NCHS should negotiate with the Department of Labor to get improved ratings of the six dimensions of work demands and conditions identified here as central to studies of activity, fitness, and health.

Finally, we will recommend that multiple titles covering various housework or homemaking roles be introduced into the DOT.

Alternatives to the DOT

At the present time there is no alternative to the DOT for observation of job demands and conditions for national samples of occupations. There is, however, a variant of the DOT, keyed to Census occupational codes, developed by Karasek et al. (51) using *aggregated self-reports or descriptions* of jobs by workers themselves. (See cell B of figure 1.) We will turn shortly to the Karasek procedure.

The utility of the DOT (and the Karasek aggregate job description method) varies directly with the breadth and extensiveness of the jobs and occupations under study. As the focus of research moves from national samples of individuals and occupations to geographically, organizationally, or occupationally limited samples of jobs and individuals, alternative procedures become more feasible and even advisable. Neither the DOT nor the Census classifications of occupations will adequately describe the full variation in types of jobs in a particular community, organization, or occupational group.

Thus, community or organizational studies of job demands and conditions and health may need to develop new job classifications and hence new ratings of them. The same would be true for studies of broad occupational groups such as nurses, police, or teachers.

Development of new rating scales and procedures will also improve the relevance of the ratings of job demands and conditions to the unique occupational health problems of such populations. House (71) provides an example of the development and assessment of unique job classifications and ratings in a study of working conditions and health in the rubber industry, and the University of Michigan Survey Research Center (49) has reported results of an extensive project on external observation of work environments, derived from earlier work by Hackman and Lawler (72). Those who are willing and able to develop new ratings procedures tailored to specific community, organizational, or occupational groups should consult these sources.

It is in the interest of users of the DOT and related systems to encourage such developmental research and to urge that standard DOT ratings be included in such research for comparison with the rating methods developed for the population. Such studies of more limited populations can hence become a major mechanism for better understanding and improving the validity and reliability of the DOT or any other system applicable to national samples.

Aggregated job description: Alternative or complement to observation of demands and conditions of work

It is often assumed that ratings of job demands and conditions by external observers are somehow more objective than workers' own ratings or descriptions of their jobs. On many grounds, it is reasonable to assume that workers are, in fact, better able than external observers to rate or describe the demands and working conditions associated with their jobs. A worker, after all, has much more extensive and detailed knowledge of his or her job than any observer can obtain in a brief period of time. When an external rater and a worker disagree in describing a job, the external rater is no more likely to be right than the worker. Thus, workers can provide accurate and objective information on their jobs and represent an alternative, often a more economical one, to external raters as a source of information on demands and conditions.

The problem with workers' ratings or descriptions of their own jobs is not that they are in any absolute sense less accurate or more biased than external raters. Both sources of ratings are likely to contain errors and biases. The problem arises when the errors and biases in the measures of job demands or conditions are correlated with errors and biases in the other measures of variables (e.g., health) to which the job demands and conditions variables are being related. Such correlated errors are more likely to occur if all measures come from the same source or method—for example, if all are based on self-reports of individuals (24, 26). Thus, at least in the area of health research, there is a continuing need for measures of job demands and conditions independent of the individual worker. It is equally important, however, to utilize research designs that allow us to relate individual perceptions of jobs to individual health while minimizing such contamination or correlated error.

Measures of job demands and conditions independent of the individual worker can be derived, however, by aggregating the self-reports of a number of workers in the same job or occupation. In line with a long tradition in sociological research (73–75), Karasek et al. (51) have pursued such a strategy for the assessment of job demands and conditions. Recognizing that measures of job demands and conditions of work have not been, and often may not be, collected in major studies of health, they developed a method for imputing such measures to any data set in which individuals' occupations are coded into the three-digit U.S. Census Occupation and Industry Codes.

Karasek et al. began their work from a desire to develop a methodology for linking data on job demands and conditions of work to existing data bases on cardiovascular disease or other health outcomes that did not contain directly collected extensive occupational data on respondents. They utilized a set of major national sample surveys of working conditions of the U.S. labor force—the 1979, 1972, and 1977 Quality of Employment Surveys (QES), conducted by the University of Michigan Survey Research Center for the U.S. Department of Labor—to compile mean scores on job characteristics for a large number of occupations in the 1970 U.S. Census classification. For each occupation, a score on each job characteristic was generated by taking the mean of the responses of all workers in that occupation to a question or questions about their job demands or conditions. Separate scores were completed for men and women, covering about 250 occupations and industries (almost all with $N > 3$). These are aggregated into 85 occupational groups and 65 industrial groups containing at least 25 workers per group. The occupations for which Karasek et al. provided scores cover 95 percent or more of the labor force.

Table 4 shows the items and indexes (subconcepts and concepts) for which mean occupational scores are available from the work of Karasek et al. It is immediately evident that these measures cover all of the major physical and psychosocial job demands and conditions that we feel must be central to future research on work and health. Actual item wordings are presented in the appendix and discussed more fully in the text under self-report job descriptions.

Karasek et al. (51) report a great deal of information on the reliability and validity of these measures. For multiitem indexes, the individual-level internal consistency reliability coefficients range from .58 to .84 for the pooled QES data. The mean scores for occupations correlate .8 and up across survey years. The variance in individual scores among the occupational groups ranges from 2.1 percent in the case of coworker support to 33.9 percent in the case of skill discretion. Variation among occupations is greatest (20% or more of the variance) on the measures of decision latitude, hours, and physical exertion. It is moderate (10–20% of the variance) for hazards and relatively low (under 10%) for the other variables.

Table 5 shows the correlation across Census occupations (weighted by occupational size) between the Karasek et al. scores and comparable DOT scales as reported by Karasek et al. (51). With the exception of a few dichotomies that are very skewed on the DOT, the correlations range from almost .5 to almost .8, providing substantial evidence of criterion validity. Evidence of construct validity vis-a-vis health is provided by analyses of the relationship between selected job characteristic measures in table 4 and data on prevalence of myocardial infarction (MI) among white males obtained in the National Health and Nutrition Examination Survey (NHANES) of 1971–74. Consistent with other research of Karasek and associates in the United States and Sweden, high psychological job demands and low deci-

Table 4. Job characteristic dimensions assessed by Karasek et al. (1982) via individual and aggregated job descriptions

General concept	Subconcept	Brief description of constituent items
Decision latitude (control)	Skill discretion	Develop skills Learn new skills Creative job Variety Skill level Repetitive work
	Decision authority	Freedom as to how to do work A lot of say [1] Ability to make decisions [2]
Job demands	Psychological workload	No excessive demands [3] No conflicting demands [3] Have enough time to do job [3] Work hard Work fast
	Hours worked	Single item on hours worked
Job security		Steady job Good job security Likely to be laid off [4]
Social support	Supervisor support	Supervisor shows concern for Supervisor pays attention to Supervisor helpful getting work done [1] Supervisor creates good teamwork [1]
	Coworker support	Coworkers friendly Coworkers helpful Coworkers personally interested [1] Coworkers competent [1]
Physical-chemical-biological hazards [5]	Hazardous conditions or exposures	Dangerous equipment Dangerous work methods Things placed or stored dangerously Fire, burns, shocks Dirty, bad maintenance
	Toxic exposure	Dangerous chemicals Air pollution (dust, fibers, etc.) Risk of catching disease
	Noise	Single item
	Temperature	Single item
Physical workload	Single item	Job requires regular physical exertion

[1] Not asked in 1969 Quality of Employment Survey.
[2] Not asked in 1977 Quality of Employment Survey.
[3] Item reversed in scoring.
[4] Asked only in 1977 Quality of Employment Survey.
[5] Asked in different form in 1977 than in 1969 and 1972.

sion latitude (and especially the combination thereof) are associated with greater MI prevalence. Physical exertion proved negatively associated with MI prevalence (6).

The Karasek et al. occupational measures improve on the DOT by being more directly relevant to health. They are available, however, for a smaller number of cruder Census occupational classifications, and in some cases the mean occupational scores are based on small numbers of respondents. As one moves from measures that are more descriptive of the external demands and opportunities of the job (e.g., decision latitude) to measures that are closer to worker behaviors (e.g., hours worked) or reactions (e.g., psychological workload, support), the relative amount of variance among occupations declines and the need for measures tied to the individual worker's job increases. Thus, we will now turn to such measures. Nevertheless, we would recommend that NCHS utilize the aggregated job description scores of Karasek in addition to the DOT worker trait measures.

Table 5. Validation of Quality of Employment Surveys (QES) job dimension scales with *Dictionary of Occupational Titles* (DOT) job scales

Job dimension scale	DOT scale	Correlation
1. Decision latitude	Occupational self-direction[1]	.765
	Data—complexity	.751
	People—complexity	.511
	General educational development	.756
	Specific vocational preparation	.693
	Routinization[2]	−.533
	Closeness of supervision[2]	−.712
a. Skill discretion	Occupational self-direction[1]	.751
	Data—complexity	.769
	People—complexity	.476
	General educational development	.792
	Specific vocational preparation	.759
	Routinization[2]	−.544
b. Decision authority	Occupational self-direction[1]	.695
	Closeness of supervision[2]	−.652
2. Psychological workload	Performance under stress[3]	[4].060
3. Physical exertion	Physical exertion[3]	.624
4. Noise exposure	Noise[3]	.485
5. Temperature extremes	Hot/cold[3]	[4].103
6. Hazard exposure	Hazards/toxic exposure[3]	.596

[1] This is Kohn's occupational self-direction scale as approximated by Temme (1975). See Spenner (1981), p. 244.

[2] These two scales are each constructed by Spenner from two worker trait indicators. See Spenner (1981).

[3] Scores are based on a single worker trait indicator from the DOT or our equally weighted sum of two such measures.

[4] Correlation is based on very skewed dichotomous variables, which tends to reduce the correlation. The percentages for the 1-category of the DOT scale are: Performance stress, 2%; heat, 1%; cold, 5%. These correlations are not significant.

Notes: Correlation coefficients between occupation-level scores from DOT (Spenner, 1981) and occupation-level scores from QES. Sample weighted by occupational size (2,671 total).

All correlations are significant at the .001 level except as noted.

Source: Adapted from R.A. Karasek, J. Schwartz, and T. Theorell: Job Characteristics, Occupation and Coronary Heart Disease. Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, 1982. (Copyright; used by permission.)

Self-report data on individual workers and jobs

Although ratings and aggregated descriptions of individual jobs can be highly useful in assessing relationships between job demands or conditions and health, such occupational data are necessarily imperfect measures of the demands and conditions associated with individual jobs. The best prediction of individual-level health outcomes should come from data on individual workers and jobs, necessarily derived (for reasons already discussed) from workers' self-reports about their jobs and their behavior and reactions to their jobs. This is confirmed by Karasek et al. (51), who compared the variance explained on an individual-level health outcome (depression) by both individual-level self-reports and occupation-based scores. The occupational scores explained only about one-third as much variance in health as the individual-level self-reports, although effects of occupation-based scores proved relatively unbiased (especially if controlled for sociodemographic variables such as age, race, and sex).

Worker behavior and reactions should also be more predictive of health than descriptions of more external demands or conditions of work. The health effects of external job demands and conditions, whether physical or psychosocial, are necessarily mediated through psychological and physiological responses of the individual. High levels of physical or psychosocial demands, for example, are likely to be detrimental to health only to the extent that these demands are experienced as strains by the individual. Physically fit or psychologically "hardy" individuals, for example, are likely to be less strained and hence to suffer less adverse health effects than less fit or hardy persons. Further, we would argue that it is sustained or chronic behaviors and reactions that are most consequential for health because the etiological processes converting physiological or psychological strain into enduring health problems are usually of substantial duration.

The Karasek et al. Job Content Survey

Karasek (76) has supplemented the core QES items described in table 4 with additional items to create a Job Content Survey, which can be completed by any employed individual. Twenty-seven items from this measure—nine items on *psychosocial job demands,* nine items on *control* and decision latitude, five items on *physical work demands,* and four items on *job insecu-*

rity—are being administered in a national followup study of the 4,000 offspring of the original Framingham Study of the U.S. National Heart, Lung, and Blood Institute. Fuller versions of the survey include additional questions in some of these areas, as well as 11 items on *social support* and 9 items on *exposure to physical, chemical, or biological hazards.* Thus, Karasek's Job Content Survey provides self-report measures for all of the six key content areas that should be included in surveys of job demands and conditions in relation to health. The items are generally aimed at descriptions of the job rather than descriptions of worker behavior or reactions, although the extent to which descriptions are contaminated by individual behaviors and reactions is uncertain. Karasek and colleagues have provided evidence relating most of these measures to some health outcomes, and further information will be forthcoming from the Framingham Study. We recommend inclusion in NHANES and the National Health Interview Survey (NHIS) of 24–48 items from the Job Content Survey assessing the major concepts listed in table 4.

The first major effect at developing self-report measures of working conditions was developed by Kohn and associates (60, 61). Their measures were developed for understanding occupational self-direction in relation to personality and thus do not cover the range of variables in the Karasek et al. measures and have not been validated in relation to physical health outcomes. Their measures do, however, index closeness of supervision and self-reliance in work more adequately than the Karasek et al. Job Content Survey, and the Kohn measures have been generalized to the domain of housework (discussed later). Thus, inclusion of some of the Kohn measures in the NCHS epidemiologic surveys would be desirable if resources allow.

Measures of behavior and reactions

Because the Job Content Survey is slanted somewhat toward job description, it would also be useful to incorporate in NCHS studies measures that more clearly tap worker behavior or reactions to their work. Such measures are among the strongest correlates, and even predictors, of health in existing studies of psychosocial job demands and conditions in relation to health (29, 77). In the area of behavior, the most likely candidate is the Type A behavior pattern, which is heavily laden with occupationally related behaviors (e.g., job involvement and competitiveness). The Type A behavior pattern represents a style of behavior in response to job demands and other factors that has been linked to coronary heart disease and some other outcomes. The original and still preferred method of assessing Type A is the structured interview, which is taped and rated by the interviewer or a trained coder (78, 79). The Type A interview is a strong candidate for inclusion in NHANES. For most other NCHS surveys, such as NHIS, a questionnaire measure of Type A is necessary. Bortner (80), Jenkins et al. (81), and Haynes et al. (82) have developed such measures, with the Haynes et al. measures having a combination of brevity, previous use in a general population survey, and clear construct validity in relation to health. On the other hand, other measures also have validity and are conceptually closer to the Type A construct. Some, such as the Jenkins Activity Survey, have specific job-related content.

The Framingham measure reflects not only behavior but psychological reactions to the job. Other such job reaction measures have been developed and related to health in studies by House with rubber workers and the Tecumseh Community Health Study (11, 32, 71, 77), and they provide some supporting evidence that it is mainly chronically elevated perceptions of job stress or strain that are consequential for health. An index of 5–10 items from the House et al. (11, 77) work would constitute a further useful measure for NHANES and NHIS, although we would expect fairly high intercorrelations between these and the Karasek et al. Job Content Survey measures of job demands.

Simultaneous inclusion in the NCHS survey of the Karasek items, a Type A measure, and some measure of perceived strain is highly desirable, both for understanding the relations among these measures and for evaluating their relation to activity and health.

A note on work history and the chronicity of job demands

We have noted at several points that it is chronic exposure to overly demanding hazardous or stressful work (and similarly to salutary physical activity) that is most likely to affect health. Thus, it is important to be able to relate history of exposure to job demands and conditions to health outcomes. This can be done prospectively by obtaining measures of job demands and conditions of work for the same individuals over a series of panel surveys and using the pattern of exposure to job demands and conditions evident in that data series to predict subsequent health outcomes (net of baseline health). In such a design, more subjective self-report variables can be used just as well as more objective occupational ratings because health measures and job condition measures can be obtained at different time points and using different methods (interviews versus exams), hence reducing or eliminating correlated errors. Retrospectively, it is possible to reconstruct job histories and, by linking them to occupational ratings from the DOT or Karasek et al. (25), to do similar analyses. Self-report data on job demands are less useful in such studies, because current reports or retrospections by individuals may be colored by their current health (which is what we seek to predict). Retrospective analyses of even occupational data, however, may encounter serious methodological problems because the selection of workers into and out of jobs may be a consequence of health rather than occupational conditions being a cause of health. Again, only prospective analyses can mitigate these problems.

Occupational demands and conditions related to housework

Although not a common subject of epidemiological inquiry, certain aspects of nonpaid household work have been at least provisionally linked to chronic forms of physical ill health (83) and, more convincingly, to mental distress (83–85). This is not to suggest that involvement with housework or household duties necessarily places the role holder at risk. Rather, as with paid employment, negative as well as positive effects may spring from this form of work. Thus, we feel that household work requires the same consideration as paid employment in assessing the health effects of demanding work. Our analysis suggests more parallels and similarities between unpaid household work and paid employment than we initially expected. We think it is also important, however, to briefly outline some ways in which household-related work may be conceptually differentiated, and perhaps more importantly, experientially differentiated from paid employment. We turn first to the social context of such activity.

The social context of housework

Housework and the attendant responsibilities of child rearing have been tasks that have largely been assigned to women, whereas males have predominated in many areas of paid employment. Further, work relations within oriented places of paid employment tend to be framed by formal, relatively narrowly defined sets of rights and responsibilities, whereas work in the household is more likely to occur within a broadly defined, relatively personal network of familial relationships. Work relations in this latter context are, then, more likely to be marked by emotional as well as functional interdependence. Not surprisingly, the interpersonal relationships found within the household are generally more intense that those in the workplace (86).

Although we fully expect there to be some division of labor within all households and varying degrees of centralization of authority and responsibility for decisionmaking, we also expect these distinctions to be less formal and less explicit than those found in work establishments in the marketplace. In addition, the criteria for establishing what constitutes successful or even superior enactment of the role of houseworker are not universally clear or accepted.

This last point is, of course, related to one of the most profound differences between employment in the public and private sectors of the economy and gainful activity as a houseworker: People in the former roles are paid directly for their services; those in the latter are not. Along with the impossibility of job promotion or career advancement in such a situation, this suggests that our society does not yet fully recognize the role of houseworker as legitimate employment.

With the foregoing in mind, we might conclude that the conditions and demands associated with paid employment and unpaid housework are profoundly different. This, we suggest, is simply not so. A number of researchers have called our attention to the parallels between these two domains (87–89). More recently, an empirically based comparative study of the procedures and working conditions underlying paid employment and nonpaid housework found that all but one of the demands and conditions they assessed were within a single standard deviation of women's paid employment (86). (The exception was that of the substantive complexity of work with things, which was higher for women involved in housework.) In fact, several dimensions were nearly identical to conditions and procedures characteristic of paid employment. These researchers observe that housework should not be considered a unique work experience and conclude that "at least in terms of the dimensions we can compare, if housework were a paid job, it would not be a particularly unusual one" (86, p.106).

Operational definition of housework

One of the difficulties involved in the assessment of the job-related demands faced by the houseworker is establishing clear boundaries around what is to be considered housework. For the purposes of assessment, we propose that housework be conceptualized as involving only those jobs that are imperative to the effective functioning of the household. This reasonably includes jobs like shopping for essential household supplies, meal preparation, cleaning chores, and the repair and maintenance of household items as well as the family home and immediate grounds. Along with Schooler and Miller (86), we suggest that child care responsibilities be distinguished from the more general parental socialization of children, with only the former category of activity included within the domain of housework.

Although difficult to establish in many cases, we think it important to also exclude those activities that are not essential to the maintenance of the household, including discretionary elaborations of the foregoing jobs. Consequently, activities like gardening projects that are not necessary for the maintenance of sound familial nutrition or carpentry work that reflects a need for self-expression would be excluded from an assessment of the time investment and activity involvement of houseworkers.

Assessment of demands and conditions of housework

In the interests of clarity of presentation and continuity with earlier sections of this paper, we will allow the more general dimensions of the typology of working conditions and modes of assessment presented earlier in this paper (figure 1) to guide our discussion of the assessment of job-related demands and conditions characteristically encountered by houseworkers. Our general strategy is first to apply standard measures of job demands and conditions to houseworkers, and then to supplement these with measures that capture unique aspects of housework.

Observations or ratings of occupations

We are clearly proposing that the role of houseworker be included in future efforts to assess the job demands of important work roles in the economy. Such data collection should, we think, proceed in ways consistent with our earlier recommendations for the development of aggregated ratings of jobs and the subsequent rating of an occupational category. Multiple ratings for the houseworker role may be in order, to reflect important contextual factors, such as the presence of young children in the home or that of paid domestic help, that we think will significantly influence the relevant demands and conditions facing the houseworker. As was mentioned earlier, such occupational ratings might then be assigned to individuals within samples of interest so that they may be used to predict individual-level health outcomes.

The *Dictionary of Occupational Titles*

Because the DOT represents the only comprehensive national effort to observe and rate the external demands and conditions of different occupations in the United States, it should assess similar dimensions of the occupation of houseworker. This seems both appropriate and feasible, given that the mandate of the DOT has grown appreciably since it was undertaken 40 years ago and given that millions of American adults are engaged as houseworkers on a full-time, frequently career-length, basis.

A number of surmountable challenges may be anticipated in such an effort. For example, negotiating entry into a meaningful sample of American households may depart somewhat from the ordinary questions of entry into private firms, but they appear no more problematic than those encountered by DOT analysts in their survey of small industrial enterprises (63) or those commonly encountered and successfully resolved by social and behavioral scientists (90, 91). Further, there is obviously a need for securing samples of households that are representative of the geographical and major sociocultural areas of the country, as well as different economic strata in our society. However, the field offices of the data gathering arms of the Department of Labor have already been established in the major geographical areas of this country (63).

The manner in which the DOT is presently organized at least minimally covers several of the six dimensions of work demands and job conditions that we outlined earlier. However, as we then noted, there is room for real improvement in the areas of psychosocial job demands, work-related social support, and job uncertainty and loss, particularly in the assessment of the houseworker role. In fact, this last category of working conditions may be best pursued through self-report measures, a point that will be pursued more fully in the next section.

In sum, we recognize that the Department of Labor is in a strategic position to fruitfully undertake the external assessment of the occupational characteristics of the houseworker role through the DOT, and we recommend that they be encouraged to do so.

Self-report measures of demands and conditions of housework

Physical-chemical-biological hazards associated with housework

Although a large number of accidents of low-to-moderate severity occur within households during the course of a year, we are unaware of any coherent body of research literature that has examined configurations of health hazards encountered by nonpaid houseworkers. However, indexes of such hazards are present in the U.S. Quality of Employment Surveys (92–94) and in Karasek et al.'s Job Content Survey (76). The items composing these indexes appear readily adaptable to the study of working conditions within the household and, at least in the realm of paid employment, they appear to have acceptable levels of reliability.

Physical activity and housework

Because so little empirical evidence now exists concerning the range and intensity of physical activity in this very common mode of work, we suggest that such data be developed and that self-reports of physical exertion among houseworkers represent a relatively economical way to gather such information. Still there are a number of issues that are worth attending to in this area.

On the positive side of things, little innovation may be necessary in the development of new survey measures of activity levels in housework because current scales incorporate many items that appear germane to activity in this sphere—e.g., walking on the level or up stairs, lifting light objects, operation of machinery or appliances, gardening or household maintenance (95). However, if scores from scales of this sort are to be used as predictors of health outcomes, they will need to be assessed in relation to activity levels existent in other roles along with other known risk factors, including diet and health behaviors (96). Finally, although it may seem on the face of it that housework will require a relatively stable level of exertion for all participants, accumulating evidence suggests that even with relatively sedentary occupations, the variance in physical activity levels within an occupation may exceed the range that exists between different occupations (16). As with other occupations, we would be wise to pay careful attention to adequately assessing the range of activity levels characterizing individuals engaged in housework.

Job insecurity and loss (analogs in the role of houseworker)

The feelings of uncertainty accompanying the threat of job loss or the acute sense of psychological distress associated with its actuality (97, 98) are ordinarily associated with the sphere of paid employment, but what of the sense of dependence and feelings of uncertainty and distress (87, 99) experienced by unpaid houseworkers who are dependent on their gainfully employed spouses for financial support and even their social standing (100, 101)?

The literatures on employment and sex roles not infrequently comment on the psychological distress experienced by full-time houseworkers in connection with the threat of their spouses' loss of employment or the prospect of the loss of financial support and social status because of divorce from or death of their mates (102, 103). Although not a standard part of epidemiological surveys of health, considering the addition of appropriate items to studies in this area would go a long way toward establishing the health risks associated with resource dependence of this kind.

Psychosocial demands of housework

Apart from relatively global analyses of the role of housewife entailing high inference measures of the demands confronting the houseworker (83, 104, 105), there has not been a great deal of epidemiological research on whether psychosocial demands of housework are consequential for health and which demands are consequential. However, there are a growing number of exceptions to this rule.

Kohn and Schooler's work (60, 106, 107) over the last 13 years has been oriented toward identifying those dimensions of the work process that appear to influence an individual's psychological functioning. For the purposes of the present discussion, their measures of work pressures may prove to be important additions to epidemiological surveys. Interestingly, Schooler et al. (86) identify not only time pressures but the dirtiness and heaviness of housework, as well as being responsible for things outside one's control, as constituting work pressures. Their findings indicate that all of these housework-related pressures were predictive of psychological distress for men, and all indicators except heaviness of physical labor were predictive of distress for women (86).

Time-use data indicate one type of condition under which a houseworker may experience time pressure and eventual distress (table 1). Housewives employed outside the home do not decrease their involvement with housework at a level that is commensurate with their work involvement. Rather, they simply work more hours (88, 91). This is not inconsistent with Gove's (83, 104) somewhat controversial thesis that the higher incidence of depression among women is linked to chronic, low-level fatigue and ill health associated with overwork.

An increasingly sophisticated literature has linked the experience of role overload and role conflict in household work to mental health (83, 99, 100, 104, 105, 108). For example, in an epidemiological study of family roles and depression, Aneshensel (99, 108) shows that employed men and women who are in the high-demands condition of parenthood (i.e., young children present in the home) are more likely to be depressed than couples without children. Consistent with what has been observed elsewhere, women in this category tended to report higher levels of depression than males. Although one might expect some aspects of the parenting role to help compensate for the increased time spent on household duties when children are present in the home, there is growing evidence that the role overload and role conflict associated with substantial involvement with paid employment and housework, or in the houseworker role singly (105), may engender significant levels of psychological distress. We suggest, therefore, that researchers consider the inclusion of such measures in epidemiological studies of mental health or those studies focusing on possible links among role demands, activity levels, and physical health.

Control and housework

In their investigation of the nature and consequences of housework, Schooler and his colleagues (86) found that among the aspects of the houseworker role that they assessed, the strongest predictor of psychological distress among women was the "frequency of being held responsible for things outside one's control" (table 4 in reference 86). Although study of the direct relationship between personal control and mental health among houseworkers has only recently begun, we think that, taken with the large body of findings from research concerned with personal control and stress in other life roles, it is quite possible that this variable will prove to be an important predictor of mental and even physical health outcomes among houseworkers, particularly if such measures are carefully formulated to reflect the responsiveness of the immediate environment to the individual's needs for control.

Social support

A number of researchers have noted that people engaged in higher levels of housework will eventually experience some form of relative chronic overload, especially for adults employed both in and outside the home (83, 108), and that certain jobs within this domain are noncomplementary and stress inducing (105). Consequently, we suggest that both the main and buffering effects of social support (31) are germane to epidemiological studies of housework and health.

More specifically, in certain situations, social support may help maintain or even enhance health and well-being by meeting basic human needs or obviating the possibility of role overload through the provision of appropriate instrumental support, as when a person's spouse renders significant aid with child care or household maintenance. On the other hand, when an individual experiences stress associated with the sustained enactment of adult responsibilities in several domains of life, the presence of a caring mate offering understanding and affection may help reduce the tensions of the day and facilitate the stressed individual's efforts to regroup.

Of course, a houseworker's mate may not be the only viable adult source of instrumental or socioemotional support. Friends and relatives may be called on to render such services if at least their physical proximity makes this possible. It should also be noted that there is increasing evidence that it is the perceived quality of the support rendered rather than strictly the quantity of social support received that is most beneficial to people in stressful situations (109).

In a potentially useful departure from the direct assessment of the receipt of social support, Gove (110) focused more on what he argues is the relatively common occurrence of the *social isolation* of full-time houseworkers from the mainstream of adult society. It is suggested that this makes possible a "self-defeating self absorption" with regard to personal problems and the diminished possibility of receiving timely social support from other adults. Aneshensel and her colleagues (108) are able to more directly assess the relevant effects of social support on one health outcome, the incidence of depression. Although not detailing the exact measures used, these researchers present strong evidence that women with major responsibilities in the home who also report low levels of social support are especially vulnerable to depressive symptoms.

Strategies for the assessment of the social support available to houseworkers through self-report measures may be borrowed from a number of sources. Published accounts of Gove's (110) measures of social isolation and Aneshensel's (108) measures of social support are instructive. In addition, measures may be borrowed and adapted from those developed by researchers involved in the study of paid employment and stress (6, 31, 36, 71). For example, Karasek (6) operationalizes social support as the overall helpfulness of one's coworkers and supervisors. His secondary analyses of the Quality of Employment Surveys (92–94) used two subscales, which may be equally appropriate to housework, involving the availability of instrumental and socioemotional support.

In sum, we strongly urge that researchers involved in epidemiological studies of the demands and conditions of housework in relation to health seriously consider inclusion of social support measures in their surveys.

Measures of behaviors and reactions in housework

As noted in our discussion of paid employment, it is desirable to supplement measures, such as the Job Content Survey, that focus on job descriptions with measures that more clearly tap workers' behavior or reactions to their work. This holds for housework as well as paid employment. Both the measures of Type A behavior pattern and the measures of sustained affective tension and arousal in reaction to work, discussed in the section on paid employment, seem applicable to housework. For houseworkers who also hold paying jobs, these might be supplemented by a set of items focusing on affective reactions to the demands (and conflicts) of combining paid employment and housework.

Recommendations and conclusions

We have argued that housework is, for the most part, not conceptually or experientially distinct from paid employment. Consequently, we expect housework to be as consequential for certain health outcomes as is paid employment. It is important to keep in mind, however, that the houseworker role may or may not be configured with other roles. As the previous discussion suggests, this fact will have important implications for the stressors encountered by people engaged in housework. Thus, in formulating the measures for the assessment of the foregoing demands and conditions, the researcher should consider whether or not health-related interactive effects are anticipated between stressors associated with the houseworker role and those associated with employment, marital, or parental roles. If so, items assessing stresses and strains in similar dimensions in other roles must be included in surveys relying on self-report measures. Interaction items would be subsequently entered into analyses examining the predictive relationship between stressors and health outcomes. An alternative strategy is simply to tap the subjective assessments of respondents regarding such interrole stress effects. There are, of course, tradeoffs between the two approaches and no universal way of resolving them. Our recommendations focus on assessment of work demands and conditions of housework, but these must be considered in the context of the demands and conditions of other jobs held simultaneously by the houseworker.

Development of measures of the demands and conditions of housework should proceed as follows:

1. Measures designed for workers in paid employment should, wherever feasible, be applied to housework. We see no reason, for example, that the *Dictionary of Occupational Titles* classification and rating system cannot be extended to housework. With minor wording changes, most of the Karasek Job Content Survey can also be applied to houseworkers. The measures of Schooler et al. (86) also deserve consideration because they have already been applied to both housework and paid employment. With the exception of measures of job insecurity and social support, the measures of Type A behavior pattern and of subjective reactions to work discussed under paid employment seem equally applicable to housework.
2. Existing measures of demands and conditions of paid employment are likely, then, to adequately assess physical activity, psychosocial demands, physical-chemical hazards, and control in housework. Special instruments are needed, however, to assess the following aspects of housework:

- *Job insecurity and loss*—As noted earlier, the health effects of the dependence of houseworkers on their spouses for financial security and social status do not appear to have been thoroughly investigated. In fact, the stress associated with such uncertainty may not reach significant levels until the spouse's job is in jeopardy or the houseworker's marriage is threatened. However, we think that consideration should be given to the inclusion of a small number of items of this sort in epidemiological surveys of occupational stress. Such questions should probably focus on both the intensity of the uncertainty and distress associated with resource dependence of this kind and the incidence of such experiences.
- *Social support*—Of the psychosocial determinants of health outcomes, we would argue that social support is one of the most important. We recommend, therefore, that social epidemiological studies of health include measures appropriate for the assessment of the social support available to houseworkers. A number of relatively general scales already exist that appear adaptable for such purposes.

  As mentioned earlier, the efforts of Gove (110) and Aneshensel (108) have been directed toward the assessment of important aspects of social support. Caplan et al. (111) also developed a 12-item measure of perceived social support from work supervisors, coworkers, and persons outside of work with respect to work-related problems. House (71) adapted and extended these measures, enabling him to distinguish the support of a respondent's spouse from friends and relatives. This last point seems crucial to studies of housework, as is our previous point that it is also important that the quality of support be distinguished from the quantity of support received (109).

Summary and recommendations

This paper discusses the measurement and analysis issues in assessing the physical and psychological demands of work on and off the job in epidemiologic studies of physical fitness and activity level among the general population. Because work activity occupies almost half of most adults' working hours, the physical and psychosocial aspects of work activity are closely related, and both have been shown to be consequential for health, we argue that NCHS general population surveys should include both individual self-report data and occupation-based ratings of the external demands and worker reactions to physical and psychosocial demands and conditions of work, whether that work is paid employment, unpaid housework, or both.

The paper develops a typology of measures of job demands and conditions. The typology is defined by three variables: (a) The *object* of assessment (individual workers or jobs versus groups of jobs or workers, usually termed occupations); (b) the *source* of the assessment (self-reports of workers versus observations or ratings by persons external to an individual worker or job); and (c) the *content* of the assessment (external conditions of work versus worker behavior versus worker reactions). Observations of individual workers or jobs are unfeasible. Therefore, the paper recommends that NCHS collect data on *observations or ratings of occupations* and *self-report-based ratings and descriptions* of individual jobs. Six types of job demands and conditions are important: Physical-chemical-biological hazards; physical activity at work; job insecurity and loss; psychosocial job demands; worker control, autonomy, and participation; and work-related social support. Specifically the report recommends inclusion in NCHS studies of the following measures:

1. Both U.S. Census Bureau and *Dictionary of Occupational Titles* codings of occupations. These can be linked to external ratings of job conditions developed by the DOT and to a set of ratings based on aggregated self-report data developed by Karasek et al. (51) for Census occupations. An occupational history would then allow retrospective analyses of the impact of lifetime job experiences, although such retrospective analyses have methodological problems because of the impact of health on selection into and out of jobs.
2. Portions of the Karasek et al. (51) Job Content Survey, which provides individual self-report descriptions of workers' jobs for the six types of job demands and conditions specified earlier.
3. Measures of the Type A behavior pattern (structured interviews for NHANES and the Framingham Type A scale for NHANES and NHIS).
4. Measures of workers' subjective feelings of pressure, tension, or stress in work derived from measures used in Framingham, Tecumseh, and studies by House (71, 112, 113).

Measures of unemployment experience and overtime should also be included. These same measures should, for the most part, be applicable to housework. Efforts should be made to get the DOT to include housework as well as health-related dimensions of work. Self-report measures of demands and conditions of housework need to focus on resource dependency as an analogy of job insecurity in paid employment and on social support from spouses, friends, and relatives as well as co-workers. The impact of combinations of paid employment and housework can be assessed both by looking at the combined additive and interactive effects of demands and conditions of both types of work and by direct questions on the special demands posed by combining paid employment and housework. Finally, attention should be paid to work history and the chronicity of job demands.

References

1. Juster, T., and Stafford, F. P. (eds.) (1984). *Time, Goods, and Well-Being.* Ann Arbor, MI: Institute for Social Research, University of Michigan.
2. Department of Health, Education, and Welfare (DHEW) (1973). *Work in America.* Cambridge, MA: MIT Press.
3. Kagan, A. R., and Levi, L. (1974). Health and environment—Psychosocial stimuli: A review. *Social Science and Medicine,* 8, 225–241.
4. Kahn, R. L. (1981). *Work and Health.* New York: Wiley.
5. House, J. S., and Cottington, E. M. (1986). Health and the workplace. In Linda Aiken and David Mechanic (eds.), *Application of Social Science to Clinical Medicine and Health Policy.* (pps. 312–416.) New Brunswick, NJ: Rutgers University Press.
6. Karasek, R. A., Schwartz, J., and Pieper, C. (1983). Validation of a survey instrument for job-related cardiovascular illness. New York, NY: Department of Industrial Engineering and Operations Research, Columbia University.
7. Ashford, N. A. (1976). *Crisis in the Workplace: Occupational Disease and Injury.* Cambridge, MA: MIT Press.
8. Rom, W. N. (ed.) (1983). *Environmental and Occupational Medicine.* Boston: Little, Brown & Co.
9. Selikoff, I. J., Hammond, E. C., and Churg, J. (1968). Asbestos exposure, smoking, and neoplasia. *Journal of the American Medical Association,* 204, 106.
10. Andersson, G. B. J. (1981). Epidemiologic aspects on lower-back pain in industry. *Spine,* 6(1), 53–60.
11. House, J. S., McMichael, A., Wells, J. A., Kaplan, B. H., and Landerman, L. R. (1979). Occupational stress and health among factory workers. *Journal of Health and Social Behavior,* 20 (June), 139–160.
12. Pasmore, W., and Friedlander, F. (1982). An action-research program for increasing employee involvement in problem solving. *Administrative Science Quarterly,* 27, 343–362.
13. Oberman, A. (1980). The role of exercise in preventing coronary heart disease. In E. Rapaport (Ed.) *Current Controversies in Cardiovascular Disease.* Philadelphia: W. B. Saunders.
14. Oglesby, P. (1980). Exercise and the prevention of coronary artery disease: The evidence is inconclusive. In E. Rapaport (Ed.) *Current Controversies in Cardiovascular Disease.* Philadelphia: W. B. Saunders.
15. Eichner, E. R. (1983). Exercise and heart disease. *The American Journal of Medicine,* 75, 1008–1023.
16. Rigotti, N. A., Thomas, G. S., Leaf, A. (1983). Exercise and coronary heart disease. *Annual Review of Medicine,* 34, 391–412.
17. Haw, M. A. (1982). Women, work and stress: A review and agenda for the future. *Journal of Health and Social Behavior,* 23, 132–144.
18. House, J. S. (1974a). Occupational stress and coronary heart disease: A review and theoretical integration. *Journal of Health and Social Behavior*, 15, 12–27.
19. House, J. S. (1974b). Occupational stress and physical health. In J. O'Toole (ed.), *Work and Quality of Life: Resource Papers for "Work in America."* Cambridge, MA: MIT Press.
20. House, J. S., and Jackman, M. (1979). Occupational stress and health. In P. Ahmed and G. Coelho (eds.), *Toward a New Definition of Health* (pps. 135–158). New York: Plenum Publishing Co.
21. Jenkins, C. D. (1971). Psychologic and social precursors of coronary disease. *New England Journal of Medicine,* 284, 244–255; 307–317.
22. Jenkins, C. D. (1976). Recent evidence supporting psychologic and social risk factors for coronary disease. *New England Journal of Medicine,* 294, 987.
23. Kasl, S. V. (1974). Work and mental health. In J. O'Toole, *Work and the Quality of Life* (pp. 171–196). Cambridge, MA: MIT Press.
24. Kasl, S. V. (1978). Epidemiological contributions to the study of work stress. In C. L. Cooper and R. Payne (eds.), *Stress at Work* (pp. 3–48). New York: Wiley.
25. Karasek, R. A., Baker, D., Marxer, F., Ahlbom, A., Theorell, T. (1981). Job decision latitude, job demands, and cardiovascular disease: A prospective study of Swedish men. *American Journal of Public Health,* 71(7), 694–705.
26. Kasl, S. V. (1981). The challenge of studying the disease effects of stressful work conditions. *American Journal of Public Health,* 71, 682–684.
27. Cobb, S., and Kasl, S. V. (1977). Termination: The consequences of job loss. U. S. Department of Health, Education, and Welfare, HEW Publication No. (NIOSH) 77–224.
28. Gordus, J. P., and McAlinden, S. (1984). Economic change, mental illness, physical illness and social deviance. (Joint Committee of the U.S. Congress.) Washington, DC: U.S. Government Printing Office.
29. Jenkins, C. D. (1983). Psychosocial and behavioral factors. In N. M. Kaplan and J. Stamler (eds.), *Prevention of Coronary Heart Disease* (Ch. 8, pp. 98–112). Philadelphia: W. B. Saunders.
30. Friedman, M. (1984). The history of an idea. In M. Friedman and D. Ulmer (eds.), *Treating Type A Behavior—and Your Heart* (pp. 3–36). New York: Alfred A. Knopf.
31. House, J. S. (1981). *Work Stress and Social Support.* Reading, MA: Addison-Wesley.
32. House, J. S., Robbins, C., and Metzner, H. M. (1982). The association of social relationships and activities with mortality: Prospective evidence from the Tecumseh Community Health Study. *American Journal of Epidemiology,* 116, 123–140.
33. Ruberman, W., Weinblatt, E., Goldberg, J. D., and Chaudhary, B. S. (1984). Psychosocial influences on mortality after myocardial infarction. *The New England Journal of Medicine,* 311(9), 552–559.
34. House, J. S., and Wells, J. A. (1978). Occupational stress, social support and health. In A. McLean, G. Black, and M. Colligan (eds.), *Reducing Occupational Stress:* Proceedings of a conference. U.S. Department of Health, Education, and Welfare, HEW Publication No. (N1OSH) 78–140.
35. Karasek, R. A., Triandis, K. P., and Chaudry, S. S. (1982a). Coworker and supervisor support as moderators of associations between task characteristics and mental strain. *Journal of Occupational Behavior,* 2, 181–200.
36. LaRocco, J. M., House, J. S., and French, J. R. P., Jr. (1980). Social support, occupational stress and health. *Journal of Health and Social Behavior,* 21, 202–218.
37. Winnubst, J. A. M., Marcelissen, F. H. G., and Kleber, R. J. (1982). Effects of social support in the stressor-strain relationship: A Dutch sample. *Social Science and Medicine,* 16, 475–482.
38. Eaker, E. D., and Feinleib, M. (1983). Psychosocial factors and the ten-year incidence of cerebrovascular accident in the Framingham Heart Study. *Psychosomatic Medicine,* 45, (abstract).
39. Haynes S. G., and Feinleib, M. (1980). Women, work and coronary heart disease: Prospective findings from the Framingham Heart Study. *American Journal of Public Health,* 70, 133–141.
40. Medalie, J. H., Snyder, M., Groen, J. J., Neufeld, H. N., Goldbourt, U., and Riss, E. (1973a). Angina pectoris among 10,000 men. *American Journal of Medicine,* 55, 583–589.
41. Medalie, J. H., Kahn, H. A., Neufeld, H. A., Riss, E., and Goldbourt, U. (1973b). Five-year myocardial infarction incidence—II. Association of single variables to age and birthplace. *Journal of Chronic Disease,* 26, 329–349.
42. Sutton, R., and Kahn, R. L. (1986). Prediction, understanding, and control as antidotes to organizational stress. In J. Lorsch, *Handbook of Organizational Behavior.* Boston: Harvard University Press.
43. Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly,* 24, 285–307.

44. Karasek, R. A., Theorell, T., Schwartz, J., and Alfredsson, L. (1982b). Job, psychological factors and coronary heart disease. *Advanced Cardiology,* 29, 62–67.
45. Kobasa, S. C., Maddi, S. R., and Zola, M. A. (1983). Type A and hardiness. *Journal of Behavioral Medicine,* 6, 41–51.
46. Pearlin, L. I., Lieberman, M. A., Managhan, E. G., and Mullan, J. T. (1981). The stress process. *Journal of Health and Social Behavior,* 22, 337–356.
47. Levinson, D. J. (1959). Role, personality and social structure in the organizational setting. *Journal of Abnormal and Social Psychology,* 58, 170–180.
48. Katz, D., and Kahn, R. L. (1978). *The Social Psychology of Organizations.* 2nd edition. New York: John Wiley & Sons, Inc.
49. Survey Research Center (1977). *Effectiveness on Work Roles: Employee Responses to Work Environments.* Ann Arbor: Institute for Social Research (2 vols.).
50. U.S. Department of Labor (1977). *Dictionary of Occupational Titles.* 4th edition. Washington, DC: U.S. Government Printing Office.
51. Karasek R. A., Schwartz, J., and Theorell, T. (1982). *Job Characteristics, Occupation, and Coronary Heart Disease.* New York, NY: Department of Industrial Engineering and Operations Research, Columbia University.
52. U. S. Department of Labor (1939). *Dictionary of Occupational Titles.* 1st edition. Washington, DC: U.S. Government Printing Office.
53. U. S. Department of Labor (1949a). *Dictionary of Occupational Titles.* 2nd edition. Vol. 1: Definitions of Titles. Washington, DC: U.S. Government Printing Office.
54. U.S. Department of Labor (1949b). *Dictionary of Occupational Titles.* 2nd edition. Vol. II: Occupational Classification and Industry Index. Washington, DC: U.S. Government Printing Office.
55. U. S. Department of Labor (1965a). *Dictionary of Occupational Titles.* 3rd edition. Vol. I. Washington, DC: U.S. Government Printing Office.
56. U. S. Department of Labor (1965b). *Dictionary of Occupational Titles.* 3rd edition. Vol. II. Washington, DC: U.S. Government Printing Office.
57. Miller, A. R., Treiman, D. J., Cain, P. S., and Roos, P. A. (eds.) (1980). *Work, Jobs, and Occupations: A Critical Review of the Dictionary of Occupational Titles.* (Final report to the U.S. Department of Labor by the Committee on Occupational Classification and Analysis, National Research Council.) Washington, DC: National Academy Press.
58. Cain, P. S., and Treiman, D. J. (1981). The dictionary of occupational titles as a source of occupational data. *American Sociological Review,* 46, 253–278.
59. Berg, I. (1970). *Education and Jobs: The Great Training Robbery.* New York: Praeger.
60. Kohn, M. L. (1969). *Class and Conformity: A Study in Values.* Homewood, IL: Dorsey Press.
61. Kohn, M. L., and Schooler, C. (1983). *Work and Personality: An Inquiry into the Impact of Social Stratification.* Norwood, NJ: Ablex Publishing Corp.
62. Miller, K. A., and Kohn, M. L. (1979). Women and work: The psychological effects of occupational conditions. *American Journal of Sociology,* 85, 66–94.
63. Mortimer, J. T., and Lorence, J. (1979). Work experience and occupational value socialization: A longitudinal study. *American Journal of Sociology,* 84, 1361–1385.
64. Scoville, J. G. (1969). *The Job Content of the U.S. Economy 1940–1970.* New York: McGraw-Hill.
65. Spenner, K. I. (1980). Occupational characteristics and classification systems: New uses of the Dictionary of Occupational Titles in social research. *Sociological Methods and Research,* 9, 239–264.
66. Snyder, D., Hayward, M. D., and Hudis, P. M. (1978). The location of change in the sexual structure of occupations, 1950–1970: Insights from labor market segmentation theory. *American Journal of Sociology,* 84, 706–717.
67. Coburn, D. (1975). Job-worker incongruence: Consequences for health. *Journal of Health and Social Behavior,* 16, 198–212.
68. Schilling, R. S. F., Letal, A. D., Hui, S. L., Beck, G. J., Schoenberg, J. B., and Bouhuys, A. (1977). Lung function, respiratory disease and smoking in families. *American Journal of Epidemiology,* 106, 274–283.
69. Cain, P. S., and Green, B. F., Jr. (1980). Rating DOT worker functions and worker traits. In A. R. Miller, D. J. Treiman, P.S. Cain, and P. A. Roos (Eds.), *Work, Jobs, and Occupations: A Critical Review of the Dictionary of Occupational Titles.* Washington, D.C.: National Academy Press.
70. Cain, P. S., and Green, B. F. (1983). Reliabilities of selected ratings available from the Dictionary of Occupational Titles. *Journal of Applied Psychology,* 68(1), 155–165.
71. House, J. S. (1980). Occupational stress and the physical and mental health of factory workers. Report on NIMH Grant No. 1RO2MH28902. Research Report Series: Institute for Social Research, University of Michigan, Ann Arbor.
72. Hackman, J. R., and Lawler, E. E., III. (1971). Employee reactions to job characteristics. *Journal of Applied Psychology,* 55(June), 259–286.
73. Blau, P. (1960). Structural effects. *American Sociological Review,* 25, 178–193.
74. Davis, J. A., Spaeth, J. L., and Huson, C. (1961). A technique for analyzing the effect of group composition. *American Sociological Review,* 26, 215–225.
75. Tannenbaum, A. S., and Bachman, J. G. (1964). Structural versus individual effects. *American Journal of Sociology,* 6, 585–595.
76. Karasek, R. A., et al. (1985). Job content instrument: Questionnaire and user's guide. University of Southern California, Los Angeles, CA.
77. House, J. S., Strecher, V., Metzner, H. L., and Robbins, C. (1986). Occupational stress and health among men and women in the Tecumseh Community Health Study. *Journal of Health and Social Behavior,* 27, 62–77. Engineering, University of Southern California.
78. Review Panel on Coronary-Prone Behavior and Coronary Heart Disease. (1981). Coronary-prone behavior and coronary heart disease. *Circulation,* 63, 1199–1215.
79. Chesney, M. A., Eagleston, J. R., and Rosenman, R. H. (1981). Type A behavior: Assessment and intervention. In C. H. Prokop and L. A. Bradley (eds.), *Medical Psychology: Contributions to Behavioral Medicine* (pp. 19–36). New York: Academic Press.
80. Bortner, R. W. (1969). A short rating scale as a potential measure of pattern A behavior. *Journal of Chronic Diseases,* 22, 87–91.
81. Jenkins, C. D., Rosenman, R. H., and Zyzanski, S. J. (1974). Prediction of clinical coronary heart disease by a test for the coronary-prone behavior pattern. *New England Journal of Medicine,* 290, 1271–1275.
82. Haynes, S. G., Levine, S., Scotch, N., Feinleib, M., and Kannell, W. B. (1978a and b). The relationship of psychosocial factors to coronary heart disease in the Framingham Study. I. Methods and risk factors. II. Prevalence of coronary heart disease. *American Journal of Epidemiology,* 107(5), 362–402.
83. Gove, W. R., and Hughes, M. (1979). Possible causes of the apparent sex differences in physical health: An empirical investigation. *American Sociological Review,* 44(Feb.), 126–146.
84. Gurin, G., Veroff, J., and Feld, S. C. (1960). *Americans View Their Mental Health.* New York: Basic Books.
85. Schooler, C., Kohn, M. L., Miller, K. A., and Miller, J. (1983). Housework as a Work. In M. L. Kohn and C. Schooler (eds.), *Work and Personality* (pp. 242–260). Norwood, NJ: Ablex Publishing Corp.
86. Schooler, C., Miller, J., Miller, K. A., and Richtand, C. N. (1984). Work for the household: Its nature and consequences for husbands and wives. *American Journal of Sociology,* 90(1), 97–124.

87. Ferree, M. M. (1976). Working class jobs: Housework and paid work as sources of satisfaction. *Social Problems,* 23(April), 431–441.
88. Meissner, M., Humphreys, E. W., Meis, S. M., and Scheu, W. J. (1975). No exit for wives: Sexual division of labor and the cumulation of household demands. *Review of Canadian Sociology and Anthropology,* 12(4).
89. Hartman, H. I. The family as the locus of gender, class, and political struggle: The example of housework. *Signs,* 6(3), 366–395.
90. Berk, S. F. (1977). Going backstage: Gaining access to observe household work. *Sociology of Work and Occupations,* 4(1), 27–48.
91. Berk, R. A., and Berk, S. F. (1979). *Labor and Leisure at Home: Content and Organization of the Household Day.* Beverly Hills, CA: Sage.
92. Quinn, R. P., and Mangione, T. W. (1973). The 1969–1970 survey of working conditions: Chronicles of an unfinished enterprise. Ann Arbor, MI: Survey Research Center, University of Michigan.
93. Quinn, R. P., and Shepard, L. J. (1974). The 1972–73 quality of employment survey: Descriptive statistics with comparison data from the 1969–70 survey of working conditions. Ann Arbor, MI: Survey Research Center, University of Michigan.
94. Quinn, R. P., and Staines, G. L. (1979). The 1977 quality of employment survey: Descriptive statistics, with comparison data from the 1969–70 and the 1972–73 surveys. Ann Arbor, MI: Survey Research Center, University of Michigan.
95. Haskell, W. L., and Fox, S. M. (1983). Physical activity in the prevention and therapy of cardiovascular disease (pp. 455–468).
96. Blackburn, H. (1983). Physical activity and coronary heart disease: A brief update and population view. Parts I and II. *Journal of Cardiac Rehabilitation,* 3, 101–111 and 171–174.
97. Price, R. H. (1985). Work and community. *American Journal of Community Psychology,* 13(1), 1–12.
98. Ferman, L., and Gordus, J. (Eds.) (1979). *Mental health and the economy.* Kalamazoo, MI: Upjohn Institute.
99. Aneshensel, C. S., Frerichs, R. R., and Clark, U. A. (1981). Family roles and sex differences in depression. *Journal of Health and Social Behavior,* 22, 379–393.
100. Nilson, L. B. (1978). The social standing of a housewife. *Journal of Marriage and the Family,* 40(Aug.), 541–548.
101. Spendlove, D. C., Gavelek, J. R., and MacMurray, V. (1981). Learned helplessness and the depressed housewife. *Social Work,* Nov., 474–479.
102. Glazer-Malbin (1976). Housework. *Signs,* 1(Summer), 905–922.
103. Bergmann, B. R. (1981). The economic risks of being a housewife. *AEA Papers and Proceedings.* Vol. 71(2), 81–86.
104. Gove, W. R. (1972). The relationship between sex roles, marital status, and mental illness. *Social Forces,* 51, 34–44.
105. Olson, J. T. (1979). Role conflict between housework and child care. *Sociology of Work and Occupations,* 6(Nov.), 430–458.
106. Kohn, M. L., and Schooler, C. (1973). Occupational experience and psychological functioning: An assessment of reciprocal effects. *American Sociological Review,* 38(Feb.), 97–118.
107. Kohn, M. L., and Schooler, C. (1982). Job conditions and personality: A longitudinal assessment of their reciprocal effects. *American Journal of Sociology,* 87(May), 1257–1286.
108. Aneshensel, C. S. (1986). Marital and employment role-strain, social support, and depression among adult women. In S. E. Hobfoll, (ed.) *Stress, Social Support, and Women* (pp. 99–114). New York: Hemisphere Publishing Corp.
109. Kahn, R. L., and Antonucci, T. (1980). Convoys over the life course: Attachments, roles and social support. In P. B. Baltes and O. Brim (eds.), *Life Span Development and Behavior.* Vol. 3. New York: Academic Press.
110. Gove, W. R. (1973). Sex, marital status and mortality. *American Journal of Sociology.* 79, 45–67.
111. Caplan, R. D., Cobb, S., French, J. R. P., Jr., Harrison, R. V., and Pinneau, S. R., Jr. (1975). Job demands and worker health: Main effects and occupational differences. (USGPO No. HE7111-J57, USPO Stock 1733-0083.) Washington, DC: U.S. Government Printing Office. Reprinted (1980) by the University of Michigan Institute for Social Research, Ann Arbor, Michigan.
112. House, J. S. (1985). Chronic life situations and life change events: Content discussion. In Adrian M. Ostfeld and Elain D. Eaker (eds.), *Measuring Psychosocial Variables in Epidemiologic Studies of Cardiovascular Disease: Proceedings of a Workshop.* (pps. 129–135.) U.S. Dept. of Health and Human Services, NIH Publication No. 85-2270.
113. House, J. S., and Kahn, R. L. (1985). Measures and concepts of social support. In. S. Cohen and S. L. Syme (eds.), *Social Support and Health.* New York: Academic Press.

Appendix: Karasek et al. (1985) Job Content Survey item pool for selecting items for National Health and Nutrition Examination Surveys and National Health Interview Surveys

The following pages contain relevant items from the Karasek et al. (1985) Job Content Survey. The items belong to scales as follows. (See also table 4.) Circled items are the core Quality of Employment Surveys items shown in table 4.

Major concept	Subconcept	Item
1. Decision latitude	a. Skill discretion	Items 3–5, 7, 9, 11
	b. Decision authority (job level)	Items 6, 8, 10
	c. Decision authority (organizational level)	Items 12–18
2. Job demands	Psychological workload	Items 19, 20, 22, 23, 26, 27–29, 32
3. Job security	—	Items 33–36
4. Social support	a. Supervisor support	Items 48–52
	b. Coworker support	Items 54–58
5. Physical-chemical-biological hazards	—	Items 39–47
6. Physical demand	a. Activity/exertion	Items 21, 24, 25
	b. Awkward positions	Items 30, 31

FOR THE QUESTIONS BELOW, PLEASE CHECK THE BOX WITH THE ANSWER THAT COMES CLOSEST.

③ My job requires that I learn new things.

Q3--Skill Discretion

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

④ My job involves a lot of repetitive work.

Q4--Skill Discretion

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

⑤ My job requires me to be creative.

Q5--Skill Discretion

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

Note: This instrument is copyrighted by the Job-Heart Project. Permission for research uses can be obtained from Professor Robert Karasek, Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089 (Phone 213–743–2705). (Copyright; used by permission.)

Job Content Survey Page 2

⑥ My job allows me to make a lot of decisions on my own.
Q6--Decision Authority
☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

⑦ My job requires a high level of skill.
Q7--Skill Discretion
☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

⑧ On my job, I have very little freedom to decide how I do my work.
Q8--Decision Authority
☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

⑨ I get to do a variety of different things on my job.
Q9--Skill Discretion
☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

⑩ I have a lot of say about what happens on my job.
Q10--Decision Authority
☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

⑪ I have an opportunity to develop my own special abilities.
Q11--Skill Discretion
☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

12. How many people are in your work group or unit?

☐ I Work Alone (1) ☐ 2-5 People (3) ☐ 6-10 People (8) ☐ 10-20 People (15) ☐ 20 or More People (30)

13A. I have significant influence over decisions in my work group or unit.

☐ I Work Alone (8) ☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

13B. My work group or unit makes decisions democratically.

☐ I Work Alone (8) ☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

14. I have at least some chance that my ideas will be considered about company policy (e.g., hiring, firing, wage levels, plant closings, new machinery purchases, etc.).

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

15. I supervise other people as part of my job.

☐ No (0) ☐ Yes 1-4 People (3) ☐ Yes 5-10 People (8) ☐ Yes 11-20 People (15) ☐ Yes More Than 20 People (30)

16. I am a member of a union or employee association.

☐ Yes (2) ☐ No (1)

17. My union or employee association is influential in affecting company policy.

☐ I Am Not a Member (8) ☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

18. I have influence over the policies of the union or employee association.

☐ I Am Not a Member (8) ☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

⑲ My job requires working very fast.

Q19--Psychological Workload

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

⑳ My job requires working very hard.

Q20--Psychological Workload

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

㉑ My job requires lots of physical effort.

Q21--Physical Exertion

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

㉒ I am not asked to do an excessive amount of work.

Q22--Psychological Workload

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

㉓ I have enough time to get the job done.

Q23--Psychological Workload

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

24. I am often required to move or lift very heavy loads on my job.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

25. My work requires rapid and continuous physical activity.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

㉖ I am free from conflicting demands that others make.

Q26--rPsychological Workload

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

27. My job requires long periods of intense concentration on the task.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

28. My tasks are often interrupted before they can be completed, requiring attention at a later time.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

29. My job is very hectic.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

30. I am often required to work for long periods with my body in physically awkward positions.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

31. I am required to work for long periods with my head or arms in physically awkward positions.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

Job Content Survey Page 5

32. Waiting on work from other people or departments often slows me down on my job.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

33 How steady is your work? (Check one.)

Q33--Job Insecurity

☐ Regular and steady. (1)
☐ Seasonal. (4)
☐ Frequent layoffs (4)
☐ Both seasonal and frequent layoffs. (4)
☐ Other. (9)

34 My job security is good.

Q34--Job Insecurity

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

35. During the past year, how often were you in a situation where you faced job loss or layoff?

☐ Never ☐ Faced the possibility once ☐ Faced the possibility more than once ☐ Constantly ☐ Actually layed off

36 Sometimes people permanently lose jobs they want to keep. How likely is it that during the next couple of years you will lose your present job with your employer?

Q36--Job Insecurity

☐ Not at all likely (1) ☐ Not too likely (2) ☐ Somewhat likely (3) ☐ Very likely (4)

37. My prospects for career development and promotions are good.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

38. In five years, my skills will still be valuable.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

㊴ Do you have a problem with exposure to dangerous chemicals on your job?

Q39--Toxic Exposures

☐ Not exposed (0)

☐ I am exposed but (1) but it is a slight problem

☐ I am exposed and it (2) is a sizeable or great problem

㊵ Do you have a problem with exposure to air pollution from dusts, smoke, gas, fumes, fibers, or other things on your job?

Q40--Toxic Exposures

☐ Not exposed (0)

☐ I am exposed but but it is a slight (1) problem

☐ I am exposed and it is a sizeable or great problem (2)

㊶ Do you have a problem with exposure to things placed or stored dangerously on your job?

Q41--Hazardous Conditions

☐ Not exposed (0)

☐ I am exposed but but it is a slight (1) problem

☐ I am exposed and it is a sizeable or great problem (2)

㊷ Do you have a problem with exposure to dirty or badly maintained areas at your workplace?

Q42--Hazardous Conditions

☐ Not exposed (0)

☐ I am exposed but but it is a slight (1) problem

☐ I am exposed and it is a sizeable or great problem (2)

㊸ Do you have a problem with risk of catching diseases on your job?

Q43--Toxic Exposures

☐ Not exposed (0)

☐ I am exposed but but it is a slight (1) problem

☐ I am exposed and it is a sizeable or great problem (2)

㊹ Do you have a problem with dangerous tools, machinery, or equipment?

Q44--Hazardous Conditions

☐ Not exposed (0)

☐ I am exposed but but it is a slight (1) problem

☐ I am exposed and it is a sizeable or great problem (2)

㊺ Do you have a problem with exposure to fire, burns, or shocks?

Q45--Hazardous Conditions

☐ Not exposed (0)

☐ I am exposed but but it is a slight (1) problem

☐ I am exposed and it is a sizeable or great problem (2)

Job Content Survey Page 7

46. While you are working, how loud would you have to talk to be heard by someone standing next to you?

☐ Whisper ☐ Normal Voice ☐ Loud Voice ☐ Shout

㊼ Do you have a problem with exposure to dangerous work methods on your job?

Q47--Hazardous Conditions

☐ Not exposed (0) ☐ I am exposed but but it is a slight (1) problem ☐ I am exposed and it is a sizeable or great problem (2)

㊽ My supervisor is concerned about the welfare of those under him.

Q48--Supervisor Support

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4) ☐ I have no supervisor (8)

㊾ My supervisor pays attention to what I am saying.

Q49--Supervisor Support

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4) ☐ I have no supervisor (8)

50. I am exposed to hostility or conflict from my supervisor.

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4) ☐ I have no supervisor (8)

(51) My supervisor is helpful in getting the job done.

Q51--Supervisor Support

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4) ☐ I have no supervisor (8)

(52) My supervisor is successful in getting people to work together.

Q52--Supervisor Support

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4) ☐ I have no supervisor (8)

(53) People I work with are competent in doing their jobs.

Q53--Coworker Support

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

(54) People I work with take a personal interest in me.

Q54--Coworker Support

☐ Strongly Disagree (1) ☐ Disagree (2) ☐ Agree (3) ☐ Strongly Agree (4)

Chapter **22**

Effects of Physical Activity and Fitness on Health

Arthur S. Leon, M.D.
University of Minnesota

In this chapter I will review the currently available evidence relating *physical activity* status and *physical fitness* to health and longevity. Definitions of associated terms are appropriate (1, 2).

Definition of terms

Physical activity

Physical activity may be defined as *dynamic* or *static* skeletal muscle exertion that increases the body's energy expenditure and results in cardiorespiratory adjustments. Dynamic physical activities involve body movements through rhythmic contraction and relaxation of large skeletal muscle groups. Examples are walking or running. If a dynamic movement is accomplished against a resistance, this is referred to as an isotonic activity, e.g., repeatedly lifting a bar bell. *Static* or *isometric* activity consists of increased muscular tension against a fixed resistance with no change in fiber length. Examples are attempting to lift a heavy object, pushing against an immovable object, or attempting to open a stuck window. Many activities have both static and dynamic components. Physical activities also are commonly identified in studies as *occupational* (work related) or *recreational* (leisure time).

Dynamic activities if mild to moderate in intensity and lasting more than 2 minutes require a steady supply of oxygen by the body to provide continuous energy for the contracting muscles. For this reason this type of activity is often referred to as *aerobic* (with air). There is an associated increase in respiratory rate, heart rate, stroke volume, systolic blood pressure, and coronary blood flow. This type of physical activity is generally considered to be more beneficial to one's health than static activities.

Energy is generated by anaerobic metabolism (without air) during the first few minutes of prolonged dynamic activities; during short duration, high-intensity, dynamic activities (such as jumping, sprinting, or throwing); and during static activities. During anaerobic activities energy is provided for muscular function without the use of oxygen. This results in an increase in muscle lactic acid formation from breakdown of stored carbohydrate (glycogen). Further, when the intensity of a dynamic exercise exceeds 70 percent of maximal aerobic capacity, anaerobic metabolism is progressively stimulated contributing to fatigue and discomfort.

Static activities result in a reflex increase in resistance in small arteries through increased sympathetic nervous system activity, which may cause an excessive elevation in blood pressure out of proportion to the increase in cardiac output (heart rate × stroke volume) compared with exercise. Myocardial oxygen demands and coronary blood flow requirements are thereby accentuated.

Exercise

Exercise is a category of physical activity that is "planned, structured, and repetitive" (1). Common intended purposes or determinants of exercise performance include desire to participate in competitive sports, social reasons, diversions from daily tasks, relief of muscle tension, "positive feelings," health promotion and disease prevention, weight management, and improvement or maintenance of physical fitness. Exercise, like physical activity, may be subcategorized into dynamic or static and aerobic or anaerobic. Currently the term *aerobic exercise* is commonly used to refer to dynamic exercise of sufficient intensity, duration, and frequency to improve cardiorespiratory endurance or *aerobic capacity.*

Exercise training

Exercise training is the regular chronic performance of exercise for the purpose of improving physical fitness. *Conditioning* is a synonymous term preferred by many physiologists. However, it should be noted that nonstructured physical activity also is positively related to improved physical fitness and health.

Physical fitness

The President's Council on Physical Fitness and Sports defines physical fitness as the ability to carry out daily tasks with vigor and alertness without undue fatigue and with ample reserve energy to enjoy active leisure pursuits. Physical fitness allows one to respond to physical and emotional stress without an excessive increase in heart rate and blood pressure, and therefore, it reduces the work of the heart. It is important for successful adaptation to a wide range of circumstances in modern living patterns.

The basic components of physical fitness may be classified as *health-related* or *skill-related* fitness and are as follows (1).

Health-related components

Cardiorespiratory endurance or aerobic capacity. Cardiorespiratory endurance is the ability to perform prolonged, moderate-intensity, rhythmic contractions and relaxation of large muscle groups that result in transporting the body over distance or against gravity. This is accompanied by stimulation of the circulatory and respiratory systems in order to maintain sufficient blood and oxygen delivery to the exercising muscles. Cardiorespiratory endurance is probably the most important component of fitness in terms of health and quality of life. It is improved by regular performance of dynamic physical activities if of sufficient intensity, duration, and frequency.

Muscular endurance. This reflects the ability of skeletal muscles to do prolonged static or dynamic work.

Muscular strength. Muscular strength is the contracting power of skeletal muscles and is improved by regular static or isotonic work or exercise training.

Flexibility. This relates to range of motion in a joint or sequence of joints.

Body composition. Body composition refers to the amount of lean and fat tissue in the body mass. Dynamic activities when of sufficient duration and frequency reduce body fat.

Other metabolic effects. In recent years there has been growing interest in potentially beneficial metabolic effects of dynamic physical activities on blood lipid-lipoproteins and glucose-insulin dynamics. Although these changes appear to be direct or independent effects of conditioning, an additive contribution is derived from associated loss of body weight and fat.

Skill-related components. These include agility, speed, balance, coordination, muscular power, and reaction time.

The health-related fitness components generally are believed to play a role in reduction in premature disability and mortality from chronic diseases (1). However, it cannot be excluded that health improvement with regular physical activity may be the result of additional biologic changes unrelated to physical fitness measurements (3).

Assessment of physical activity and physical fitness

More than 30 different approaches have been used to assess physical activity. These have recently been extensively reviewed (4). These approaches can be grouped into seven major categories: calorimetry, job titles or classification, survey questionnaires or interview procedures, observer monitoring, mechanical and electronic movement detectors, dietary measures, and physiological markers. Criteria to be considered for evaluation of appropriateness for use in studies such as NHANES include the following (4):

1. *Validity.* Does the instrument measure what it is intended to measure?
2. *Reliability.* Does the instrument consistently give the same results under the same circumstances?
3. *Practicality.* Is it practical in terms of personnel and subject time, cost, personnel required, and the convenience of the participants to use the instrument?
4. *Reactivity.* Does the instrument alter the behavior it seeks to measure or is it nonobtrusive?

Direct calorimetry

Direct calorimetry determines energy expenditure by measurement of associated heat production. This is a highly accurate method to measure energy costs of given activities; however, it requires the use of special chambers or an insulated "space suit," making it impractical for monitoring physical activity in large populations.

Indirect calorimetry

This approach measures the oxygen consumption during activities. Oxygen uptake increases with the intensity of dynamic physical exertion. It correlates well with heat production and energy expenditure. However, it requires that a participant wear a face mask or a mouth piece and nose clip and carry a collection container to collect expired air. It is not practical for use with large populations because of cost for equipment and personnel; however, it could be employed in systematic samples of the study population to validate other survey methods.

Job classification

In the past, job titles were often used to index jobs according to supposed levels of physical activity. Although this method is easy to do, low cost, and nonreactive, it is not a valid and reliable measurement of physical activity. Problems include marked variability and fluctuations in energy expenditure within job classes. Furthermore because of extensive mechanization few jobs any longer require much physical exertion.

Survey procedures

Survey procedures consist of questionnaires either self-administered or by personal or phone interview. Respondents are asked to recall physical activities at work and/or during leisure for periods of 1–7 days or up to a year. Another approach is for participants to maintain a physical activity diary. This has been used both in adults and children (5). Diary surveys appear to be the most accurate of these survey approaches for estimating daily energy expenditure. However, these are more expensive than self-administered questionnaires, require more participant time and a reasonable education level, and may thereby be unacceptable for use in large general populations. They also may result in alteration of usual physical activity patterns to simplify the recording procedure.

The Minnesota questionnaire (6) and a similar questionnaire by Montoye (7), used in the Tecumseh Heart Study, contain an extensive list of specific activities. Participants quantitate estimated frequency and duration of each activity over the previous year. Through the use of tables giving known values of energy costs from physiologic studies for specific activities, total energy expenditure for these activities can be calculated as well as the energy expenditure for activities subclassified as light, moderate, and high intensity. These questionnaires may be self-administered or administered by an interviewer and take 45–60 minutes for completion. The Minnesota questionnaire was used in the recently completed NHLBI Multiple Risk Factor Intervention Trial, and a similar questionnaire was used in the Evans County study. The Minnesota questionnaire has been partially validated against physical fitness measurements and caloric intake on dietary records as well as against HDL cholesterol, body mass index, and heart rate (8, 9). Short-term repeatability is high.

The usefulness of these quantitative activity history questionnaires in obtaining detailed information on physical activity habits has been demonstrated in population surveys; however, the time required for completion and the need for trained interviewers or reviewers may limit their general applicability in NHANES. They could be used on a systematic subsample, however.

A modified 7-day version of the Minnesota questionnaire developed at Stanford University, which includes work-related activities, has been partially validated and has been shown to have a short-term repeatability of 0.67 (10). This questionnaire may be useful in regions of the country in which there are no great seasonal fluctuations in physical activities; otherwise the questionnaires need to be repeated in different seasons of the year.

Paffenbarger et al. (11) used a simple recall survey sent through the mail to college alumni, which attempted to elicit frequency of major contributors to the weekly physical activity pattern. Similarly in the Netherlands, Magnus et al. (12) determined hours per week spent walking, cycling, or gardening per year. Unfortunately, there is little information available on the reliability or validity of these simplified questionnaires. In the NHLBI Lipid Research Clinic's Prevalence Study and Coronary Primary Prevention Trial, an even simpler two-question survey was used prior to exercise testing. This elicited whether the participants performed heavy physical labor or vigorous exercise, and if so, whether they performed it more than twice a week. Those who reported the most heavy exertion had longer durations on the treadmill exercise test and higher HDL cholesterol concentrations than those who reported no heavy exertion (13).

Several large scale prospective epidemiologic trials have included assessments of both work and leisure time physical activity. These include the Health Insurance Plan of New York study, the Framingham Heart Study, and the North Karelia Study in Finland. These instruments lack validation and reliability studies, although a questionnaire similar to that used in the North Karelia Study was significantly related to estimated maximal oxygen uptake in another population (14).

Heart rate, mechanical, and electronic movement detectors

Heart rate monitoring and electronic motion sensors are available as objective means of assessing physical activity. Heart rate monitoring theoretically seems an attractive approach to monitoring physical activity of individuals in a population. The technological capabilities are available for continuous 24-hour recordings of heart rate. Heart rate is a physiological parameter known to be directly related to intensity of physical activity and in the laboratory is linearly related to oxygen consumption. However, it is impractical to use heart rate monitoring to assess physical activity status in a large scale study because individual laboratory assessments are needed to determine the relation of heart rate to workload and oxygen consumption; because of the cost and time required for monitoring; and because of the large number of confounding factors affecting heart rate, including response to emotional stress (4).

A number of devices are available that directly measure body movement. These include an instrument affixed to a subject's thigh used to compare the amount of time seated or lying down versus standing or moving about; and pedometers and shoe step counters that fit into the shoe both for recording walking behavior (4). None of these instruments have been used before in large-scale population studies to determine activity levels.

Electronic motion sensors have recently received much attention and are now commercially available. They are easy to use, nonintrusive ways to monitor an individual's daily movements. These include the large-scale integrated (LSI) motor activity monitor and accelerometers. The LSI is an instrument the size of a pocket watch that can be placed on various body sites to measure body movements. The LSI appears to accurately measure movement over time in diverse populations (15). Accelerometers, which measure both frequency and intensity of movement, are also available (16). Their usefulness for population studies remains to be confirmed. Currently the cost of these electronic monitors of movement is still prohibitive for large-scale population studies, but may be expected to decrease with technological advances.

Behavioral observations

This approach consists of an observer watching an individual and rating the observed activity at specific time intervals (4, 17). Such an approach is impractical on a population basis. It also is unacceptable to many people; and, in addition, the presence of an observer may influence activity levels.

Dietary measures

The caloric intake from food may be used to assess energy expenditure in subjects in energy balance who have stable body weight. Caloric intake can be assessed from dietary recall or records, as recently done in NHANES. In our laboratory, significant positive correlations have been found between caloric intake from 3-day dietary records expressed in kcal per day or kcal per kg of body weight per day and leisure time physical activity as determined by the Minnesota questionnaire (6, 18). In addition, a significant correlation was demonstrated between calorie intake and maximal oxygen consumption and a significant negative correlation with percent body fat. However, the considerable intra- and interindividual variability in a 24-hour recall may reduce its accuracy for estimating an individual's physical activity status. Nevertheless, because this data is routinely collected in NHANES, it should be compared with other measures of physical activity.

Physiological markers

Assessment of physical fitness was thoroughly reviewed by Wilmore in chapter 4. Essentially, the cardiorespiratory endurance as measured by maximal oxygen uptake ($\dot{V}O_2$ max) is generally thought to be the most important component of physical fitness for health promotion and disease prevention. It can be measured directly in the laboratory using steps, bicycle ergometer, or treadmill. It also can be accurately estimted by duration to exhaustion on a standard exercise test protocol during a submaximal exercise load using a treadmill or bicycle ergometer, or from the heart rate achieved by field tests to determine the maximal distance that can be achieved walking or running in a fixed time (e.g. 12 minutes), or from the time to cover a fixed distance (1, 1.5, or 2.0 miles). Relative cardiorespiratory fitness in epidemiologic surveys can also be obtained by measuring the workload achieved before reaching a heart rate of 140, 150, or 170 bpm (WL 140, 150, or 170) on a bicycle ergometer or treadmill test.

A significant negative relationship has also been demonstrated between aerobic power and heart rate at rest or during submaximal exercise loads. Further, a positive association exists between $\dot{V}O_2$ max and vital capacity or forced 1-second expiratory volume (FEV_1 sec).

Although regular dynamic physical activity can increase cardiorespiratory endurance, direct or indirect indicators of maximal oxygen consumption only have a modest relationship to physical activity status. This is because genetic factors are important contributors to an individual's $\dot{V}O_2$ max. Resting heart rate and heart rate during submaximal exercise and FEV_1 also have been found to have a weak but significant association with habitual physical activity in some studies.

Evidence for health benefits of regular physical activity

Increased regular exercise to improve physical fitness has been officially recommended as part of any strategy for public health promotion and disease prevention by numerous governmental and professional organizations around the world. They include the World Health Organization, the U.S. Public Health Service, the American Public Health Association, the American Heart Association, the Canadian Government, the Royal College of Physicians, and the British Cardiac Society. The basis for these recommendations is the growing body of evidence of the beneficial effects of physical activity on physical and mental health. These benefits have recently been reviewed and are summarized in table 1 (19–22). The evidence supporting these relationships comes from a variety of sources: anecdotal, phylogenic observational-epidemiologic, observational-epidemiologic, and experimental studies in animals and humans.

Anecdotal

Regular physical activity has been regarded as an adjunct to good health down through history by sages and respected observers and in folklore as evidenced by the writings of Cicero, Socrates, and Hippocrates in Ancient Greece (about 100 BC), and Avicenna (908–1037) and Moses Maimonides (1135–1204), both practicing medicine in ancient Moslem countries (21). In the 18th and 19th centuries, several prominent British physician-scientists, including Thomas Addison, advocated regular exercise. In 1799, Easton obtained the first supporting observational data. After studying 1,712 people in England who were over 100 years of age, he concluded that one of the things they had in common was the use of "much exercise." In 1864, a British physiologist-physician, Edward Smith, observed a higher mortality rate in people in sedentary occupations compared with those who were physically active.

Table 1. Benefits of regular dynamic exercise

- Skeletal muscle changes
  - Increased blood supply (capillaries)
  - Increased oxygen uptake and utilization
  - Increased metabolic capacity
  - Increased myoglobin content
  - Increased energy (glycogen) stores
  - Decreased lactic acid production
- Cardiovascular-respiratory changes
  - Increased cardiorespiratory capacity and endurance
  - Reduced heart rate
  - Reduced blood pressure
  - Reduced cardiac work and coronary blood flow requirements
  - Increased heart muscle (myocardial) blood supply
  - Increased output per beat (stroke volume)
  - Decreased cardiac irritability
  - Increased blood volume and hemoglobin
  - Reduced blood coagulability
  - Reduced work of breathing
- Metabolic adaptations
  - Reduced body fat stores
  - Appropriate appetite adjustment
  - Increased tissue sensitivity to insulin and improved glucose tolerance
  - Improvement in blood lipid profile: Reduced triglycerides and increased high-density lipoproteins
  - Improved bone mineralization and cortical thickness
- Psychosocial benefits
  - Feeling of well being
  - Tranquilizing and muscle relaxing effect
  - Reduced mental depression
  - Promotes sound sleep
  - Increased stamina and resistance to fatigue
  - Motivation to improve other health habits
  - Provides a means for fellowship

Early in the history of this country, prominent leaders such as Benjamin Franklin, Thomas Jefferson, and John Quincy Adams expressed strong opinions in favor of habitual exercise and physical fitness.

In contrast, early in this century following the industrial revolution, the medical community and lay public not only questioned the need for exercise but also thought that vigorous exercise might be detrimental to health. This concept was later challenged by a pioneer in American cardiology, Paul Dudley White, who regarded the sedentary way of life in this country "a pity" and harmful to physical and mental health and especially to the "organic soundness of the cardiovascular system." Kraus and Raab, during the same period, developed the concept of "hypo-kinetic diseases," which they defined as "the whole spectrum of inactivity-induced somatic and mental dearrangements." Based on information from many sources, they hypothesized that many disorders of the body were more frequent in the sedentary than in the active person, including obesity, diabetes mellitus, low back pain, emotional difficulties, and coronary heart disease. Supporting evidence of these associations are now available and will be discussed later in this review.

Phylogenic

Teleological evidence can be used to support the claim of a human requirement for regular vigorous physical activity (23). This is based on the fact that, for over 99% of the 2 million years or so of existence, our ancestors were hunter-gatherers. The argument goes that such a life style requires a greater deal of physical activity, thereby physical fitness conferred survival advantages through natural selection. Thus we are believed to have inherited a legacy of bodies mechanically well designed for walking, running, lifting, and carrying, plus metabolic adaptations to sufficiently supply energy for prolonged physical activity. Regular physical activity is then postulated to be required to maintain proper functioning of our bodies. However, civilization promotes mass sedentariness through agriculture, urbanization, mechanization, automation, and motorized transportation. This mass sedentariness is postulated to contribute to mass physiologic and metabolic maladaptations, including obesity, hyperlipidemia, hyperglycemia, hyperinsulinemia, hypercoagulability, thinning of bone, and weakness of muscles (23). These maladaptations in turn are believed to be precursors for the mass occurrence of adult diseases and health problems, including coronary heart disease, cerebrovascular accident, adult-onset diabetes, osteoporosis, chronic low back trouble, and perhaps certain types of malignancies. Supporting evidence for phylogenic relationship of physical activity to the structure and function of the human body includes examination of fossil human leg bones. This reveals great cortical thickness and distinctive protuberance, indicating great muscularity and strength (23). Observation of surviving groups of hunter-gatherers in Africa and Australia give us further clues to the past. Their usual activity pattern consists of 16–24 hours a week of mostly moderate intensity activities to acquire food, with periods of heavier exertion during dancing and other recreational activities.

Experimental

It is widely assumed by the lay public and many health professionals that regular vigorous physical activity will reduce the rate of aging and increase the life span. Experimentally, rats exercised 10 minutes per day on a motor driven drum for their entire life span are reported to live 25 percent longer than nonexercised litter mates (24). However, forced exercise begun late in life appears to hasten the death of the animal. A great deal more work is needed in additional species of animals to be certain of the effects of exercise on life span.

The consequences of extreme physical inactivity (bed or chair rest) or immobilization have been systematically studied in humans. The physiologic changes that result are substantial, reproducible, and reversible, and closely mimic those of human aging. This supports the hypothesis that physical activity can retard the aging process and perhaps prolong the life span. Reversible physiologic changes and metabolic changes with inactivity that mimic aging include a rapid and marked deterioration in $\dot{V}O_2$ max, a reduction in blood volume, postural hypotension, skeletal muscle atrophy with a loss in strength, and disturbances in central nervous system function (25–27). Metabolic dysfunctions include negative nitrogen balance, hyperinsulinemia, reduced glucose tolerance, negative calcium balance associated with bone demineralization and osteoporosis, blood clotting defects, and perhaps even reduced immunologic response (28).

In contrast cross-sectional comparison studies show that physically active older people have $\dot{V}O_2$ max levels similar to sedentary people of much younger ages (29–31). However, the rate of decline of $\dot{V}O_2$ with age appears to be about 4.5% to 10% per decade; however, there is a great deal of variability between people. The rate of decline in strength with age appears to be slower than for $\dot{V}O_2$ max. and may be minimal in active men up to age 60 (31, p. 114; 32). Short-term exercise training studies also have demonstrated the ability to reverse, improve, or stabilize most physiologic and metabolic dysfunctions associated with inactivity and/or advancing age (20, 27, 28, 33–36).

Longevity of athletes, epidemiologic evidence

Another approach in attempt to study the relationship between physical activity and longevity has been to compare the longevity and health of former athletes with nonathletes. Most of these studies have involved college athletes and a few Olympic champions and professional athletes. The results of these studies have previously been reviewed (36–39).

Generally, studies that compare life span of former athletes with the general population or matched insurance holders found that the ex-athletes tended to live slightly longer. However, the increased life span cannot be attributed to exercise because, for most of the former athletes, strenuous activities were undertaken only for a limited period of time relatively early in life. Other variables that may have contributed to the differences in mortality rates include physique, psychological factors involved in selection for sports, socioeconomic status (especially if college graduates are compared with nongraduates), and life habits including exercise later in life, diet habits, and the use of cigarettes and alcohol. In this regard, it is relevant that former major league baseball players, managers, and umpires are reported to live longer than men in the general population, confirming that factors other than vigorous exercise are involved in promoting longevity in sports figures (39). Further proof of this is that occupations requiring unusually hard physical labor such as lumbering are not associated with a long life expectancy perhaps because of confounding by socioeconomic status and other life habit variables (37).

More recent studies have more appropriately compared the longevity of college sports letter winners relative to their classmates. These studies have either shown no differences between ex-athletes and their classmates or that honor graduates fared slightly better than letter winners and that minor sports letter winners lived longer than major letter winners. The difference in life span between these various groups was a modest 2 years or less. There is some evidence that recent physical activity, exercise, and/or sports activity is more likely to reduce premature mortality than remote college sports (40). These findings suggest that current physical activity and associated habits may be more important protective factors than genetic factors leading to selection of sports earlier in life. However, this remains to be confirmed.

Physical activity and longevity in general population

Rose and Cohen (41), as part of one of the few long-term longitudinal aging studies, looked at recent "on-job" and "off-job" activity at the time of death as determined by interviewing survivors. They found significant, but low-order, positive correlations between both categories of activity and age at death. However, when the correlation was adjusted for education only the association with off-job activity persisted and actually showed a stronger relationship with longevity (0.14–0.18). Other variables found to be more strongly related to age of death than physical activity variables included youthful appearance (0.47), mean number of illnesses (−0.46), smoking (−0.29), worry (−0.28), and urban vs. rural residence (−0.24). The 10 best predictors, which included the physical activity variables, accounted for 44% of the variance after initial control for education.

Belloc and Breslow (42) studied the association of health habits with health status and longevity in a probability-based sample of almost 7,000 adults in Alameda County, California. Data was obtained by questionnaires. A relationship was found between the following health practices and a health-vigor index: regular physical activity, no cigarette smoking, no alcohol consumption or moderation, regularity of meals, about average weight for height and gender, and adequate sleep. Those practicing all of these health practices had an average health-vigor index, the same as people 30 years younger who practiced fewer favorable habits than they, and in a followup study, the vigorous were found to live 15 years longer than the careless or immoderate individuals.

Physical activity levels and coronary heart disease

In the absence of randomized clinical trial, epidemiologic studies are important to test whether regular exercise protects against coronary heart disease (CHD), morbidity, and mortality, and to justify the use of exercise as a preventive measure. Relevant epidemiologic studies have recently been critically reviewed and will be summarized here along with supporting experimental data (43, 44).

Over 30 years ago Morris and colleagues (45) reported that physically active London conductors on double-decker buses and postmen who walked or cycled on the job had lower rates of CHD events and sudden death compared with sedentary bus drivers, postal clerks, and government civil servants. Followup of the busmen showed that the conductors throughout middle-age had one-half the CHD death rate of the drivers (46). However, it subsequently was determined that, at the time of recruitment, the bus drivers were heavier and apparently at greater risk for CHD than the conductors (47). Thus selection may have contributed to the differences in CHD rates. Despite this shortcoming, the reports of Morris and colleagues stimulated worldwide interest in the relationship between habitual physical activity and CHD rates.

The studies that followed continued to have certain limitations, including problems similar to the studies in athletes of self-selection versus protection; confounding risk factors that made it difficult to determine the independent effects of exercise; shifts from more active to less active jobs and leisure time pursuits for health reasons, concentrating those with health problems in the least active category; and difficulties in assessing levels of habitual physical activity in populations and in accurately diagnosing CHD (23, 36). More recent epidemiologic studies have attempted to deal with these problems with varying degrees of success.

Occupational activity

Comparisons of incidence or prevalence of CHD events in men performing jobs requiring different levels of physical activity have been made in a variety of industries, including postal workers (45, 48); railroad workers (49–51); farm workers in Iowa (52), North Dakota (53, 54), and Georgia (55); employees of utility companies (55–60); police and firefighters (61); civil servants (62–66); and longshoremen (44, 67, 68). Most physically active workers, as determined primarily by occupational classifications, had one-third to three-fourths fewer total or fatal CHD events than the least active. However, not all such studies found this effect. No association was noted among civil service employees in Los Angeles (62) nor among utility company employees (56–58), perhaps because within these light industries jobs lacked contrast in energy expenditure requirements. In addition, leisure-time activities were not considered. Although a prevalence study by Taylor et al. (50) of Midwest railroad switchmen versus clerks and executives showed an inverse relation between job activity and CHD, it was believed by the investigators that job transfers by sicker workers from more active to inactive jobs may have contributed to the apparent association; moreover, a 5-year followup study that failed to find a relation between supposed job activity was not considered (49). Furthermore, direct observations revealed that job titles used to assess physical activity did not accurately reflect actual energy expenditure on the job.

In Finland, lumberjacks were found to have a higher rate of mortality from CHD and greater frequency of abnormal electrocardiograms than did less active farmers from the same region (69). However, lumberjacks consumed more saturated fat, smoked slightly more, and were of lower socioeconomic status than farmers, which may have "overwhelmed" protective effects of vigorous physical exertion.

Farmers consistently have been shown to have a lower CHD rate than nonfarmers. For example, a recent study by Pomrehn et al. (52) found a 10 percent lower CHD and total mortality rate in Iowa farmers than in nonfarmers. These farmers were twice as likely to be more physically active than nonfarmers, were more fit by exercise testing, and had a higher caloric intake consistent with their greater energy output. However, a significantly lower consumption of tobacco and alcohol in Iowa farmers as compared with nonfarmers confounded the results. Similarly, lower smoking rates among farmers in North Dakota (54) and Evans County, Georgia (55) probably also contributed to their lower CHD rates.

Among the better controlled occupational studies were a retrospective study of Israeli workers in collective settlements called kibbutzim (70) and a 22-year prospective study of San Francisco longshoremen (67, 68). The Israeli study is of special interest because many of the usual confounding variables were eliminated as a result of the subjects' relatively homogenous ethnic origin and uniform environment, including a communal dining facility and similar medical care. In addition, further evaluation of a representative sample of the study population revealed no significant difference in average body weight and blood lipid levels between the two activity groups. In this study, a survey of 15-year incidence rates of fatal and nonfatal coronary events was made of 5,288 men and 5,229 women, of from 40 to 64 years of age, from 58 collective settlements. Based on the percentage of the working day spent seated or performing manual labor, workers were classified as either sedentary or nonsedentary. The relative risk of total CHD for men engaged in sedentary work was 2.5 times that of men engaged in physical work. The corresponding risk ratio in women was 3.1.

Paffenbarger and Hale (67, 68) followed 6,351 longshoremen, age 35–74 years at entry, for 22 years (or to death or age 75). Physical activity status was estimated from analysis of job situations; changes in work assignments were checked annually and adjusted for; and mortality rates were based on official death certificates and person-years of observation. Cargo handlers who loaded and unloaded ships were classified as physically active in contrast to foremen and clerks, considered less vigorous. Self-selection as a factor affecting outcome was excluded because union rules required all workers to serve at least their first 5 years as cargo handlers (the average being 13 years). The CHD death rate per 10,000 man-years of work was about one-half for the group of men. The rate for sudden death was about one-third as high for the more active workers. The differential in CHD deaths remained statistically significant when adjustments were made for other risk factors.

Leisure time exercise

Because mechanization and automation have reduced the job-related physical exertion in most industries to a minimum, recent studies exploring the association between physical activity and CHD rate have focused on the leisure-time activities of the workers. The apparent protective effects of regular leisure-time exercise against CHD in this type of study are more consistent than those found in studies that considered only occupational activities. The risk in the more active men was generally about one-half to two-thirds of the least active.

Chave, Epstein, and Morris (63–66) considered leisure-time physical activity obtained by a simple questionnaire only in studying a cohort of about 18,000 middle-aged male British civil servants. During the next 8½ years of followup, those who reported vigorous exercise, such as sports, jogging, rapid walking, hiking, hill climbing, or heavy work around the house, garden, or garage had about one-half the rate of initial heart attacks and about one-third the rate of CHD mortality as did their colleagues who reported no vigorous exercise. Moreover, the CHD incidence and mortality rates increased significantly with age only in the physically inactive group. An evaluation of a representative sample from this cohort revealed no significant difference in other conventional risk factors between the exercise and non-exercise groups (64). However, the frequency of resting electrocardiographic abnormalities was twice as high in the inactive group. The latter finding is in agreement with Rose (71), who found an inverse association between duration of walk to work and resting electrocardiographic abnormalities in British civil servants.

Paffenbarger et al. (11, 72, 73), as part of a long-term cohort study of precursors of chronic disease, assessed the association between remote and recent leisure-time activity determined by a mail questionnaire to CHD rates in 17,000 male alumni of Harvard University. Most of these men were in sedentary occupations or were retired. The age-adjusted incidence rate of CHD was inversely related to the energy expenditure by walking, stair climbing, and playing sports and to the composite energy expenditure in kcal per week as determined by responses to mail questionnaires. Men expending fewer than 2,000 kcal per week were at 64 percent higher risk than their ex-classmates who were less active. No additional reduction in CHD rate was found in men expending more than 2,000 kcal per week. CHD risk was further increased in men with low levels of physical activity if other risk factors such as cigarette smoking, hypertension, diabetes mellitus, obesity, or a parental history of heart attack were present. The association between risk of CHD and level of physical activity remained strong when adjustments were made for these confounders. Another important finding from this study relevant to the selection versus protection issue was that alumni who had been athletes during college and did not continue exercising were at excess risk for CHD compared with physically active alumni who were not college athletes. It was concluded that only current physical activity is associated with reduced CHD events, independent of previous athletic status. Further, the apparent protective effects of exercise appeared independent of constitutional factors involved in selection for formal athletics during youth.

A study from Holland supports the finding of Paffenbarger and colleagues that moderate intensity activities reduce CHD risk. Magnus et al. (12) obtained information about physical activity by interviewing 473 subjects or close relatives within 4 weeks of their coronary events and 815 control referents from the same communities in central Holland. A significant inverse association was found between acute coronary events and habitual (defined as >8 months per year) walking, cycling, and gardening. This was not true if these activities were performed only occasionally or on a seasonal basis (4–8 months per year). Regular vigorous exercise did not appear to offer any more protection from CHD than did moderate activities.

We studied the relation of duration (min/yr) and intensity of leisure-time physical activity assessed by the

Minnesota questionnaire to age-adjusted, 7-year CHD and total mortality rates in over 1,200 high risk men, aged 35 to 57 years, participating in the NHLBI Multiple Risk Factor Intervention Trial (74). A significant inverse relationship was found between physical activity duration and estimated energy expenditure duration at baseline and CHD mortality in the Special Intervention (SI) group that received intensive hypertension, dietary, and smoking intervention, as well as the group that received usual medical care in the community (UC). The most active quintile had about 0.6–0.7 the CHD mortality rate as the least active. The principal reduction in risk occurred between quintile 1 and 3 or 4. More active men appeared more fit on exercise testing, smoked less, and had slightly higher HDL cholesterol and lower overall CHD risk. When adjustments were made for confounders, the association of physical activity to CHD mortality remained significant although diminished.

Another approach is to study the relationship between physical fitness status and CHD. Peters et al. (61) assessed the physical fitness by bicycle ergometry of 2,779 healthy male fire and law enforcement workers younger than 55 years of age in Los Angeles County, who were then followed for an average of about 5 years. The relative CHD risk adjusted for other conventional CHD risk factors was 2.2 for those workers who had below average exercise capacity at baseline, reflecting below average physical fitness. Below average physical fitness in combination with at least two other risk factors (that is, above median serum cholesterol level, elevation of systolic blood pressure, and/or cigarette smoking) increased the risk ratio to 6.6. A significant positive correlation between leisure-time physical activity and work capacity on exercise testing has previously been demonstrated (8, 75). Further, a negative relationship has been found between work capacity and other CHD risk factors (75, 76).

Population studies

Early population studies used death certificates to assess both physical activity levels (from job titles) and death from CHD. The majority of these studies showed a direct association between jobs requiring low levels of physical activity and CHD, but the problems with this approach are numerous (36).

One of the first population studies to assess both leisure-time and occupational physical activity was a retrospective study of 55,000 men aged 25–64 years enrolled in the Health Insurance Plan (HIP) of New York (77–79). On-the-job and leisure-time physical activities were assessed by a questionnaire and interviews with the insured or his widow. Subjects' activity levels were classified as light, moderate, or heavy. Considerations were made for time spent walking and sitting at work, transportation to and from work, total working hours and time spent lifting and carrying objects. The incidence of myocardial infarction in both the heavy and moderate activity category was about one-half that of the light activity group; furthermore, the least active men had a 4.5 times greater mortality rate following infarction than the most active men. These differentials persisted after adjustments were made for body weight and smoking.

Siscovick et al. (80) compared the prior physical activity status of 163 cases of primary cardiac arrest aged 25–75 years with matched controls in Seattle and suburban King County, Washington. Spouses of subjects and controls were interviewed to quantify vigorous leisure-time activity during the prior year using the Minnesota questionnaire. The risk of cardiac arrest was 55–65 percent lower in the persons in the two upper quartiles of high-intensity physical activity than in persons doing no high-intensity physical activity. Similarly, Hennekins et al. (81) collected retrospective data from wives of 568 men from two Florida counties, aged 30–70 years who died of CHD, and a matched living neighborhood control sample. Physical activity status was classified according to the HIP study criteria. Increased leisure-time physical activity, but not job activity, was found to be associated with decreased risk of CHD death even when cigarette smoking and hypertension were accounted for. These latter two studies, like the HIP study, suffer from the potential bias previously demonstrated that widows underestimate prior physical activity status of their deceased husbands.

A number of major prospective (longitudinal) studies over the past 30 years in the United States and Europe had followed initially healthy people from specific geographic areas for long periods to determine CHD risk characteristics. In these studies, responses to questions about occupational and/or leisure time activity were considered in classifying people as sedentary, moderately active, or very active. The results have been mixed.

The Seven Country Collaborative Study involved 16 cohorts of men in seven countries who were aged 40 through 59 years and free of clinical CHD at the start of the study (82). Differences between 10-year CHD incidence and mortality rate between populations were unrelated to the proportion of sedentary men in the population. Other risk factors, particularly serum cholesterol levels and dietary intake of saturated fat, appeared to better explain differences in CHD rates among countries. Only three of the seven countries showed an inverse association between physical activity (mostly occupational) and CHD, and in the others (e.g., the United States and Finland), there was no apparent association. In Finland, the country with the highest incidence and mortality from CHD in the study, the 10-year followup revealed no difference in CHD mortality between men classified as sedentary as compared with those classified most active, and the rate in the moderately active men was unexplainably twice that of the other two classes. However, there were relatively few men classified as sedentary in the Finnish cohorts. Karvonen (83) followed the two Finnish cohorts in the Seven Country Study for an additional 5 years and reassessed the original 10-year data after studying in greater detail the men's physical activity habits by an extensive structured interview. Reevaluation of the original 10-year data revealed that, for men age 50–69 years, CHD incidence was clearly associated with sedentary habits. Subsequent 5-year CHD mortality and combined fatal and nonfatal myocardial infarction rates in this age group were inversely related to physical activity status; however, the majority of men dying of CHD already had CHD diagnosed prior to the 5-year followup. Thus it is unclear whether physical activity had a protective effect or whether the high incidence of CHD among the least active men was related to already present CHD.

In a more recent 7-year longitudinal study from two counties in eastern Finland, Salonen et al. (14) showed that low levels of physical activity at work were associated with increased risk for myocardial infarction, stroke, and overall mortality. After adjustments were made for age, relative weight, serum total cholesterol, diastolic blood pressure, and smoking, the relative risk of low physical activity at work was 1.5 in men and 2.4 in women compared with more active people. Risk of CHD was also increased to those with low levels of combined job and leisure-time physical activity. However, physical inactivity during leisure by itself was not associated with increased risk of myocardial infarction or stroke, but it was with overall mortality. The investigators speculated that this may reflect a strong positive association of leisure-time physical activity with smoking in Eastern Finland.

In another study from Scandinavia, Wilhelmsen et al. (84) performed an 8-year followup of 973 middle-aged men originally examined in 1967 in Goteborg, Sweden, and found no relationship between occupational activity level and the 49 documented CHD events. However, there was a decreased risk of CHD among men active during leisure time for the year prior to the event, but this association disappeared when adjustments were made for other conventional risk factors. In fact, most of the association between leisure activity and decreased CHD rate could be explained by the negative association between leisure activity and cigarette smoking in this population. Further, the lack of an association between job activity and CHD may be explained, in part, by heavier smoking in physically active workers.

In the Framingham Heart Study (85) an inverse relationship was found between an index of overall job and leisure-time physical activity determined by a questionnaire and 14-year CHD mortality in men, but not in women. The relative risk of development of cardiovascular disease in the least active as compared with the most active level of physical activity index was 1.3 for men 45–64 years of age.

The Puerto Rico Heart Health Program (86) used a similar physical activity rating index as the Framingham Study in evaluating 8,793 men, initially 45–65 years of age. The 8.5-year followup showed an inverse association between incidence of CHD events, other than angina pectoris; the range of risk was twofold from the lowest to highest. Multivariate analysis to correct for confounding by other CHD risk factors confirmed that reduced physical activity contributed independently to risk of CHD. The association persisted after the data was reanalyzed to exclude CHD events in the first 2½ years to eliminate men whose low levels of physical activity might have been caused by subclinical disease.

The Western Collaborative Group Study was a prospective study of employed California men age 39–59 years. The main emphasis in this study was on the relation of Type A behavior pattern to CHD. Occupational physical activity and leisure exercise habits were determined by personal interview. At the 4½ year followup, the incidence of CHD and fatal myocardial infarction was significantly lower in the men with regular exercise habits (87). However, by the final followup after 8–9 years, no significant relation was apparent between physical activity level and CHD for the cohort (88). Subgroup analysis did show that men over 50 years of age who reported exercising daily (including walking) had significantly lower CHD rates than those reporting only occasional exercise.

Pathologic studies

There have been a small number of large-scale postmortem studies correlating physical activity levels with extent of coronary atherosclerosis and myocardial damage. These include studies in Westchester County, New York (89, 90), in Britain (91), in Israel (92), and in Finland (93) as well as an autopsy on Clarence DeMar, an outstanding life-long marathon runner (94).

The following conclusions can be derived:

- Autopsy evidence of death from CHD was twice as frequent in men doing light work as compared with men doing heavy work. In addition, *sedentary men tended to die of CHD at an earlier age than active men.*
- In most studies the frequency of severe coronary atherosclerosis tended to be similar in sedentary, moderately active, and very active men. An exception was the study by Rissonen (93) in which active Finnish men had less severe coronary atherosclerosis than less active or sedentary men. The latter findings are supported by recent experimental data from nonhuman primates receiving atherogenic diets in which regular exercise (jogging on a treadmill 30 minutes three times a week) reduced the severity of coronary atherosclerosis (95).
- The physically active men had significantly larger coronary artery luminal areas as compared with light workers in agreement with experimental findings in exercised rats and monkeys (95–97). In addition, the active men were less likely than inactive men to have total occlusions of major coronary arteries despite advanced atherosclerosis.
- Physically active men had significantly less ischemic myocardial fibrosis and healed infarcts than inactive

men, even in the presence of similar severe coronary atherosclerosis.

Clinical trials

Clinical trials are required to determine whether associations uncovered in epidemiologic observations represent cause-and-effect relations. The validity of the trial depends on having a representative population sample and matching treatment and control groups with respect to characteristics thought to affect outcome. The random assignment of subjects to treatment or control group is essential to equally distribute known and unknown confounding variables between groups. Estimates of the required number of middle-aged men for a clinical trial to test the role of exercise in primary prevention of CHD range from several thousand to several hundred thousand, depending on whether exercise is hypothesized to reduce acute coronary events 5 or 25 percent over an 8-year period (98). There have been no such definitive primary prevention trials nor are any anticipated because of difficulties in design, the large sample size required, costs, and difficulty in maintaining long-term compliance.

An alternative approach is secondary prevention trials on heart attack survivors with the goal of determining whether or not exercise conditioning reduces recurrence rates of myocardial infarctions and associated mortality. A definitive trial using this approach is estimated to require 3,000 myocardial infarction patients free of contraindications to exercise randomized into an exercise or control group and studied for at least 5 years (99). No such definitive large-scale trials have been undertaken. However, there have been four smaller scale randomized secondary prevention trials, two in North America (one in the United States (100), and one in Canada (101)) and two in Scandinavia (one in Goteborg, Sweden (102), and one in Finland (103)). The results of the Canadian and the Swedish studies are inconclusive; however, in the multicenter U.S. National Exercise Heart Disease Project, there was an encouraging trend toward reduction in the exercise group of the cumulative 4-year cardiovascular mortality rates and a significant reduction in overall mortality (100). However, in the Canadian study (101) the trend was in the opposite direction. In Finland, Kallio et al. (103) randomized 375 men and women survivors of myocardial infarction at the time of hospitalization into a multiple risk factor intervention group (including exercise) or control group. The intervention group experienced a significant reduction in total cardiovascular mortality and cardiovascular and sudden death, but not in reinfarction. However, there was no evidence of improved physical fitness on bicycle ergometer testing in the intervention group. Although this study does not provide answers on the independent role of exercise in secondary prevention of CHD, it does foster the hope that exercise as part of a multiple factor intervention program will improve survival.

Effect of physical activity on other CHD risk factors

Both epidemiologic and applied physiologic studies provide evidence that exercise favorably alters other CHD risk factors. This has led to the suggestion by Salonen et al. (14) that controlling for other risk factors in epidemiologic studies of the association between physical activity and CHD may be too conservative since it masks ways that exercise might protect against CHD, that is indirectly by attenuating other CHD risk factors. The effects of exercise on CHD risk factors have previously been reviewed, including effects on blood lipids (104–108), blood pressure (107–110), glucose tolerance (107, 108, 111–113) and cigarette smoking (107, 108).

Physical inactivity has been unfavorably associated with the following risk factors in population studies: obesity, hypertension, blood lipid abnormalities, and maturity-onset (type II) diabetes mellitus. The relation of both physical activity at work and leisure to cigarette smoking habits is mixed. Although the association between serum total cholesterol and physical activity is inconclusive, physically inactive people usually have higher triglyceride and VLDL cholesterol levels and lower HDL cholesterol levels than active people. Increased body fat in inactive people and perhaps heavier smoking may contribute to these differences. Increased relative body weight only partially explains the differences in diabetes rates in physically inactive as compared with active people.

Although exercise is commonly proposed as an aid to smoking cessation, there is little evidence from controlled studies that exercise programs result in long-term cessation of cigarette smoking (107, 108); on the contrary, cigarette smoking appears to be a predictor of noncompliance to exercise (114).

Physiological effects of exercise conditioning that may help to protect against CHD

This topic recently was extensively reviewed (107, 108, 115). The major health-related physiological effects of exercise training are summarized in table 1. These provide plausible mechanisms for the apparent partial protective effect of physical activity or exercise against CHD.

Based on the available data, I recently proposed the mechanisms outlined in figure 1 to explain the multifactoral role of physical activity in reducing risk of clinical manifestations of CHD (43). Physical activity clearly

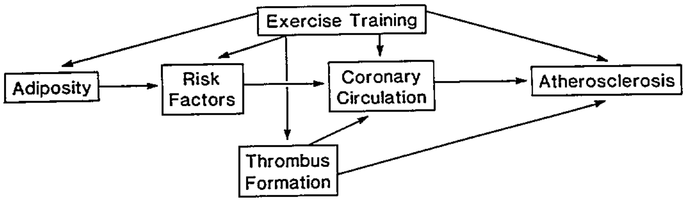


Figure 1.

reduces body adiposity which in turn favorably affects blood pressure and metabolic risk factors. Exercise also directly can favorably alter CHD risk factors. In addition, exercise training favorably affects the balance between coronary blood flow supply and demand and may reduce the likelihood for thrombosis in narrowed coronary vessels by reducing platelet aggregation and promoting fibrinolysis. Further, exercise training may directly, as well as indirectly, attenuate the atherosclerosis process.

Quality and quantity of physical activity apparently required to protect against CHD

Much work is required to define the optimal amounts of exercise to protect against CHD events, and NHANES data may prove useful. However, careful analysis of the available epidemiologic data provides us suggestive evidence of the intensity and quantity of exercise required for reducing risk of CHD. Intermittent spurts of strenuous physical exertion on the job, including lifting, carrying, shoveling, and climbing throughout the working day such as performed by farmers and longshoremen, appear to reduce CHD rates (53, 54, 67, 68, 72). An encouraging finding from the study of North Dakota farm workers suggests that most of this reduced risk can be achieved with as little as 1 hour per day of physical labor (53). The work of Morris et al. (64–66) further suggests that as little as 30 minutes a day of continuous vigorous leisure time activity at an intensity of 7.5 kcal per minute or more can be used as a substitute for or adjunct to job activity to reduce risk of CHD. Data from Paffenbarger et al. (72, 73) and Magnus et al. (12) indicate that moderate intensity exercise, such as brisk walking, stair climbing, cycling, and gardening, if done regularly year round, also will reduce chances for CHD. NHANES data may provide more information on quantity and quality of physical exertion required to protect against CHD.

Effect of physical activity status on other health problems

Dynamic exercise has been advocated in the prevention and treatment of several other important health problems in our society in addition to CHD. These include obesity, hypertension, and diabetes mellitus. It also may be beneficial as adjunct therapy in other chronic conditions, including peripheral vascular disease, chronic obstructive pulmonary disease, cystic fibrosis, and osteoporosis. However, there have only been a limited number of studies evaluating the effects of exercise in any of these conditions. I will review briefly some of the evidence of a relationship of physical activity to some of these conditions, which have been better studied.

Obesity

NHANES I defined as obese any "adult with a triceps skinfold measurement greater than the 85th percentile measurement for people 20–29 years of the same sex." According to this criterion, 13 percent of all men and 23 percent of all women in the age range 20–74 years were obese in NHANES I. Other studies estimate that from 40–80 million individuals in our country are significantly overweight and obese. The prevalence of obesity increases with age and appears to parallel a reduction in job and recreational physical activity during this period of life (20, p. 61). Also obese adolescents and children have been observed to be less physically active than their nonobese peers. Data from NHANES III may be useful for verification of these apparent relationships.

Studies on the effects of exercise alone on obesity reveal a modest short-term weight loss; however, none to date have addressed the more critical problem of long-term weight control. Several mechanisms appear to contribute to the loss of weight and body fat through regular physical activity. The principal mechanism is increased energy expenditure during the activity. Because 1 pound of body fat contains approximately 3,500 kcal of latent energy and because walking or running a mile burns up approximately 100 kcal, theoretically one would lose about 1 pound for each 35 miles of walking or running if food intake were not increased. A small additional increased energy expenditure occurs during recovery (116). Regular exercise may also help reduce appetite and preserve lean body mass during associated weight loss (112). Reduced circulating insulin levels during prolonged activity along with elevations of counterregulatory hormones accelerate fat mobilization from adipose tissue (111–113).

The high prevalence of overweight and obesity in our society is of public health concern because of the association of 20% or more overweight with excess mortality. In addition, obesity is a risk factor for hypertension, maturity-onset diabetes mellitus, and CHD.

Hypertension

About 25 million American adults in the United States have hypertension. Blood pressure measurements made on people 6–74 years of age as part of NHANES I showed a prevalence rate of hypertension of 18 percent among those 18–74 years of age. In addition, black people had higher rates than white people. Hypertension is of considerable public health concern because it is a major predictor and precursor of heart attack and stroke. Whether level of physical activity is related to hypertension independent of body weight is controversial. This topic was recently reviewed in detail by Blackburn and me (109). No consistent relationship has been noted between physical activity at the job and either blood pressure levels or prevalence or incidence of hypertension. Similarly, no consistent association is evident between leisure-time physical activity and blood pressure levels or prevalence of hypertension. However, in most studies, risk of hypertension increased with relative weight which, as we have indicated, may be inversely related to physical activity. Further, a long term followup study in Harvard alumni showed that exercise significantly reduced risk of hypertension in men 25% or more overweight for their height, but not normal weight men (73). In addition, a lower incidence of acute coronary events was observed in physically active as compared with inactive hypertensives. There have been few controlled studies on the use of dynamic aerobic exercise training in the management of hypertension. In these studies, its value appeared limited as compared with the blood pressure lowering effects of weight reduction, a low sodium diet, and antihypertensive drugs.

Diabetes mellitus

Diabetes is another major public health problem, with about 11 million diabetics in the United States. About 90 percent of these have adult or maturity-onset diabetes. Epidemiologic observational studies have shown that about 80 percent of maturity-onset diabetics are obese (111). Sedentary life style is another important risk factor. At least part of the latter association is through the contribution of inactivity to obesity; however, experimentally forced inactivity as provided by bed rest within days leads to cell resistance to insulin with resulting hyperinsulinemia and glucose intolerance (117), the hallmarks of maturity-onset diabetes (117), suggesting that inactivity directly contributes to etiology. In addition, physical activity appears to increase numbers of cell insulin receptors, which reduce insulin secretion. Further, exercise conditioning, with and without associated weight loss in obese nondiabetic subjects, has been shown to decrease glucose stimulated blood insulin levels (113). Thus, a physically active life style may protect against diabetes both indirectly through its influence on weight and directly through improved cell insulin sensitivity. NHANES should further explore this relationship between physical activity status and diabetes.

Adverse effects of exercise

Scope of the problem

Irrespective of the benefits, strenuous exercise carries some risks. Remarkably little population-based data exist on morbidity and mortality associated with exercise or associated predisposing factors (118). Careful epidemiologic surveys, cohort studies, and case comparison studies are necessary to define rates of injuries with different modes of exercise and sports in defined populations at risk, during a given time frame, along with factors associated with the injuries. Potential factors include host characteristics, intensity and duration of different forms of exercise and sports, and the host's perception of the exercise benefits as compared with risks. NHANES may be able to provide such data.

The existing literature on the risks of exercise is comprised primarily of anecdotal case reports or collection of case reports or "clinical series." Such reports fail to define the population at risk. Koplan et al. (118) outlined the following problems that need to be addressed in order to collect meaningful data on injuries from surveys:

Definitions of injuries and what is meant by regular exercise.
Characterization of subgroups within the population most susceptible to injuries with various forms of exercise.
Correlations of injuries to quantity and intensity of different types of exercise.
Define cases of injuries or events in the context of populations at risk, i.e., the denominator population.
Comparison of cases or events with matched controls.
Determinations of the injury rate in terms of a time interval, permitting determination of an incidence, e.g., 10 injuries per 1,000 participants per year.
Comparison of risks and benefits for specific forms of exercise using a utility analysis.

Musculoskeletal injuries

What is apparent from the current clinical series are the injuries most likely to be associated with various types of exercise. It is not surprising that during dynamic exercise, skeletal muscle and connective tissue structures that attach them to bones and joints may be injured and physically damaged. Common exercise-specific musculoskeletal or orthopedic injuries have recently been reviewed (118–121).

Risk of orthopedic injuries appears to be related to the type, intensity, frequency, and duration of exercise. It is apparent that "overuse or abuse" of exercise is commonly an associated factor. Other potential sources of injuries and accidents are related to sports, e.g., head, neck, and knee injuries with football; traumatic injuries from falling from bicycles or horses or being struck by a car while jogging, walking, or cycling; eye injuries during racquet sports; dehydration; heat and cold injuries; and amenorrhea with habitual vigorous exercise in women.

Myocardial infarction and sudden death

The occurrence of a myocardial infarction or sudden cardiac death (SCD) during or shortly following exercise, particularly jogging, is newsworthy because of the paradox involved as well as because of its rarity. SCD can be defined as a witnessed, unexpected, nontraumatic fatality in a subject, with or without preexisting cardiac disease, who succumbs within an hour or less up to 24 hours from the onset of the terminal event (122). Coronary heart disease is implicated in over 75% of all cases of SCD, including those occurring during jogging or sports (123–125). Standard coronary risk factors are usually present and often prodromal symptoms were ignored by both joggers and inactive people dying suddenly. Hypertrophic obstructive cardiomyopathy is the most common cause of SCD in young athletes under age 30, with other structural cardiovascular abnormalities accounting for most of the rest (126, 127). The mechanism of death in SCD is generally ventricular fibrillation usually secondary to underlying coronary heart disease (122).

Siscovick et al. (80) found in a case control study that the risk of cardiac arrest during exercise is inversely related to usual levels of vigorous habitual activity. These findings are compatible with other observational epidemiologic studies that persons who regularly engage in physical activity have a reduced rate of SCD. Nevertheless, it appears from the Siscovick study, as well as from another case control study from Rhode Island (123) in which an estimate was made of the base population at risk, that even a habitually active person is transiently at greater risk of SCD while performing vigorous exercise compared with resting.

Although the greater the habitual vigorous activity the lower the risk (80), the excess risk while jogging compared with resting in Rhode Island was about seven times. The incidence of jogging deaths for men in Rhode Island aged 30 through 64 years was one per 7,620 joggers or about one death per 396,000 hours of jogging. From this review of the literature, it appears that risk of sudden cardiac death while exercising is small but definite even in experienced runners. This requires verification in a national survey such as NHANES as does the determination in general of the relative risk: benefit ratio of exercising regularly.

Summary and conclusions

Observational epidemiologic studies suggest that living habits and other environmental factors are important causative factors in most of the chronic diseases in adults that dominate the morbidity and mortality statistics in the United States. Regular vigorous physical activity, maintenance of normal weight, proper diet, no smoking, abstinence or limited use of alcohol are postulated to contribute to maintenance or improvement of health. Increased regular exercise to improve physical fitness has been officially recommended as part of any strategy for public health promotion and disease prevention by numerous official governmental and professional health organizations around the world. However, if exercise and physical fitness are to be promoted as a preventive medicine or public health measure, documentation is required that they have a favorable effect on health and disease prevention, that the benefits exceed the risks of exercise (and of inactivity), and of the optimal amounts of exercise to favorably affect the risk to benefit balance. To study these relationships requires accurate ways of documenting levels of physical activity and physical fitness during population surveys, such as NHANES.

More than 30 different approaches are currently available to assess physical activity. These can be grouped into seven major categories: direct or indirect calorimetry; job classifications; activity questionnaires, interviews, or diaries; observer monitoring; mechanical and electronic movement or heart rate detectors; calories per day or per kilogram of body weight from dietary measurements, and by physical fitness assessment or physiological markers. A good instrument should accurately measure physical activity and be practical in terms of time required, cost, and convenience. Most studies assessing the association of physical activity to disease and mortality rates have employed job and/or leisure-time physical activity questionnaires or interviews. However, most of these survey instruments have only been partially validated, if at all.

Few major observational epidemiologic studies have actually employed physiological measures of physical fitness. Cardiorespiratory endurance or aerobic capacity is the most important health-related component of physical fitness. It can be determined in the laboratory by measuring maximal oxygen uptake ($\dot{V}O_2$ max) during all-out exercise on a step test, treadmill, or bicycle ergometer. It can be estimated with a fair degree of accuracy by safer, easier to administer, submaximal step, treadmill, or bicycle ergometer exercise tests or by field tests (e.g., a 2.0 mile run). Also fitness can be assessed indirectly by measuring resting heart rates, heart rates during submaximal exercise, vital capacity or forced 1-second respiratory volume (FEV_1), and HDL cholesterol levels. Relative weight and skinfold thickness measurements are also easy to obtain indirect indicators of fitness. However, the association between fitness variables and physical activity in population surveys are usually weak.

Evidence for the possible health benefits of vigorous physical activity come from a variety of sources. These include folklore and anecdotal information, physiologic, observational-epidemiologic, and experimental studies. Our bodies through evolution appear to be well adapted for regular, moderate-intensity, dynamic physical activity and intermittent high-intensity exertion.

Because physical *inactivity* results in substantial, reproducible, and reversible physiologic changes that mimic human aging, it has been hypothesized that exercise can retard the aging process and perhaps prolong the life span. In support of this hypothesis are findings in elderly athletes of levels of maximal oxygen uptake and strength similar to much younger sedentary people. Furthermore, physical fitness can be improved at all ages by exercise conditioning; and exercise can partially reverse physiological and metabolic changes usually attributed to advancing age. However, there is no definitive evidence available that exercise improves longevity. Observational studies comparing the longevity of college, Olympic, and professional former athletes with nonathletes have yielded mixed results. Interpretation of such studies is also fraught with difficulty because of the large number of confounding variables in comparing athletes with nonathletes. Further, most of the college athletes in these studies did not continue exercise after graduation, and it is well known that the training effects of exercise are only transitory.

There is stronger, but still inconclusive, evidence from observational epidemiologic studies of an inverse association between physical activity, generally assessed by questionnaires, and fatal and nonfatal coronary events. The strength of this relationship is less than the three major risk factors, serum cholesterol, blood pressure, and cigarette smoking. Plausible mechanisms to explain this relationship have come from experimental studies. These include a direct conditioning effect on the heart improving functional capacity; improvement in the blood lipid/lipoprotein profile, and perhaps in other coronary risk factors; and reduced body weight and fat. The amount and intensity of physical activity to provide such protection is still uncertain. There is suggestive epidemiologic and experimental evidence that 2,000 kcal a week of moderate intensity dynamic activities is sufficient. However, some studies suggest that vigorous exercise is required for protection. This needs to be resolved in order to give an appropriate public health message about exercise. There is also suggestive evidence that physical activity may protect against other chronic diseases associated with the modern life style that contribute to premature mortality, including maturity-onset diabetes mellitus, hypertension and osteoporosis. This also requires confirmation. The risks of sensible amounts of exercise for most people in our society are minimal and appear to be less than that associated with a sedentary life style. However, in the presence of significant coronary atherosclerosis, there is an increased risk of sudden cardiac death during vigorous exercise as compared with sedentary activities. If less vigorous physical activity offered comparable protective effects against these medical problems at less risk, this would have important public health implications.

It is concluded that there is sufficient evidence that regular physical activity and exercise can improve physiological and metabolic functions and perhaps reduce the risk of common chronic diseases that are the main contributors to premature mortality in our society. Therefore, it is warranted to monitor physical activity and physical fitness in NHANES to determine the effects of chronic disease prevalance and mortality rates on national trends and to help define the quantity and quality of physical activity required.

Surveillance of physical activity should include the general use of questionnaires of leisure and job activities and the determination of energy intake in kcal per day (from dietary recalls) per kg. of body weight. It does not appear to me to be feasible in NHANES to assess cardiorespiratory endurance, the most important component of physical fitness, by *direct measurement* of $\dot{V}O_2$ max. However, this component of physical fitness could be estimated in NHANES using a submaximal exercise test on a treadmill or bicycle ergometer, for example, by determining the workload to achieve a heart rate of 140–170 bpm depending on age. A systematic sample of the study population could have more intensive physical activity interviews and monitoring of heart rate by ambulatory ECG monitoring and/or movement by mechanical detectors to further validate the physical activity questionnaires and the submaximal exercise tests. In addition, body mass index, skinfold thickness measurements, FEV_1, and HDL cholesterol levels will provide other indirect indexes of physical fitness.

References

1. Casperson, C.J., Powell, K.E., and Christenson, G.M.: Physical activity, exercise and physical fitness: Definition and distinctions of health-related research. *Public Health Reports.* 100:126–130, 1985.
2. Leon, A.S., and Blackburn, H.: Physical Inactivity. In Kaplan, N.M., and Stamler, J. (eds), *Prevention of Coronary Heart Disease. Practical Management of the Risk Factors.* W. B. Saunders, Philadelphia, 1983, pp. 86–97.
3. Haskell, W.L., Montoye, H.J., and Orenstein, D.: Physical activity and exercise to achieve health-related physical fitness components. *Public Health Reports.* 100:202–211, 1985.
4. LaPorte, R.E., Montoye, H.J., and Casperson, C.J.: Assessment of physical activity in research: Problems and Prospects. *Public Health Reports.* 100:131–146, 1985.
5. Baranowski, T., Dworkin, R.J., Cieslik, et al: Reliability and validity of self report of aerobic activity; Family Health Project. *Res. Quarterly for Exer. & Sport.* 55:309–317, 1984.
6. Taylor, H.L., Jacobs, D.R., Jr., Schucker, B., Knudsen, J., Leon, A.S., and DeBacker, G.: A questionnaire for assessment of leisure time physical activity. *J. Chronic Dis.* 31:741–755, 1978.
7. Montoye, H.J.: Estimation of habitual physical activity by questionnaire and interview. *Am. J. Clin. Nutri.* 24:1113–1118, 1971.
8. Leon, A.S., Jacobs, D.R.,. Jr., DeBacker, G., and Taylor, H.L.: Relationship of physical characteristics and life habits to treadmill exercise capacity. *Am. J. Epidemiol.* 113:653–660, 1981.
9. Folsom, A.R., Casperson, C.J., Taylor, H.L., et al: Leisure time physical activity and its relationship to coronary risk factors in a population-based sample. The Minnesota Heart Survey. *Am. J. Epidemiol.* 121:570–579, 1985.
10. Blair, S.: How to assess energy expenditure and physical fitness. In Matarazzo, J.D., et al (eds), *Behavior and Health. A Handbook of Health Enhancement and Disease Prevention.* John Wiley and Sons, New York, 1984, pp. 424–447.
11. Paffenbarger, R.S., and Wing, A.L.: Chronic disease in former college students. X. The effects of single and multiple characteristics on risk of fatal coronary heart disease. *Am. J. Epidemiol.* 90:527–535, 1969.
12. Magnus, K., Matroos, A., and Strackee, J.: Walking, cycling, and gardening, with or without seasonal interruption, in relation to acute coronary events. *Am. J. Epidemiol.* 110:724–733, 1979.
13. Gordon, D.J., Probstfield, J.L., Rubenstein, C., Bremmer, W.F., Third, J., Karon, J., Bryan, H., Insull, W., Leon, A.S., Schwartz, L., and Sheffield, L.T.: Proportional hazards analysis of the relationship of coronary risk factors and exercise test performance in asymptomatic hypercholesterolemic men. *Am. J. Epidemiol.* 120(2):210–224, 1984.
14. Salonen, J.T., Puska, P., and Tuomieato, J.: Physical activity and risk of myocardial infarction, cerebral stroke, and death. A longitudinal study in East Finland. *Am. J. Epidemiol.* 115:526–537, 1982.
15. LaPorte, R.E., Cauley, J.A., Kinney, C.M., Corbett, W., Robertson, R., Black-Sander, R., Kuller, L.H., et al: The epidemiology of physical activity in children, college students, middle aged men, menopausal females and monkeys. *J. Chronic Dis.* 35:787–795, 1982.
16. Montoye, H.J., Washburn, R., Servais, S., Ertl, A., Webster, J.G., and Nagle, F.J.: Estimation of energy expenditure by a portable accelerometer. *Med. Sci. Sports Exer.* 15:403–407, 1963.

17. Kleges, R.C., Coates, T.J., and Moldenhauer, L.M.: The fats: An observational system for assessing physical activity in children and associated parent behavior. *Behav. Assessment* 6:333–345, 1984.
18. Sopko, G., Jacobs, D.R., Jr., and Taylor, H.: Dietary measures of physical activity. *Am. J. Epidemiol.* 120:900–911, 1984.
19. Leon, A.S. (ed): Forum Exercise and Health. *Preventive Med.* 13:1–99, 1984.
20. Thomas, G.S., Lee, P.R., Franks, P., and Paffenbarger, R.S., Jr.: *Exercise and Health. The Evidence and the Implications.* Oeleschlager, Grunn, and Hain, Publishers, Cambridge, MA, 1981.
21. Leon, A.S. and Fox, S.: Physical Fitness. In Wynder, E.L., Ertzberg, S., and Parker, E. (eds), *The Book of Health.* Franklin Watts, New York, 1981, pp. 283–341.
22. Stromme, S.B., Frey, H., Harlem, O.K., Stokke, O., Vellar, O.D., Aaro, L.E., and Johnson, J.E.: Physical activity and health. Part I and Part II. *J. Cardiac Rehab.* 4:316–326 and 364–375, 1984.
23. Blackburn, H.: Physical activity and coronary heart disease: a brief update and population view (Part 1). *J. Cardiac Rehab.* 3:101–111, 1983.
24. Retzlaff, E.J., Fontaine, J., and Furuta, W.: Effect of daily exercise on life-span of albino rats. *Geriatrics* 21:171–177, 1966.
25. Taylor, H., Herschel, A., Brozek, J., and Keys, A.: Effects of bed rest on cardiovascular function and work performance. *J. Appl. Physiol.* 2:223–239, 1949.
26. Saltin, B., Blomqvist, G., Mitchel, J., Johnson, L., Jr., et al: Responses to exercise after bed rest and after training: a longitudinal study of adaptive changes in oxygen transport and composition. *Circulation.* 33(Suppl. 7):1–78, 1968.
27. Bortz, W.M., II: Disuse and aging. *JAMA* 248:1203–1208, 1982.
28. Simon, H.B.: The immunology of exercise. A brief review. *JAMA* 252:2735–2738, 1984.
29. Taylor, H.L., and Montoye, H.J.: Physical fitness, cardiovascular function and age. In Ostfeld, A. and Gibson, D. (eds), *Epidemiology of Aging,* U.S. Dept. HEW, U.S. Gov't. Printing Office, Washington, D.C., 1973, pp. 223–241.
30. Hodgson, J.L., and Buskirk, J.: Physical fitness and age, with emphasis on cardiovascular function in the elderly. *J. Am. Geriatrics Soc.* 25:385–392, 1977.
31. Shephard, R.J.: *Physical activity and aging.* Year Book Medical Publishers, Chicago, 1978, p. 103.
32. Shock, N.W., and Morris, A.H.: Neuromuscular coordination as a factor in age changes in muscular exercise. In Brunner, D. and Jokl, E. (eds), *Physical Activity and Aging, Medicine and Sports,* Vol. 4. Karger, Basel/New York, 1970, pp. 92–99.
33. Leon, A.S.: Exercise and risk of coronary heart disease. In Eckert, H.M., and Montoye, H.J. (eds), *Exercise and Health. American Academy of Physical Education Papers.* Champaign, Ill., Human Kinetics Publishing, 1984, pp. 14–29.
34. Leon, A.S.: Exercise and coronary heart disease. *Hosp. Med.* 4:38–57, 1983.
35. Smith, E.L., Reddan, W., and Smith, P.E.: Physical activity and calcium modalities for bone mineral increase in aged women. *Med. Sci. Sports Exerc.* 13:60–64, 1981.
36. Leon, A.S. and Blackburn, H.: The relationship of physical activity to coronary heart disease and life expectancy. *Ann. N.Y. Acad. Sci.* 301:561–578, 1977.
37. Karvonen, M.J.: Endurance sports, longevity and health. *Ann. N.Y. Acad. Sci.* 301:653–655, 1977.
38. Polednak, A.P. (ed): *The Longevity of Athletes.* Charles C. Thomas, Springfield, Ill., 1979.
39. Metropolitan Life Insurance Co. Longevity of major league baseball players. *Statistical Bull.,* 204, Apr. 1975.
40. Paffenbarger, R.S., Jr., Wing, A.L., and Hyde, R.T.: Chronic disease in former college students. XVI. Physical activity as an index of heart attack risk in college alumni. *Am. J. Epidemiol.* 108:161–175, 1978.
41. Rose, C.L., and Cohen, M.L.: Relative importance of physical activity for longevity. *N.Y. Acad. Sci.* 30:671–677, 1977.
42. Belloc, M.B., and Breslow, L.: Relationship of physical health status and health practices. *Preventive Med.* 1:409–421, 1972.
43. Leon, A.S.: Physical activity levels and coronary heart disease. Analysis of epidemiologic and supporting studies. *Med. Clinics N.A.* 69:3–40, 1985.
44. Paffenbarger, R.S. and Hyde, R.T.: Exercise in the prevention of coronary heart disease. *Preventive Med.* 13:3–22, 1984.
45. Morris, J.N., Heady, J.A., Raffle, P.A.B., et al: Coronary heart disease and physical activity of work. *Lancet* 2:1053–1057, 1111–1120, 1953.
46. Morris, J.N.: *Uses of Epidemiology.* Edition 3, New York, Churchill Livingstone, 1975.
47. Morris, J.N., Kagan, A., Pattison, D.C., et al: Incidence and prediction of ischaemic heart-disease in London busmen. *Lancet* 2:552–559, 1966.
48. Kahn, H.A.: The relationship of reported coronary disease mortality to physical activity of work. *Am. J. Public Health* 53:1058–1067, 1963.
49. Taylor, H.B., Blackburn, H., Keys, A., et al: Five-year follow-up of employees of selected U.S. Railroad companies. *Circulation.* 41(Suppl. 1):120–139, 1970.
50. Taylor, H.L., Kleptar, E., Keys, A., et al: Death rates among physically active and sedentary employees of the railroad industry. *Am. J. Public Health* 52:1697–1707, 1962.
51. Taylor, H.L., Menotti, A., and Puddu, V.: Five year of follow-up of railroad men in Italy. *Circulation.* 41(Suppl. 1):113–122, 1970.
52. Pomrehn, P.R., Wallace, R.B. and Burmeister, L.F.: Ischemic heart disease mortality in Iowa farmers. The influence of life-style. *JAMA* 248:1073–1076, 1982.
53. Fox, S.M., III, and Haskell, W.L.: Physical activity and health maintenance. *J. Rehabil.* 32:89–92, 1966.
54. Zukel, W.J., Lewis, R.H., Enterline, P.E., et al: A short-term community study of the epidemiology of coronary heart disease. *Am. J. Public Health* 49:1630–1639, 1959.
55. Cassell, J., Heyden, S., Bartel, A.C., et al: Occupation and physical activity and coronary heart disease. *Arch. Intern. Med.* 128:920–928, 1971.
56. Hinkle, L.E., Whitney, L.A., Lehman, E.W., et al: Occupation, education, and coronary heart disease. *Science* 161:238–246, 1968.
57. Mortensen, J.M., Stevensen, T.T. and Whitney, L.H.: Mortality due to coronary heart disease analyzed by broad occupational groups. *Arch. Indust. Health* 19:1–4, 1959.
58. Paul, O.M., Lepper, M.H., Phelan, W.H., et al: A longitudinal study of coronary heart disease. *Circulation.* 28:20–31, 1963.
59. Stamler, J.D., Berkson, D.M., Lindberg, H.A., et al: Long term epidemiologic studies on the possible role of physical activity and physical fitness in the prevention of premature clinical coronary heart disease. In Brunner, D. and Jokl, E. (eds), Medicine and Sports. Vol. 4. *Physical Activity and Aging.* Baltimore, University Park Press, 1970.
60. Stamler, J., Lindberg, H.A., Berkson, D.M., et al: Prevalence and incidence of coronary heart disease in strata of the labor force of a Chicago industrial corporation. *J. Chronic Dis.* 11:405–420, 1960.
61. Peters, R.K.L., Cady, L.D., Jr., Bischoff, D.P., et al: Physical fitness and subsequent myocardial infarction in healthy workers. *JAMA* 249:3052–3066, 1983.
62. Chapman, J.M., and Massey, F.J.: The inter-relationship of serum cholesterol, hypertension, body weight and risk of coronary heart disease: Results of the first ten years follow-up in the Los Angeles heart Study. *J. Chronic Dis.* 18:933–947, 1964.
63. Chave, S.P.W., Morris, J.N., Moss, S., et al: Vigorous exercise in leisure time and the death rate: A study of male civil servants. *J. Epidemiol. Community Health.* 32:239–243, 1978.
64. Epstein, L., Miller, G.T., Sitt, F.W., et al: Vigorous exercise in leisure-time coronary-risk factors, and resting electrocardiograms in middle-aged civil servants. *Br. Heart J.* 38:403–409, 1976.

65. Morris, J.N., Adams, C., Chave, S.P.W., et al: Vigorous exercise in leisure-time and the incidence of coronary heart disease. *Lancet* 21:333–339, 1973.
66. Morris, J.N., Pollard, R., Everitt, M.D., et al: Vigorous exercise in leisure-time: Protection against coronary heart disease. *Lancet* 2:1207–1210, 1980.
67. Paffenbarger, R.S., Jr., and Hale, W.E.: Work-activity and coronary heart disease mortality. *N. Engl. J. Med.* 292:545–550, 1975.
68. Paffenbarger, R.S., Jr., Hale, W.E., Brand, R.J., et al: Work-energy level, personal characteristics, and fatal heart attack: A birth cohort effect. *Am. J. Epidemiol.* 105:200–213, 1977.
69. Punsar, S., and Karvonen, M.: Physical activity and coronary disease in population from East and West Finland. *Adv. Cardiol.* 18:196–207, 1976.
70. Brunner, D., Manens, G., Modan, M., et al: Physical activity at work and the incidence of myocardial infarction, angina pectoris and death rate due to ischemic heart disease: An epidemiological study in Israeli collective statements (kibbutzim). *J. Chronic Dis.* 27:217–233, 1974.
71. Rose, G.: Physical activity and coronary heart disease. *Proc. R. Soc. Med.* 62:1183–1188, 1969.
72. Paffenbarger, R.S., Jr., and Hyde, R.T.: Exercise in the prevention of coronary heart disease. *Prev. Med.* 13:3–22, 1984.
73. Paffenbarger, R.S., Jr., Hyde, R.T., Jung, D.L., and Wing, A.L.: Epidemiology of exercise and coronary heart disease. *Clinics Sports Med.* 3:297–318, 1984.
74. Leon, A.S., Connett, J., Jacobs, D.R., Jr., and Taylor, H.L.: Relation of leisure time physical activity to mortality in the Multiple Risk Factor Intervention Trial (MRFIT). *Circulation* 70(Suppl. 2):II 64, 1984.
75. Leon, A.S., Gordon, D.J., Ekelund, L-G, Sopko, G., Probstfield, J., Rubenstein, C., and Sheffield, L.T.: Relation of treadmill exercise test performance to habitual vigorous exercise and coronary risk factors in apparently healthy hypercholesteremic men. *Med. Sci. Sports Exer.* 17(2):221, 1985.
76. Cooper, K.H., Pollock, M.L., Martin, R.P., White, S.R., Linnerud, A.C., and Jackson, A.: Physical fitness vs selected coronary risk factors. A cross-sectional study. *JAMA* 236:166–169, 1976.
77. Frank, C.W.: The course of coronary heart disease. Factors relating to prognosis. *Bull. N.Y. Acad. Med.* 49:900–915, 1968.
78. Frank, C.W., Weinblatt, E., Shapiro, S., et al: Physical inactivity as a lethal factor in myocardial infarction among men. *Circulation.* 34:1022–1033, 1966.
79. Shapiro, S., Weinblatt, E., Frank, C.W., et al: Incidence of coronary heart disease in a population insured for medical care (HIP): Myocardial infarction, angina pectoris, and possible myocardial infarction. *Am. J. Public Health.* 59(Suppl. 2):1–101, 1969.
80. Siscovick, D.S., Weiss, N.S., Fletcher, R.H. and Lasky, T.: The incidence of primary cardiac arrest during vigorous exercise. *N. Engl. J. Med.* 311:874–877, 1984.
81. Hennekins, C.H., Rosner, J., Jesse, M.J., et al: A retrospective study of physical activity and coronary deaths. *Int. J. Epidemiol.* 6:243–246, 1977.
82. Keys, A.: *Seven Countries: A Multivariate Analysis of Death and Coronary Disease.* Cambridge, Mass., Harvard University Press, 1980.
83. Karvonen, M.J.: Physical activity in work and leisure time in relation to cardiovascular diseases. *Ann. Clin. Res.* 14(Suppl. 34):118–123, 1982.
84. Wilhelmsen, L., Tibblin, G., Aurell, M., et al: Physical activity, fitness and risk of myocardial infarction. *Adv. Cardiol.* 18:217–230, 1976.
85. Kannel, W.B., and Sorlie, P.: Some health benefits of physical activity. The Framingham Study. *Arch. Intern. Med.* 139:857–861, 1979.
86. Garcia-Palmieri, M.R., Costas, R., Jr., Cruz-Vidal, M., et al: Increased physical activity: A protective factor against heart attacks in Puerto Rico. *Am. J. Cardiol.* 50:749–755, 1982.
87. Rosenman, R.H., Brand, R.J., Jenkins, C.D., et al: Coronary heart disease in Western Collaborative Group Study: Final follow-up experience of 8½ years. *JAMA* 23:871–877, 1975.
88. Rosenman, R.H., Friedman, M., Straus, R., et al: Coronary heart disease in the Western Group Study: A follow-up experience of 4½ years. *J. Chronic Dis.* 28:178–190, 1970.
89. Spain, D.M., and Bradess, V.A.: Sudden death from coronary atherosclerosis, age, race, sex, physical activity and alcohol. *Arch. Intern. Med.* 100:228–231, 1957.
90. Spain, D.M., and Bradess, V.A.: Occupational physical activity and the degree of coronary atherosclerosis in "normal" men. A postmortem study. *Circulation.* 22:239–242, 1960.
91. Morris, J.N., and Crawford, M.D.: Coronary heart disease and physical activity of work-evidence of a national necropsy survey. *Br. Med. J.* 2:1488–1496, 1958.
92. Mitrani, Y., Karplus, H., and Bunner, D.: Coronary atherosclerosis in cases of traumatic death. In Bruner, D. and Jokl, E. (eds): Medicine and Sport. Vol. 4, *Physical Activity and Aging.* Baltimore, University Park Press, 1970.
93. Rissanen, V.: Occupational physical activity and coronary artery disease. A clinicopathologic appraisal. *Adv. Cardiol.* 18:113–121, 1976.
94. Currens, J.H., and White, P.D.: Half century of running: Clinical, physiological, and pathological findings in the case of Clarence DeMar ("Mr. Marathon"). *N. Engl. J. Med.* 265:988–993, 1961.
95. Kramsch, D.M., Aspen, A.J., Abramowtiz, B.M., et al: Reduction of coronary artherosclerosis by moderate conditioning exercise in monkeys on an atherogenci diet. *N. Engl. J. Med.* 305:1483–1488, 1981.
96. Leon, A.S.: Comparative cardiovascular adaptation to exercise in animals and man and its relevance to coronary heart disease. In Bloor, C.M. (ed), *Comparative Pathophysiology of Circulatory Disturbances.* New York, Plenum Publishing Co., 1972.
97. Leon, A.S., and Bloor, C.M.: The effects of complete and partial deconditioning on exercise-induced cardiovascular changes in the rat. *Adv. Cardiol.* 18:81–92, 1977.
98. Leon, A.S., and Blackburn, H.: Physical activity in the prevention of coronary heart disease: An update, 1981. In Arnold, C.B., Kuller, L.H. and Greenlick, M.R. (eds), *Advances in Disease Prevention, Vol. 1.* New York, Springer Publishing Co., 1981.
99. Leon, A.S., and Blackburn, H.: Exercise and coronary heart disease. *Minn. Med.* 57:106–107, 1974.
100. Naughton, J.: The National Exercise and Heart Disease Project. Effect of prescribed, supervised exercise program on mortality and cardiovascular morbidity in patients after myocardial infarction. *Am. J. Cardiol.* 48:39–46, 1981.
101. Rechnitzer, P.A., Cunningham, D.A. and Andrew, G.M., et al: Relation of exercise to the recurrence rate of myocardial infarction in men: The Ontario Exercise-Heart Collaborative Study. *Am. J. Cardiol.* 48:39–46, 1981.
102. Wilhelmsen, L., Sanne, H., Elmfeldt, D., et al: A controlled trial of physical training after myocardial infarction: Effects on risk factors, non-fatal infarction and death. *Prev. Med.* 4:491–508, 1975.
103. Kallio, V., Hamalainen, H., Hakkila, J., et al: Reduction in sudden deaths in multifactoral intervention programme after acute myocardial infarction. *Lancet* 2:1091–1094, 1979.
104. Haskell, W.L.: The influence of exercise on concentrations of triglycerides and cholesterol in human plasma. In Terjung, R.L. (ed.), *Exercise and Sports Sciences Reviews, Vol. 12,* Lexington, Mass., Collamore Press, 1984.
105. Haskell, W.L.: Exercise-induced changes in plasma lipids and lipoproteins. *Prev. Med.* 13:23–26, 1984.
106. Sopko, G., Leon, A.S., Jacobs, D.R., Jr., Foster, N., et al: The effects of exercise and weight loss on plasma lipids in young obese men. *Metabolism.* 34:227–236, 1985.
107. Leon, A.S.: Exercise and coronary heart disease. *Hosp. Med.* 4:38–57, 1983.
108. Leon, A.S.: Exercise and risk of coronary heart disease. In Eckert, H.M. and Montoye, H.J. (eds), *Exercise and Health.* American

Academy of Physical Education Papers. Champaign, Illinois, Human Kinetics Pub. Inc. 1984.

109. Leon, A.S., and Blackburn, H.: Physical activity and hypertension. In Sleight, P. and Freis, E. (eds), *Cardiology.* Vol. 1: *Hypertension.* London and New York, Butterworth Scientific, 1982.

110. McMahon, M., and Palmer, R.M.: Exercise and Hypertension. *Med. Clin. N.A.* 69:57–70, 1985.

111. Leon, A.S.: Diabetic patient and athletic performance. In Haskell, W., Scala, J. and Whitman J. (eds), *Nutrition and Athletic Performance.* Palo Alto, California, Bull Publishing Co., 1982.

112. Leon, A.S., Conrad, J, Hunninghake, D.B., and Serfass, R.: Effects of a vigorous walking program on body composition, and carbohydrate and lipid metabolism in obese young men. *Am. J. Clin. Nutr.* 32:1776–1787, 1979.

113. Rauramaa, R.: Relationship of physical activity, glucose tolerance, and weight management. *Prev. Med.* 13:37–46, 1984.

114. Oldridge, N.B., Doitner, A.P., Burk, C.W., et al: Prediction of dropout from cardiac exercise rehabilitation: Ontario Exercise-Heart Collaborative Study. *Am. J. Cardiol.* 51:70–74, 1983.

115. Hammond, H.K., and Froehlicher, V.F.: The physiologic sequellae of chronic dynamic exercise. *Med. Clin. N.A.* 69:21–39, 1985.

116. Shaw, G.S., and Leon, A.S.: Resting metabolic rate (RMR) following prolonged exercise. *Fed. Proc.* 44:1732, 1985.

117. Lipman, R.J., Raskin, P., Love, T., Trebwasser, J., Lecoco, F.R., and Schinke, J.J.: Impairment of peripheral glucose utilization in normal subjects by prolonged bed rest. *J. Lab. Clin. Med.* 76:221–230, 1970.

118. Koplan, J.P., Siscovick, D.S., and Goldbaum, G.M.: The risk of exercise: A public health view of injuries and hazards. *Public Health Reports* 100:189–195, 1985.

119. Koplan, J., Powell, K.E., Silkes, R.K., Shirley, R.W., and Campbell, C.C.: An epidemiologic study of the benefits and risks of running. *JAMA* 248:3118–3121, 1982.

120. Roy, S.: Injuries of exercise. *Med. Clin. N.A.* 69:197–209, 1985.

121. Kraus, J.F., and Conroy, C.: Mortality and morbidity from injuries in sports and recreation. *Ann. Rev. Public Health* 5:163–192, 1985.

122. Lown, B., and Lampert, S.: Sudden cardiac death. In Connor, W.E. and Bristow, J.D. (eds), *Coronary Heart Disease. Prevention, Complication and Treatment.* J. B. Lippincott, Philadelphia 1985, pp. 351–374.

123. Thompson, P.D., Funk, E.J., Carleton, R.A., and Sturmm, W.Q.: Incidence of deaths from jogging in Rhode Island from 1975 through 1980. *JAMA* 247:2535–2570, 1982.

124. Virmani, R., Rabinowitz, M., and McAllister, H.A.: Nontraumatic death in joggers. *Am. J. Med.* 72:874–882, 1982.

125. Waller, B.F., and Roberts, W.C.: Sudden death while running in conditioned runners aged 40 years or over. *Am. J. Cardiol.* 451:1292–1300, 1980.

126. Maron, B.J., Roberts, W.C. and McAllister, H.A., et al: Sudden death in young athletes. *Circulation.* 62:218–229, 1980.

127. Opie, L.H.: Sudden death and sports. *Lancet.* 1:263–266, 1975.

Chapter **23**

Measurement and Evaluation of Health Behaviors in Relationship to Physical Fitness and Physical Activity Patterns

Steven N. Blair, P.E.D.
Harold W. Kohl, M.S.
Institute for Aerobics Research
Dallas, Texas

Introduction

The relationship of exercise and physical fitness to other aspects of a healthy lifestyle is an important issue, but one which has received superficial attention and study. A better understanding of the interrelationships among health behaviors is important for at least two reasons. First, many of the current major health problems in the United States are chronic diseases, and lifestyle factors are major contributors to most of these problems. More than 75% of all mortality in the United States is attributed to only five groups of diseases (1). An analysis of morbidity and disability data supports the same conclusion: that most ill health is caused by relatively few disease conditions (1–4). The link between these disease conditions and lifestyle is supported by analyses that show that 40–60% of deaths from these problems are directly attributed to factors such as smoking, diet, sedentary living, and lack of safety precautions (5–7). The importance of lifestyle in disease etiology and progression requires a better understanding of the interrelationships among health habits. For example, the possible association between sedentary living and risk of developing hypertension may be partly due to the association of inactivity and fatness. It is also possible that individuals who engage in vigorous exercise may be less likely to smoke, may eat a different diet, or may be more effective in managing their stress. Exercisers may be more health conscious in general, and may make more effective use of the medical care system. Another possibility is that related behaviors are linked by a third variable such as educational achievement. Second, by understanding the interrelationships among health behaviors, health promotion planning and health education activities can become more effective. Kottke et al. have shown that mass intervention offers greater hope for a reduction in disease burden than special, high-risk group intervention does (8). The magnitude of this task requires not only an efficient approach to intervention but also an understanding of how changes in one health

---

Note: We thank Mike Smith for data processing support, Kathryn Albrecht for typing the manuscript, Laura Montgomery for providing the National Survey of Personal Health Practices and Consequences data tape, and technicians at the Cooper Clinic and the Institute for Aerobics Research for collecting the data for the ACLS and PALS studies.

behavior affect another. If it is true, for example, that people who start exercising stop smoking, vigorous promotion of exercise may aid antismoking programs. If starting a regular exercise program is likely to lead to fat loss, less emphasis can be placed on caloric restriction. Conversely, if heavy smokers are less likely to start to exercise, a physical fitness program should target smokers for special efforts.

The issues discussed above have important direct implications for the Public Health Service and the Department of Health and Human Services. A complete evaluation of progress toward the 1990 objectives requires both the monitoring of individual health behaviors, as well as the systematic examination of their interrelationships. Obviously, a thorough study of the various important behaviors is also important to help interpret national morbidity and mortality data.

The purpose of this report is to examine the association of exercise or physical fitness with other health behaviors. We consider exercise to be a subset of physical activity, and in this report we consider the two terms interchangeable. Although physical fitness is a status and not a behavior, it is a status achieved through the behavior of regular exercise. Physical fitness is used in this report as a reasonable marker for exercise participation. That is, groups with high levels of physical fitness are assumed to be generally more active than groups with low levels of physical fitness. In addition to a literature review, we also performed several new analyses on existing data sets. The data sets in the analyses include the National Survey of Personal Health Practices and Consequences (NSPHPC) from the National Center for Health Statistics, the Aerobics Center Longitudinal Study (ACLS) from the Institute for Aerobics Research (IAR), and data from worksite health promotion programs conducted by the IAR.

Review of literature

Refereed journals

We reviewed papers in the scientific literature pertaining to the relationships between exercise or physical fitness and smoking patterns, alcohol intake, substance abuse, diet, weight control, stress management, risk-taking behavior, and preventive health examinations. The original list of papers for review was developed from reference lists in known articles and a MEDLINE search. Many of the papers reviewed here are somewhat indirectly related to our topic. Few of the papers were written with the data analysis specifically performed for the issue under study. Most of the reports were written for some other primary purpose and incidentally included some data on interrelationships of health behaviors. Several dozen papers were initially identified. More than 40 were selected for a thorough review based on the article title, contents of the reference, or contents of the abstract. Papers from the initial list were not included for review if the definition of exercise was unclear or uncertain, if the definition or description of other health behaviors was unclear, or if there was an incomplete description of the characteristics of the study group. In addition to the problem mentioned above that many of the articles were not specifically focused on the topic of exercise and other health behaviors, several other problems were encountered in the literature review. First, many of the reports dealt with small convenience samples, so selection bias and the generalizability of results must be questioned. A major problem encountered was that assessments of exercise patterns were usually crude and imprecise, potentially leading to substantial misclassification. Also, many other health behaviors, such as diet and stress management, are equally difficult to assess and dual misclassifications exist. This lack of precision in measuring both exercise and other health behaviors may obscure relationships between the variables.

Results of the literature review are grouped by the association of exercise or physical fitness with other health behaviors.

Weight control

Weight is a physical attribute and not a behavior. However, weight control is determined by the behaviors of diet and exercise patterns. We discuss weight control here separately from diet because overweight is a significant health problem for Americans. This section will focus primarily on the impact of exercise behavior and physical fitness on weight control, but it is certainly widely recognized that dietary behavior is also an important component of the weight control equation.

The evidence is strong that physical activity level is an important determinant of body weight. More active individuals weigh less than their sedentary peers, and this relationship is supported by numerous epidemiologic and clinical studies. The role of inactivity in the etiology of obesity is not well understood. The specific contributions and mechanisms whereby activity relates to weight control need further description. Several cross-sectional epidemiologic studies clearly indicate an inverse association between habitual physical activity and body weight (9–12). This association has been demonstrated in both men and women across the age range.

The cross-sectional results are confirmed by prospective and intervention studies. Literally dozens of exercise training studies have shown an association between exercise participation and weight loss. Four recent reviews on physical activity and weight have been published (13–16). A metaanalysis with careful selection of articles reviewed 16 studies meeting stringent inclusion criteria (14). Conclusions from this metaanalysis are that overweight individuals are more likely to be underexercised than overfed; exercise produces a reliable and measurable weight reduction; little is known about what kind of individual benefits most from exercise programs; and the weight loss produced by exercise tends to be gradual. Other important physiologic benefits of exercise in a weight control program are preservation of lean body mass and the acute effect of exercise on appetite suppression (13).

Diet

There are very few good studies on the relationship between physical activity and dietary intake. The laws of thermodynamics dictate that, in the long run, individuals with higher levels of physical activity must have higher levels of caloric intake. Several studies do support this theoretical position. Cross-sectional studies have shown a positive association between estimates of physical activity and caloric intake, although the correlations are usually low (17–20). This association becomes more striking when highly active individuals are examined. For example, we found that men and women runners averaging 40 and 35 miles per week, respectively, reported caloric intakes approximately 600 kilocalories a day higher than community controls (17). Competitive athletes on university teams also have very high caloric intakes (21).

Detecting an increase in caloric intake in individuals who initiate an exercise program is difficult. One training study found no difference in change in caloric intake between exercisers and controls over a 1-year period (22). However, when the exercise group alone was examined, there was a significant correlation ($r = 0.45$) between estimated caloric intake and miles run per week and between estimates of energy expenditure and caloric intake. When a subset of the more active men from this study were followed for 2 years, significant increases in total caloric intake (367 kcals/day) were seen (23). Energy balance studies from metabolic wards show conflicting results with some studies finding a positive association between energy intake and output (24, 25) and others (26, 27) finding no association. These metabolic ward studies are methodologically superior to the epidemiologic and clinical studies reviewed above, but the results may not be generalizable because of small and highly selected groups studied.

Another confounding factor is that individuals differ in metabolic responses to diet and exercise and in the way in which fat is metabolized and stored (28). Thus, any changes (or lack of changes) in energy intake seen with increased exercise in these studies may be the result of an interaction between individual characteristics and metabolic responses. Smoking status may also influence metabolic responses (29).

Uncertainties arise because of conflicting results of the studies reviewed above. Nonetheless, thermodynamic theory and logical analysis lead to the expectation that more active individuals, over the long haul, need more calories than inactive ones. Although this relationship may be temporarily perturbed by weight gain or loss, a positive association between energy intake and expenditure is likely.

Although we conclude that exercise and caloric intake are associated, the nature of the correlation between physical activity pattern and nutrients in the diet is unclear. Several studies show no difference in nutrient intake across activity strata, at least after adjustment for total caloric intake has occurred. There is some evidence that certain eating practices, such as having a nutritious breakfast, are associated with physical activity patterns (30, 31). These studies have employed very crude dietary assessment methodology and the results should be questioned. The association between physical fitness categories and the frequency of intake of selected foods was significant in one study (32) but not in another (33).

Smoking

It is entirely logical to expect to find a strong association between exercise patterns and smoking status, because these two behaviors appear to be incompatible. Unfortunately, evidence from the studies reviewed was not so clear cut. Several studies report no association between exercise behaviors and smoking behavior (34–37). Several other cross-sectional studies have found negative but weak associations between leisure time physical activity and smoking (30, 38–40). In one study, young middle-aged men who did not smoke averaged 30–35 kilocalories per day more of high-intensity physical activity than their smoking peers (10).

Although the associations between leisure time physical activity and smoking are inconsistent, occupational physical activity and smoking status are positively associated in European studies (30, 38, 39). It seems likely, however, that this association may be confounded by socioeconomic status. Individuals from lower socioeconomic strata tend to have higher levels of physical activity on the job, and because education has been inversely associated with smoking, these individuals are likely to smoke more. The association between physical fitness and smoking status has also been examined in men and women (12, 32, 41). These studies had cross-sectional designs, thus making the direction of the association difficult to determine (that is, Does physical fitness influence smoking or does smoking influence physical fitness?).

Surprisingly, two recent prospective studies also failed to confirm an association between vigorous exercise habits and smoking status. Smokers in the exercise group in a randomized trial were no more likely than controls to stop smoking during a 1-year study (23), and increases in physical fitness in another study were not associated with smoking cessation (42).

Alcohol and substance abuse

Associations between alcohol abuse and exercise or physical fitness are inconsistent. Higher levels of physical fitness are associated with lower alcohol intake in women (41), but not in men (32). In cross-sectional surveys, physical activity is positively associated with alcohol intake in some studies (10, 17) but not in others (18). Change in physical fitness in a prospective survey (33) and change in exercise habits in an experimental study (22) were not associated with changes in alcohol intake. The findings listed above are not the result of age confounding, because age was accounted for in all analyses. Sinyor et al. report some success in treating alcoholism with exercise (43). Approximately twice as many exercisers as controls were abstinent at 3 months of follow-up. We have found no studies on controlled substance abuse and physical activity or physical fitness. A study of high school and college students found associations among alcohol use, smoking, and controlled substance use (44). Exercise and other health behaviors were not associated with this cluster.

Psychological health

There has been a lot of recent enthusiasm for using exercise as a stress management technique. Many people subjectively report that exercise makes them feel better when they are under stress. Whether this association is perceived or real, confirmatory data are lacking. We have been unable to find any evidence that exercise behavior is associated with the behavior of stress management techniques such as progressive relaxation.

A somewhat tangential issue is the association between exercise and various aspects of mental health. This topic has recently been extensively reviewed by Taylor et al. (45). Physical activity probably may alleviate some symptoms of mild and moderate depression (46). Physical activity may also improve self-image, reduce the symptoms of anxiety, and ameliorate some physiologic responses to stressors (47, 48). There is some evidence that regular exercise may affect coronary prone behavior (49). Taylor et al. also point out that some undesirable psychological impact may result from exercise (45).

Other health behaviors

It seems reasonable to assume that individuals who make the effort to engage in regular physical activity may also take other action to protect or improve their health. The evidence on this point is sparse and somewhat weak, but there is support for the hypothesis. Langlie found that active individuals were more likely to obtain dental check-ups and to maintain their immunizations (34). Another survey found that active individuals were more likely to obtain medical and dental check-ups and to have a TB test (50). Interpretation of the data is difficult, but adequate sleep has also been associated with physical activity patterns (50, 51). Physical fitness level was not associated with sleep behavior in a study conducted in a sleep laboratory (52).

Participation in regular physical activity may be associated with seatbelt use and good driving habits (34, 50, 53). These findings from several surveys may be questioned, however, because demographic variables, such as age and education, may likely confound the results. Higher levels of education are associated with physical activity and are also associated with such variables as seatbelt use.

Interrelationships among health behaviors

Heretofore we have been examining the association between physical activity or physical fitness and various health behaviors. The issue can also be viewed in a more general sense. That is, is there a general orientation to health promotion, protection, or prevention; and can we expect to find a clustering of health behaviors? A clustering of cardiovascular risk factors does occur (54), but this observation could be the result of demographic or environmental factors rather than a clustering of health behaviors. Clustering of health behaviors was examined in three studies (30, 50, 55). From three to five factors emerged from the analysis in these three studies. The results suggest that health behavior tends to be multidimensional. Most commonly, however, exercise variables loaded on a single factor just as dietary components did; although in each of the reports, at least one of the dietary measures also loaded on the exercise factor. Interpretation of these studies is made difficult by differences in assessment methodology. The representativeness of the study populations may also be questioned (30). We conclude from reviewing these papers that there is relatively little overlap between physical activity and other health behaviors. This finding is confirmed by a recent report by Stephens based on a large, representative national sample and appropriate sex and age controls in the analysis (56). Clustering of health behaviors was also studied in a large random sample (n = 2,562) representative of the Dutch population (57). The authors concluded that systematic clustering of behavioral risk factors did not occur. Individuals with multiple unhealthy behaviors were more likely to be men with low educational and occupational achievement. Reports from the Alameda study support the conclusion that several key health behaviors operate independently in regard to risk of all-cause mortality (58).

Summary of literature review

This review is based on approximately 45 studies found in the scientific literature. The findings are summarized in table 1. In this table, we have classified the studies by design (note that few cohort or intervention studies are included). The type of associations observed in the studies are given along with the references. Several points were developed from this review.

- Even when associations between physical activity and other health behaviors are found, the correlations are low.
- Regular participation in physical activity is positively associated with better weight control.
- There is little evidence that dietary composition is associated with physical activity patterns, although high levels of activity are associated with high caloric intake.
- The possible association between smoking status and physical activity is confounded by the source of physical activity. Occupation-related physical activity is positively associated with smoking habit, although this is in turn likely to be confounded by socioeconomic status. Associations between leisure time phys-

ical activity and exercise and smoking pattern are inconsistent.
- There is some evidence that physically active individuals may be more likely to engage in certain preventive health behaviors.
- Studies should be interpreted cautiously because control for potential confounders such as age, sex, and education was sometimes lacking.

Although several studies related to the topic were found, the scientific data base is weak. Many of the studies were not designed or written specifically to examine the association between exercise or fitness and other health behaviors. The small and/or statistically insignificant correlations reported for exercise and other health behaviors may be disappointing to some, but several explanations are possible. The findings may be true, and exercise may be generally unassociated with other health behaviors. On the other hand, the lack of strong associations may be more the result of weaknesses in methodology. Misclassification, of both exercise behavior and other behaviors, may be because of our crude and imprecise measuring instruments. It is likely that these misclassifications would act to diminish the strength of any associations that might exist. It should also be noted that some of the observed associations may not be causal, and alternative explanations for the relationships are plausible. It may be that we expect too much in looking for associations among health behaviors. Human behavior is exceedingly complex, and the determinants of health behavior change are not well understood. Because it is likely that multiple unknown factors affect an individual's decisions to practice or not to practice a particular health behavior, perhaps we should not expect exercise to be a strong correlate of other behaviors. Exercise also has a recreational or "fun" component that is presumably not a motivator for most of the other behaviors. It may be this fun component and not potential health benefits that motivates people to participate in regular physical activity.

National surveys

In addition to the peer-reviewed scientific literature, three recent national surveys on exercise are relevant to this paper. These are all cross-sectional surveys that used scientific sampling protocols to ensure a representative sample. Thus, the strength of these surveys is that the results may be generalized to the U.S. adult population, unlike the studies previously reviewed. Weaknesses are that the classifications of exercise status are crude and imprecise and other health behaviors were also measured in a somewhat cursory fashion. In addition, data analyses have not been very extensive, and they are typically limited to cross-tabulations, in many cases without controls for age, education, or other potential confounders.

Perrier Study

The Perrier Study was conducted for the Perrier company by Louis Harris and Associates in 1978 (35). Data were collected by personal interview with 1,510 men and women 18 years and over. Physical activity classification was based on reported regular participation in any one of several listed sports. This crude classification yielded 59% classified as active, and 41% as inactive. In a more detailed classification, actives were further categorized as high, moderate, and low based on estimated weekly caloric expenditure in sports activity.

Table 1. Summary of reported associations between physical activity and other health behaviors by study design

Health behavior	Type of associations in cross-sectional studies	Type of associations in cohort or intervention studies
Weight control	Positive—(9, 10) leisure time activity Positive—(11, 12) fitness test	Positive—(13–16) review articles
Caloric intake	Positive—(17–20) leisure time activity Positive—(18–20) occupational activity None detected—(66) leisure time activity	Positive—(23, 25) leisure time activity Positive—(25) occupational activity None detected—(22) leisure time activity None detected—(26, 27) other measure of activity
Smoking	Positive—(38, 39) occupational activity None detected—(34–37) leisure time activity None detected—(36) occupational activity Negative—(10, 30, 38–40) leisure time activity Negative—(12, 32, 41) fitness test	None detected—(22) leisure time activity None detected—(42) fitness test
Alcohol consumption	Positive—(10, 17) leisure time activity None detected—(18) leisure time activity None detected—(18) occupational activity Negative—(41) fitness test	None detected—(23) leisure time activity None detected—(34) fitness test
Alcoholism		Negative—(43) leisure time activity
Preventive behavior	Positive—(34, 49) leisure time activity Positive—(49) occupational activity	
Risk-taking behavior	Negative—(34, 49, 51) leisure time activity Negative—(49) occupational activity	

Note: Numbers in parentheses refer to references listed at the end of the paper.
Source: Public Health Reports 100(2):172–180, 1985.

This breakdown was 15% high, 16% moderate, and 28% low.

Several questions on diet were included in the Perrier survey. Differences in frequency between activity groups are small, but actives reported higher frequencies and amounts of use in 10 of the 16 food groups. Nonactives were higher in three categories and two of these categories (candies, cookies, or cake and fried food) might be considered as less healthy choices. The Perrier Study also reported a breakdown of respondents' perceptions of whether athletic participation has affected their diets. This analysis was limited to actives. The more active individuals reported increased intake of most items. The greatest differences are frequently seen in the comparisons of runners versus low actives. The pattern suggests that the more active individuals consume a somewhat more healthful diet. The greater apparent intake in the more active individuals is consistent with other reports (17, 21) previously mentioned.

Smoking status and activity was reported in the Perrier Study (35). The percent of nonsmokers was remarkably consistent across groups, with the exception that the moderately active were slightly more likely to be nonsmokers. These data are somewhat surprising, and no ready explanation for the findings is available. Smoking pattern did vary by type of leisure time activity. Runners, participants in racquet sports, and walkers had lower smoking rates when compared with bowlers, swimmers, basketball players, and those who do calisthenics. It is possible that participants in these activities were more likely to be exercising for their health, thereby are more health conscious and were thus less likely to smoke. Participants in the other sports have participated primarily for social or recreational reasons instead of for health.

There was no difference in the amount of sleep each day between actives and nonactives, but nonactives reported slightly more difficulty in getting to sleep or remaining asleep until morning.

Canada Fitness Survey

The Canadian survey used a stratified random sampling procedure to select 13,440 households representative of the Canadian population (31). A survey response rate of 88% was obtained (11,884 households) that yielded 21,658 individuals over 10 years of age on whom data were collected in 1981. The survey questionnaire asked more detailed questions about participation in physical activity than has been the case in several other large surveys. Questionnaire responses were used to develop a quantitative estimate of energy expenditure and classifications of activity level based on participation frequency and duration. Although more detailed data on participation is available from this survey compared with others, we are concerned that criteria for grouping participants into an activity category were too liberal. The active category contains 56% of the population; the moderate category, 31%; and the sedentary, 11%. If a more rigorous definition of activity were used, it seems unlikely that only 11% of Canadians would be classified as sedentary.

Data on activity status and selected health measures are presented in table 2. In every example, the more active individuals have a more positive status.

American Health Survey

The Gallup Organization conducted this survey for *American Health* magazine in 1984 (59). Data were obtained using the standard Gallup telephone survey, which is a probability sample of all telephone households in the United States (except Hawaii). The survey obtained data on 1,019 individuals 18 years of age or over. With this sample size, survey responses between 50 and 70% have an approximate sampling tolerance ± 4 (95% confidence interval).

Several unusual questions were asked during the survey that are related to social health items. A large group (45%) believed that exercise participation had improved their love life. Although a majority of exercisers believed that exercise had had no impact on their relationship with their spouse or loved one (69%), 20% felt that exercise made the relationship closer. Participants were asked about the association between exercise and social life. The results were about evenly split between "made no difference," and "helped make new friends." Younger individuals were more inclined to believe that exercise helped make new friends.

Exercisers who were employed ($n = 422$) were asked about creativity at work. A majority (53%) believed that exercise had no impact, but a sizable group (43%) thought that they were more creative.

The AHS also asked about exercise and dietary habits. Exercisers were asked if they changed their diets after starting to exercise; nonexercisers were asked if they had made any changes in the past year or two. Exercisers were more likely to report presumably beneficial changes, such as eating more fruits and vegetables, less sugar, and less meat, and in weight loss, and use of low-calorie foods and beverages.

Survey respondents were also asked about stopping smoking. Exercisers were more likely (13%) than nonexercisers (7%) to report smoking cessation. This is one of the only surveys to show the association between changing exercise and smoking habits concurrently. This is encouraging, although the data are based on retrospective self-report and may be suspect.

Table. 2 Physical activity status, by selected health measures: Canada Fitness Survey
[Percent]

Health measures	Active	Moderate	Sedentary
Positive emotional well-being	74	70	55
Self-rating of health, very good	23	17	14
Smoke	37	44	42
Get 7–8 hours of sleep	70	70	63
Have a good breakfast	51	45	43

Source: Canada Fitness Survey. Fitness and Lifestyle in Canada, Fitness Canada, Ottawa, 1983.

Original analyses

Results are grouped by respective data sets. Each analysis could not be performed on all data sets because of differences in the type of data collected. By asking questions relevant of each data set, we maximized the inherent strengths of each group of data. The National Survey of Personal Health Practices and Consequences (NSPHPC) data file was created by the National Center for Health Statistics, and the Aerobics Center Longitudinal Study (ACLS) and the worksite health promotion data sets were created and are maintained by the Division of Computer Services at the Institute for Aerobics Research in Dallas, Texas. All data analyses were performed using the Statistical Package for the Social Sciences (60).

Aerobics Center Longitudinal Study

The ACLS is an ongoing investigation examining the relationships between aerobic exercise and health. Clinical examination data and data from a followup survey in 1982 are the two main components to the ACLS. The study participants are patients receiving preventive medical examinations at the Cooper Clinic in Dallas, Texas. Patients of the clinic are primarily self-referred or are sent by their employers for examinations and advice on positive health behavior. Many patients return at various intervals for repeat examinations, although there is no strict protocol for followup examinations.

Clinical data

Upon visiting the Cooper Clinic, each patient gives informed consent, completes a medical history and lifestyle questionnaire, and has a comprehensive physical examination that includes a maximal exercise stress test and a complete blood profile analysis. Each patient, based on the results of his or her examination, receives an exercise and diet prescription from their clinic physician and nutritionist. These individualized recommendations are based on the patient's physical fitness level and coronary risk profile. After the visit, each patient's medical record is keyed to disk storage. The data base is continually updated in an archive fashion so as to have complete access to all records of the Cooper Clinic population.

Patients included in the present analyses were those who had at least two complete clinic visits between 1971 and June 1984. Only data from the first and last visits are used for analyses. Patients included in the analyses were those who reported no personal history of chronic disease at baseline visit. After exclusion for positive medical history, incomplete records, and outliers, the cohort under study is 3,732 patients.

Exercise classifications in these analyses result from a question asked on each visit to the Cooper Clinic: "Are you currently engaged in a regular exercise program?" Based on patient responses at baseline and followup, the four exercise classifications are: never exercise, stopped exercise, started exercise, and continued exercise.

In addition to exercise inquiries, the medical history and lifestyle questionnaire completed at each visit included other health behavior inquiries, including a battery of dietary habit questions. Each patient is asked to provide the number of servings per week they normally have over a range of food items. These items as well as weight and fitness (treadmill time) are used in the present analyses.

Followup survey

A followup survey of Cooper Clinic patients who received a medical examination between 1970 and 1981 was conducted in 1982. The 12,285 respondents represented response rate of 77%, and the respondents did not differ substantially from nonrespondents on most key variables (61).

The mail survey questionnaire focused on current health habits and health status. An extensive set of exercise habit inquiries provided the core questions of the survey. Questions on health status, health habits, health service visits, and important demographic issues were also included.

Each questionnaire was coded and keyed to disk storage. For the purposes of analyses, these followup data were then linked to the respondent's clinical data (provided at the time of visit to the Cooper Clinic). The median interval from a patient's last Cooper Clinic visit to the time of the survey was 4 years. This aspect of the study provided an opportunity to examine changes in health behaviors over an extended time period; to determine how these changes affect health status, physical fitness, and physical activity; and permitted statistical control for followup interval.

Exercise categorizations in the followup survey are based on a question asking a respondent to provide the total amount of time during the last 7 days that was spent doing vigorous activity (defined as jogging or running, swimming, racquetball, chopping wood, digging in the garden, etc.) and moderate activity (golf, doubles tennis, yard work, heavy housecleaning, etc.). After exclusion of respondents with missing data values, 9,159 observations remained for analysis.

Health behaviors included in these analyses are grouped into several categories: diet, smoking, exercise, stress, preventive health behaviors, substance abuse and risk-taking behavior. Although the questions in each of these categories are not an exhaustive inquiry into each type of behavior, those issues included are relevant to current research.

Cooper Clinic patients

Tables 3 and 4 show the baseline values and changes for dietary, weight, and fitness data by exercise categories for 3,732 Cooper Clinic patients who had at least two clinic examinations. Changes are calculated from first and last clinic visits. The mean interval between visits was nearly 3 years. Exercise categories are based on exercise status at the two clinic visits. For example, the started exercise category contains individuals who

Table 3. Baseline diet, weight, and fitness data, by exercise categories: Cooper Clinic patients

Item	All (*n* = 3,732)	Never exercised (*n* = 378)	Stopped exercise (*n* = 223)	Started exercise (*n* = 780)	Continued exercise (*n* = 2,351)
Weight (lb)	174.5 (29.8)	181.7 (34.1)	176.7 (32.3)	176.0 (31.7)	172.6 (27.8)
Eggs*	3.5 (3.2)	3.7 (3.4)	3.7 (3.3)	3.7 (3.2)	3.5 (3.1)
Beef*	4.8 (2.9)	5.7 (3.2)	4.9 (3.1)	5.5 (3.1)	4.4 (2.7)
Pork*	1.2 (1.5)	1.5 (1.8)	1.2 (1.6)	1.3 (1.5)	1.1 (1.5)
Fish*	1.6 (1.4)	1.2 (1.1)	1.5 (1.4)	1.4 (1.2)	1.8 (1.5)
Fowl*	1.9 (1.4)	1.7 (1.3)	1.8 (1.5)	1.7 (1.4)	2.0 (1.4)
Whole milk*	1.8 (4.1)	2.4 (4.4)	1.9 (3.8)	1.8 (3.8)	1.6 (4.2)
Lowfat milk*	2.8 (5.4)	1.8 (5.5)	3.4 (6.7)	1.6 (3.8)	3.3 (5.5)
Tea*	6.6 (8.7)	6.8 (8.4)	7.2 (8.9)	7.1 (8.9)	6.3 (8.7)
Coffee*	15.1 (14.5)	16.2 (15.4)	14.0 (14.2)	17.1 (14.5)	14.4 (14.3)
Beer*	2.7 (5.4)	3.1 (7.1)	2.3 (5.1)	2.7 (5.4)	2.7 (5.2)
Wine*	2.3 (4.0)	1.6 (2.7)	2.4 (4.8)	2.1 (3.7)	2.4 (4.2)
Liquor*	3.2 (5.4)	4.1 (6.7)	2.9 (4.9)	4.1 (6.2)	2.9 (4.8)
Age (years)	42.8 (9.5)	43.1 (8.9)	41.6 (10.7)	43.0 (9.1)	42.8 (9.6)
Treadmill time (sec).	1,014 (296.8)	789 (227)	970 (293)	848 (234)	1,110 (281)

*Servings per week.
Note: Values in parentheses are means (SD).

were not exercising at the first visit but reported exercising at followup. This exercise status classification is somewhat crude. It is based on patients' responses to a question asking if they are currently engaged in a regular exercise program. This imprecise classification of exercise status at each clinic visit leads to possible misclassifications when patients are assigned to exercise change categories. Moreover, using data from the first and last visits precludes analyses of behaviors undertaken between the two visits. However, these classifications have some validity, as indicated by change in treadmill time in the four exercise change categories (assuming that patients can change their maximal treadmill performance by increasing their exercise participation). Patients in the started exercise category increased their treadmill time by 167 seconds, and those in the never exercised category increased by only 40 seconds.

The data in tables 3 and 4 indicate that patients in the started exercise category lost nearly twice as much weight as the group as a whole. The entire patient group tended to make favorable dietary changes (generally reporting less fat and cholesterol intake) during the followup period. In general, the started exercise group had relatively more favorable changes in diet than the other groups, but the differences were small. Changes in weight, dietary habits, and fitness, by exercise change categories, are shown separately for men and women in table 5.

Because the exercise categorization cannot be viewed as an ordered progression, there are more discrete comparisons to highlight. In terms of relevant comparisons the started exercise group showed more beneficial changes (table 4) than the never exercisers did. The continued exercisers also showed more pronounced

Table 4. Changes in diet, weight, and fitness, by exercise categories: Cooper Clinic patients

Item	All (*n* = 3,732)	Never exercised (*n* = 378)	Stopped exercise (*n* = 223)	Started exercise (*n* = 780)	Continued exercise (*n*= 2,351)
Weight (lb)	−1.3 (9.6)	−0.2 (11.0)	−0.5 (9.4)	−2.4 (11.1)	−1.2 (8.8)
Eggs*	−0.8 (2.7)	−0.9 (2.6)	−0.9 (3.0)	−0.8 (3.0)	−0.9 (2.6)
Beef*	−1.1 (2.4)	−1.3 (2.8)	−1.1 (2.5)	−1.4 (2.8)	−1.0 (2.2)
Pork*	−0.3 (1.4)	−0.2 (1.4)	−0.2 (1.4)	−0.3 (1.4)	−0.2 (1.4)
Fish*	0.4 (1.4)	0.4 (1.2)	0.4 (1.3)	0.6 (1.4)	0.4 (1.4)
Fowl*	0.5 (1.5)	0.3 (1.4)	0.5 (1.7)	0.6 (1.6)	0.5 (1.5)
Whole milk*	−0.7 (3.4)	−1.0 (3.7)	−0.9 (3.8)	−0.6 (3.5)	−0.6 (3.3)
Lowfat milk*	1.1 (5.0)	0.6 (5.2)	−0.1 (6.6)	1.7 (4.6)	1.0 (5.0)
Tea*	−0.7 (7.5)	−0.3 (8.2)	−2.1 (6.8)	−0.4 (6.9)	−0.7 (7.7)
Coffee*	−1.7 (10.9)	−0.3 (10.9)	−2.1 (12.0)	−2.3 (11.1)	−1.6 (10.9)
Beer*	−0.2 (4.1)	−0.7 (4.4)	−0.6 (3.4)	−0.3 (4.3)	−0.1 (4.0)
Wine*	0.1 (3.6)	−0.1 (2.3)	−0.4 (4.3)	0.3 (3.6)	−0.1 (3.7)
Liquor*	−0.7 (3.8)	−1.1 (5.7)	−0.4 (3.5)	−1.0 (4.3)	−0.6 (3.3)
Age (years)	2.8 (2.2)	2.7 (2.2)	2.3 (1.7)	3.7 (2.4)	2.6 (2.1)
Treadmill time (sec)	120 (186.8)	39.7 (180.5)	−3.70 (269.5)	168 (188.9)	117 (169.4)

*Servings per week.
Note: Values in parentheses are mean changes (SD).

Table 5. Change in weight, dietary habits, and fitness in men and women, by exercise categories, first to last visits, and change category: Cooper Clinic patients

	All		Never exercised		Stopped exercise		Started exercise		Continued exercise	
Item	Men (*n* = 3,237)	Women (*n* = 495)	Men (*n* = 317)	Women (*n* = 61)	Men (*n* = 188)	Women (*n* = 35)	Men (*n* = 635)	Women (*n* = 125)	Men (*n* = 2,077)	Women (*n* = 274)
Weight (lb):										
Mean	−1.53	0.02	−0.66	1.98	−0.45	−1.12	−2.99	0.41	−1.30	−0.45
SD	(9.66)	(9.28)	(10.36)	(13.87)	(9.12)	(10.95)	(11.38)	(9.04)	(8.95)	(7.76)
Eggs*:										
Mean	−0.91	−0.39	−0.92	−0.43	−1.02	0.00	−0.84	−0.57	−0.93	−0.36
SD	(2.75)	(2.25)	(2.65)	(2.19)	(3.12)	(2.47)	(3.01)	(3.00)	(2.65)	(1.95)
Beef*:										
Mean	−1.14	−0.95	−1.31	−1.25	−1.22	−0.71	−1.47	−1.23	−1.01	−0.84
SD	(2.47)	(1.91)	(2.84)	(2.66)	(2.55)	(2.57)	(2.96)	(1.67)	(2.23)	(2.96)
Pork*:										
Mean	−0.26	−0,24	−0.27	−0.15	−0.27	0.00	−0.33	−0.23	−0.23	−0.27
SD	(1.40)	(1.20)	(1.44)	(1.37)	(1.53)	(0.59)	(1.39)	(1.10)	(1.39)	(1.24)
Fish*:										
Mean	0.45	0.42	0,42	0.28	0.38	0.35	0.58	0.56	0.41	0.40
SD	(1.40)	(1.52)	(1.20)	(1.10)	(1.31)	(1.27)	(1.40)	(1.51)	(1.43)	(1.59)
Fowl*:										
Mean	0.52	0.55	0.38	0.08	0.35	1.30	0.62	0.58	0.52	0.55
SD	(1.50)	(1.61)	(1.37)	(1.38)	(1.44)	(2.88)	(1.59)	(1.62)	(1.49)	(1.48)
Whole milk*:										
Mean	−0.73	−0.28	−1.07	−0.56	−1.03	−0.44	−0.64	−0.13	−0.68	−0.26
SD	(3.55)	(2.47)	(3.84)	(2.09)	(3.60)	(5.18)	(3.69)	(2.06)	(3.45)	(2.33)
Lowfat milk*:										
Mean	1.08	1.02	0.60	0.73	−0.32	1.00	1.66	1.76	1.05	0.82
SD	(5.11)	(4.45)	(5.41)	(3.13)	(7.14)	(3.00)	(4.61)	(4.94)	(5.07)	(4.54)
Tea*:										
Mean	−0.62	−1.36	−0.35	−0.06	−2.16	−1.78	−0.36	−1.05	−0.64	−1.61
SD	(7.52)	(7.66)	(8.49)	(5.54)	(7.03)	(5.57)	(6.73)	(7.83)	(7.65)	(8.01)
Coffee*:										
Mean	−1.70	−1.62	−0.56	1.53	−1.83	−3.89	−2.06	−4.52	−1.72	−1.00
SD	(11.19)	(9.34)	(11.19)	(8.57)	(12.62)	(7.49)	(11.10)	(10.92)	(11.13)	(8.84)
Beer*:										
Mean	−0.19	−0.10	−0.76	−0.22	−0.68	0.00	−0.35	−0.36	−0.03	0.00
SD	(4.30)	(1.84)	(4.65)	(0.76)	(3.01)	(0.78)	(4.55)	(1.58)	(4.20)	(2.09)
Wine*:										
Mean	0.07	0.12	−0.06	−0.88	−0.32	−0.58	0.33	0.31	0.01	0.26
SD	(3.70)	(2.88)	(2.22)	(2.42)	(4.56)	(2.27)	(3.68)	(3.37)	(3.81)	(2.78)
Liquor*:										
Mean	−0.76	−0.49	−1.25	−0.22	−0.35	−0.94	−1.07	−0.67	−0.62	−0.43
SD	(4.03)	(1.88)	(6.03)	(1.61)	(3.63)	(3.04)	(4.48)	(2.11)	(3.49)	(1.72)
Age (years):										
Mean	2.76	3.19	2.52	3.52	2.31	2.08	3.50	4.57	2.59	2.63
SD	(2.16)	(2.53)	(2.02)	(3.03)	(1.66)	(1.74)	(2.19)	(3.26)	(2.16)	(1.74)
Treadmill time (sec):										
Mean	114.5	105.1	36.97	54.68	−0.42	−23.03	170.87	150.16	117.81	109.73
SD	(184.5)	(201.2)	(181.2)	(177,34)	(246.99)	(381.03)	(191.61)	(171.37)	(167.79)	(181.89)

*Servings per week.

Note: Values in parentheses are means (SD).

Table 6. Associations among changes in exercise and diet variables: Cooper Clinic patients

										Runners and joggers and walkers only (n = 2,052)	
Item	Visit number	Days per week exercise	Eggs	Beef	Pork	Fowl	Fried foods	Whole milk	Coffee	Miles per day	Average time
Treadmill time (sec.)	*0.16	*0.28	*−0.04	*−0.10	*−0.04	*−0.09	*−0.07	−0.01	*−0.04	*0.34	*0.19
Visit number		*0.14	*−0.07	*−0.09	*−0.06	*−0.01	−0.02	*−0.05	−0.02	*0.09	*0.08
Days/week exercise			*0.16	*−0.08	−0.02	*0.08	−0.04	−0.01	0.00	*0.47	*0.43
Eggs				*0.14	*0.27	0.03	*0.07	*0.07	*0.08	−0.01	−0.02
Beef					*0.22	*−0.10	*−0.27	*−0.11	*0.08	−0.08	−0.03
Pork						0.02	*0.12	*0.05	*0.06	−0.05	−0.03
Fowl							*0.06	0.01	*0.03	−0.08	0.04
Fried foods								*0.11	*0.07	−0.04	0.01
Whole milk									0.04	−0.01	−0.01
Coffee										*−0.05	−0.03
Miles/day											*0.49

Notes: All values are based on differences (followup — baseline).
Dietary variables measured in servings per week.
*Significant at $p < 0.05$.
n = 3,732.

changes from baseline to followup than those who stopped exercise did.

The association between exercise or physical fitness changes and dietary variables is also shown in the correlation matrix in table 6. Individuals who had greater increases in either treadmill time or in exercise participation tended to eat fewer servings per week of most foods. Although exercisers eat more total calories (17, 21), we would expect them to be eating more servings of some foods. Perhaps these increases occurred in foods not available for the analysis (grains, fruits, and vegetables). The correlations, even where significant, were quite low, usually less than 0.2.

Smoking habit and exercise status at first and last visits to the Cooper Clinic is shown in table 7. At baseline, smoking prevalence in nonexercisers is double that of exercisers. At followup the difference in smoking prevalence between groups decreases. It is difficult to determine causality from such simple analyses. It can be said, however, that there was a decrease in smoking rates coincident with an increase in exercise participation.

Followup survey

Frequencies of health behavior responses by the amount of time spent doing vigorous activity in the last 7 days are presented in tables 8 and 9. Despite the potential for misclassification based on this question, those who said they participated in vigorous exercise 7 hours or more in the previous week were more likely to be active than other respondents. In fact, the purpose of this breakdown centers on identifying highly active people. In general, the respondents who fell in the higher activity classifications (4–6 and 7 or more hours of vigorous physical activity in the previous week) showed more positive health behaviors than those in the lower two categories (0 hours and 1–3 hours) of vigorous physical activity in the previous week. Even though it is recognized that this inquiry addresses acute and not chronic exercise exposure, the patterns seen in table 8 do lend credence to the validity of the above generalizations.

Table 7. Smoking and exercise habits at first and last clinic visits: Cooper Clinic patients
[Percent]

	Baseline		Followup	
Smoking habit	Do not exercise (n = 1,158)	Exercise (n = 2,674)	Do not exercise (n = 601)	Exercise (n = 2,651)
Do not smoke	71.2	85.1	78.2	84.7
Smoke	28.8	14.9	21.8	15.3

An effort was made to further analyze this group based on exercise participation in the last 7 days. The data shown in table 9 represent a comparison of health behaviors in those highly active people and those who reported neither moderate nor vigorous exercise in the previous week. Thus, this analysis examines the extreme groups in the exercise distribution. Differences are seen between the two groups in nearly every category. In general, the vigorous exercisers show more positive health behaviors than sedentary respondents do.

National Survey of Personal Health Practices and Consequences

The National Survey of Personal Health Practices and Consequences (NSPHPC) was a telephone survey of U.S. adults. The NSPHPC consisted of two waves of data collection, spaced 1 year apart (62–65). Wave I was conducted in spring 1979, and respondents to Wave I were reinterviewed for Wave II. The target population of this survey was all noninstitutionalized civilian adults ages 20–64 years in the United States residing in telephone households in the 48 contiguous States (66).

There were 2,436 Wave II respondents out of the possible 3,025 persons who responded to the Wave I questionnaire (62, 63). These data correspond to a Wave II response rate of 80.5%. Although the participants in both waves of the NSPHPC have been shown to differ in some respects (most notably sex composition) from the ongoing National Health Interview Survey (NHIS), it has been concluded that the data set is reasonably representative of the target population; and no major analytic difficulties can be expected (62, 63).

Study design and data collection

The NSPHPC consisted of a stratified multistage cluster sampling design. The three-stage design allowed equiprobable case selection in the first two stages and a probability proportional to size in the final stage of selection. The data in both waves of the survey were collected via a 20-minute telephone interview. No proxy information was collected. If it was determined that the targeted respondent was not at home at the time of the interview, another interview was scheduled for a later date. The NSPHPC employed extensive followup efforts in order to maintain the integrity of the original sampling plan. Analyses in this paper are limited to those respondents who reported that their activity was not limited in any way (that is, they were not disabled).

Only those who responded to both waves of the NSPHPC were included for analysis (*n* = 2,428). Exercise categorizations were adopted from responses to a battery of activity participation questions. Subjects were classified as highly active if they reported jogging or running sometimes or often, averaging more than 15 miles per week. Respondents were classified as active if they reported jogging or running sometimes or often, averaging 5–15 miles per week. Categorization into the somewhat active exercise group resulted from respondents reporting that they participated sometimes or often in one of the following activities: swimming in the summer, long walking, physically active hobby such as dancing or gardening, bicycling, calisthenics or physical exercise or other active sports. Subjects who reported running or jogging sometimes or often, but who aver-

Table 8. Health behaviors and measures by hours spent in the last week doing vigorous exercise: Aerobics Center Longitudinal Study
[Percent]

Item	0 hours (*n* = 1,819)	1–3 hours (*n* = 3,995)	4–6 hours (*n* = 2,081)	7+ hours (*n* = 1,265)
Diet				
One or fewer cups of coffee/day	35.9	79.4	85.4	32.8
One or fewer cups of tea/day	56.7	59.3	64.1	57.9
Drink alcohol	72.7	78.4	77.7	75.8
Restrict amount of salt in diet	42.7	45.0	51.2	54.9
Smoking				
Never smoked on a daily basis	44.5	51.9	50.4	54.6
Current smoker	17.9	9.0	8.3	6.9
Exercise				
Walk, run, or jog as part of a physical activity program	35.9	79.4	85.4	82.0
Watch television 10 or fewer hours/week	63.1	71.8	72.8	71.1
Stress				
High level of occupational stress	36.3	35.7	33.6	33.7
Difficulty sleeping several times/week	11.5	7.4	6.4	7.2
Health status measures				
Zero days of work missed last year due to illness	51.3	51.6	56.3	58.5
Zero doctor visits last year	49.1	50.9	53.7	54.7
Less than one year since last dentist visit	84.4	90.2	90.6	89.0
Substance abuse				
Alcohol intake pattern spread over entire week	67.0	68.3	71.6	71.5
Risk-taking behavior				
Wear seatbelts 76–100% of time	20.1	25.7	26.9	22.5

aged less than 5 miles per week, were also categorized in the somewhat active group. Respondents who reported rarely participating in any of the aforementioned activities were classified as inactive. Those who reported no participation in any of the activities in question were classified as sedentary.

Health-related behaviors were analyzed by these exercise categories. Health behaviors examined were stress management, preventive health behaviors, risk-taking behavior, weight, diet, and smoking. This set of behaviors was analyzed for both waves of the NSPHPC. We believe that the inquiries by the NSPHPC concerning health-related behavior, although not exhaustive, provide one of the best and most complete data sets on this topic.

Exploratory multivariate analyses were also performed on the NSPHPC data. Stepwise discriminant function analyses were performed in an attempt to discriminate between exercise classifications based on a set of predictor variables. Use of the stepwise method allows for predictor variables to enter and be dropped from a model based on their relative contributions given the other variables in the model. Data on diet, smoking, weight, stress, and other health behaviors for Wave I by exercise category are presented in tables 10 and 11. Wave II results (not shown) are highly congruent with those from Wave I, so only Wave I results are given in the table. The major objective of the exercise index (described above) was to create categories of truly active and truly inactive individuals. Intuitively, the distribution of individuals across categories suggests that objective was achieved. The highly active and active categories constitute approximately 10% of the sample. This is slightly less than the highly active (15%) category in the Perrier Study (35). It does seem reasonable to assume that the 10% in the top two exercise categories are likely to be active to the point of achieving health benefits. Unfortunately, the middle category encompasses the rest of the sample and is thus much larger. We have not developed a logical coding scheme to further subdivide the middle category.

In general, the highly active and active individuals have better health habits than individuals in the other categories. They are less likely to be overweight and smoke; and they are more likely to limit red meat, visit a dentist, floss their teeth, and wear seatbelts. Areas where the more active individuals have poorer habits include drinking five drinks or more at a sitting and getting medical checkups. The considerably lower smoking prevalence in the highly active and active individuals is at variance with most previous reports, but it is more in line with our expectations. This may be because the activity categories contain truly vigorously active individuals and therefore less misclassification than in earlier reports. The reader is reminded that all NSPHPC analyses excluded those who were reportedly disabled.

One weakness in current health behavior research is the lack of multivariate analyses. The interrelationship of health behaviors is an exceedingly complicated multidimensional problem, and complex statistical techniques need to be applied to sort out independent effects. Table 12 presents results from a stepwise discriminant function analysis. The object of this analysis was to identify a combination of variables that provide the best discriminatory power between groups of the dependent variable. The coefficients in table 12 can be interpreted in the same manner as beta coefficients in multiple regression.

Each coefficient represents the relative contribution of its corresponding variable to the discriminating power of the function. The strongest relative contributions came from body mass index, age, the number of glasses of fruit or vegetable juices consumed per day, and weight. Essentially these are the relative effects of the variables after controlling for all other variables in the model.

Table 9. Health behaviors: sedentary versus vigorous exercisers, ACLS followup survey
[Percent]

Item	No reported moderate or vigorous activity in last 7 days (*n* = 1,106)	7 or more hours of vigorous activity in last 7 days (*n* = 1,265)
Diet		
One or fewer cups of coffee/day	29.2	32.8
One or fewer cups of tea/day	55.5	57.9
Drink alcohol	70.5	75.8
Restrict amount of salt in diet	41.6	54.9
Smoking		
Never smoked on a daily basis	44.2	54.6
Current smoker	17.9	6.9
Exercise		
Walk, run or jog as part of a physical activity program	30.1	82.0
Watch television 10 or fewer hours per week	62.1	71.1
Stress		
High level of occupational stress	37.8	36.7
Difficulty sleeping several times per week	11.0	7.2
Preventive health behaviors		
0 days of work missed because of illness last year	51.3	58.5
0 doctor visits last year	48.9	54.7
Less than 1 year since last dentist visit	81.8	89.0
Substance abuse		
Alcohol intake pattern spread over entire week	65.2	71.5
Risk-taking behavior		
Wear seatbelts 76–100% of the time	19.9	22.5

Table 10. Stress and health behaviors by exercise categories: National Survey of Personal Health Practices and Consequences, Wave I
[Percent]

Item	Highly active (*n* = 54)	Active (*n* = 187)	Somewhat active (*n* = 2,066)	Inactive (*n* = 70)	Sedentary (*n* = 57)
Stress					
A great deal of stress in job	22.6	25.1	16.1	18.6	5.3
Preventive health behaviors					
Less than 1 year since last dental visit	75.9	69.9	61.4	44.3	45.5
Floss teeth 3–7 times/week	37.0	39.6	29.7	21.4	22.8
Less than 1 year since last eye exam	38.5	45.7	36.6	41.4	35.7
Less than 1 year since last general physical exam	48.1	51.6	44.6	41.8	50.9
Less than 1 year since last blood pressure check	75.5	83.3	79.9	80.0	84.2
Less than 1 year since last PAP test (women only)	22.2	33.2	37.6	36.2	37.0
Less than 1 year since last breast examination (women only)	22.2	34.9	40.2	36.8	38.2
Alameda 6 habit health score:					
4 or more health habits	70.3	76.4	73.4	28.6	52.6
No emotional problems					
Emotion problem score	44.4	50.3	51.5	51.4	54.4
Risk-taking behavior					
Always or nearly always wear seatbelts	25.9	22.5	19.8	10.0	15.8

Source: National Center for Health Statistics; analyzed by the Institute for Aerobics Research.

The data in table 13 amplify the effects of education and age on exercise categorization. The highly active people tended to be more educated and younger, and the inverse was true for the sedentary people, thus raising questions about the associations in tables 10–12. Unfortunately, we have not been able to extend these preliminary analyses by controlling for these latter and other potentially confounding variables. Of the highly active, 42% were college graduates, and only 1.8% of the sedentary group had comparable education. Younger people made up over 66% of the highly active group and only 15.8% of the sedentary group.

Personalized Aerobics Lifestyle System

Two worksite health promotion projects were implemented in the Dallas area between 1982 and 1984 by the Institute for Aerobics Research in a suburban school district (Group I) and for city employees (Group II). Group I consisted of 647 employees (men = 104, women = 543), for whom the average age was 40.4 years (SD = 9.3). Group II (*n* = 321) was younger (average age = 36.8 years; SD = 9.4) and more evenly split between men (*n* = 199) and women (*n* = 122). The programs were presented using the Personalized Aero-

Table 11. Diet, smoking status, and weight by exercise categories: National Survey of Personal Health Practices and Consequences, Wave I
[Percent]

Item	Highly active (*n* = 54)	Active (*n* = 187)	Somewhat active (*n* = 2,066)	Inactive (*n* = 70)	Sedentary (*n* = 57)
Weight					
Perceive themselves as overweight	25.9	36.4	45.4	59.4	43.6
Over 110% of desirable weight	28.7	28.7	42.7	67.1	52.7
Diet					
Limit red meat in diet	42.6	33.7	25.0	24.3	22.8
Take vitamins:					
Regularly	33.3	46.0	34.7	25.3	30.4
Never or rarely	51.9	35.8	48.3	61.4	60.7
Do not drink alcohol	22.2	16.2	28.3	44.3	58.2
5 or more drinks at a sitting	20.4	10.2	6.2	5.7	5.6
Smoking					
Never smoked	46.3	49.2	44.7	44.3	64.9
Current smoker	24.1	28.3	36.4	42.9	28.1
Made serious attempt to quit	53.8	58.5	48.6	30.0	40.0

Table 12. Standardized canonical discriminant function coefficients, discriminating between exercise classifications: National Survey of Personal Health Practices and Consequences, Wave I

Predictor variables	Coefficient
Daily cups of coffee	0.001
Age	−0.488
Daily glasses of milk	0.032
Daily glasses of water	0.146
Education	0.290
Daily glasses of fruit or vegetable juices	0.396
Weight	0.412
Alameda 6 habit health score	0.272
Body mass index	−0.514

Note: $n = 2{,}323$.

Source: National Center for Health Statistics; analyzed by the Institute for Aerobics Research.

bics Lifestyle System (PALS). PALS is a comprehensive health promotion program, guiding participants through a 10-week to 12-week health intervention. The process begins with baseline medical screening that includes fitness assessment, submaximal exercise stress testing, and a blood chemistry profile, as well as administration of a health and lifestyle questionnaire. A 10-week comprehensive series of educational and counseling sessions follows that emphasize exercise, stress management, and nutrition. Supervised exercise classes are offered throughout the program. Each participant receives an extensive summary of baseline tests, including a coronary risk profile. This summary provides exercise and nutrition prescriptions as well as references that can be utilized throughout the program. At followup, the testing is repeated to determine effects of the program on physical fitness, health behavior, and clinical status. The positive effects of the PALS intervention have been documented (67).

Health screen data including treadmill time and blood profile were collected by qualified technicians. Health behavior data, assessed by questionnaire, were obtained concurrently. These data were then linked so that a complete health profile for each subject was obtained for both baseline and followup.

Project participants were classified into exercise categories designed to differentiate highly active people from the rest of the subjects. At both baseline and followup in both populations, vigorous exercisers were those who reported participation in one of the following activities: regularly (for the previous 3 months) running or jogging 10 miles or more per week, playing strenuous racquet sports more than 11 hours per week, riding a bicycle 26 miles or more per week, or swimming more than 4 miles per week. Subjects not classified as vigorous exercisers were subsequently lumped into the not vigorous exerciser group for comparative purposes. As in previous analyses, responses to a set of health behavior questions were analyzed by exercise classifications. Health behaviors used for the PALS analyses include: smoking, stress, preventive health behaviors, and substance abuse.

In both of the PALS data sets, the same general trends are seen as in previous analyses. Those participants classified as vigorous exercisers tended to have more favorable health behaviors than those in the not vigorous exerciser group (tables 14 and 15). These differences are seen in both data sets, at baseline and at followup. The changes from baseline to followup are striking. A general improvement in each health-related behavior is seen across all categories. Although it is difficult to prove that changes in exercise status from baseline to followup contribute to changes in health behavior, it is possible to say that favorable changes in health behaviors occurred concurrently with an increase in exercise participation.

There are between-group as well as within-group comparisons that need to be addressed. Between-group comparisons are not entirely valid because of the different compositions of the two populations. The school district employees were mostly women (83.9%) and relatively young (mean age = 40.4 years). Only 4.2% of the school district employees had less than a high school education. Conversely, the city employees were 38.3% women, and the average age was 36.8 years. Almost 21% of the city employees had only a high school education. Vigorous exercisers smoked less, used seatbelts more often, visited the dentist more often, and had a pap smear

Table 13. Education level and age by exercise classifications: National Survey of Personal Health Practices and Consequences, Wave I
[Percent]

Item	Highly active (n = 53)	Active (n = 187)	Somewhat active (n = 2,063)	Somewhat inactive (n = 70)	Sedentary (n = 57)
Education					
Less than high school	13.0	8.0	20.2	36.8	38.6
High school graduate or equivalent	14.8	31.6	39.5	27.9	45.6
Some college	29.6	27.3	22.3	23.5	14.0
College graduate or more	42.6	33.2	18.0	11.8	1.8
Age					
20–34 years	66.7	65.8	40.5	24.3	15.8
35–49 years	29.6	27.8	30.9	38.6	24.6
50–64 years	3.7	6.4	28.6	37.1	59.6

Source: National Center for Health Statistics; analyzed by the Institute for Aerobics Research.

Table 14. Selected health behaviors, by exercise category: Personalized Aerobics Lifestyle System, school district employees
[Percent]

Item	Baseline		Followup	
	Vigorous exerciser (*n* = 67)	Not vigorous exerciser (*n* = 580)	Vigorous exerciser (*n* = 92)	Not vigorous exerciser (*n* = 555)
Smoking				
Never smoked cigarettes	64.2	63.3	64.1	67.6
Current smoker	15.6	34.3	12.8	25.2
Stress				
High occupational stress level	34.3	43.8	31.5	36.7
Difficulty sleeping several times per week	9.0	9.0	4.3	2.8
Risk-taking behavior				
Use seatbelts 75–100% of time	25.4	21.1	38.0	33.1
Preventive Health Behaviors				
Had a general check up in last year	58.2	60.1	73.9	67.2
Visit dentist at least once/year	85.1	85.8	90.2	89.2
Pap smear at least once/year (women only)	69.8	74.1	80.3	76.4
0 days of work missed last year because of illness	34.3	24.2	—	—
Substance abuse				
Alcohol intake pattern spread over 1 week	34.0	34.5	32.8	32.1
Often or sometimes use drugs to help relax	10.5	8.3	5.4	5.8

(women only) more often than the not vigorous exercisers at the end of the intervention program.

Summary

Peer-reviewed literature

More than 40 studies from the scientific literature were reviewed. Summary points from this review are as follows:

- Even when associations between exercise or physical fitness and other behaviors are found, the correlations are relatively low.
- High levels of physical activity are associated with high calorie intakes.
- Dietary composition does not appear to be associated with exercise pattern.
- High levels of habitual exercise are associated with better weight control.

Table 15. Selected health behaviors, by exercise category: Personalized Aerobics Lifestyle System, city employees
[Percent]

Item	Baseline		Followup	
	Vigorous exerciser (*n* = 25)	Not vigorous exerciser (*n* = 296)	Vigorous exerciser (*n* = 52)	Not vigorous exerciser (*n* = 269)
Smoking				
Never smoked cigarettes	36.0	58.5	65.4	59.1
Current smoker	43.8	51.2	18.2	48.1
Stress				
High occupational stress level	36.0	32.7	44.2	30.7
Difficulty sleeping several times per week	12.0	11.2	3.8	4.6
Risk-taking behavior				
Use seatbelts 75–100% of time	40.0	18.9	55.5	25.1
Preventive health behavior				
Had a general check up in last year	44.0	43.7	47.1	51.1
Visit dentist at least once/year	44.0	58.0	63.5	61.4
Pap smear at least once/year (women only)	62.5	69.7	81.8	71.4
Zero days of work missed last year because of illness	24.0	29.7	—	—
Substance abuse				
Alcohol intake pattern spread over 1 week	27.3	42.9	35.1	44.9
Often or sometimes use drugs to help relax	20.0	10.4	7.7	6.9

- Smoking status and exercise pattern associations are confounded by type of activity. Occupational physical activity is positively associated with smoking status. Leisure time physical activity is inversely associated with smoking, although the differences in smoking between activity groups are somewhat lower than expected.
- More active individuals seem to be more likely to engage in selected preventive health behaviors such as regular visits to physicians and dentists. Studies did not always control for potential confounders such as age and sex, and results should be cautiously interpreted.

National surveys

Three recent surveys, two from the United States and one from Canada, were also reviewed. These surveys used scientific sampling schemes to obtain representative samples, thus the results should be more generalizable than the findings from many of the studies in the scientific literature (which frequently relied on select groups). The findings from national surveys generally agree with the conclusions of the literature review. One of the national surveys presented data on social health variables, items not typically reported (59). These results indicate that physically active individuals believe that exercise has had either a beneficial effect on relationships or no effect. Very few believed that exercise had a negative effect. Nearly one-half of the exercisers thought that they were more creative at work since starting exercise programs. The American Health Survey also presented some of the strongest evidence that starting exercise is associated with smoking cessation.

Original analyses

We also conducted several new analyses for this report. Data sets from the National Center for Health Statistics and the Institute for Aerobics Research were used. One of the primary objectives of these analyses was to identify extreme groups in the physical activity spectrum to test the hypothesis that absence of expected correlations in existing studies may be because of the lack of precision in identifying the highly active. Scoring algorithms were developed that identified the most active and most sedentary. In both cases, these categories consisted of approximately 10% of the total group. The generalizations derived from these analyses were similar to those obtained from the literature review and from the previously published results of national surveys. However, our hypothesis regarding extreme groups was also supported in several instances. The most highly active had much lower smoking rates than were seen in the earlier studies. Positive associations between exercise and preventive health behaviors were also seen.

Calculation of age-, sex-, and education-adjusted proportions in the health behavior by exercise category tables were not presented, and these additional analyses must be done. Time constraints as well as the actual distribution of the exercise categories precluded this. Many of the individual cells in the most active and the most inactive groups had severely unequal sizes. This lack of consistency across age-sex-education groupings tends to mask real differences seen in these health behaviors between exercise groups. For each population (except NSPHPC, for which extensive demographic data are published), we have provided pertinent demographic information that can be used in interpretation of the cross-tabulations.

Recommendations

Recommendations for future research

The associations between living habits and chronic disease are strong. Most of the leading causes of death in the United States are at least partly related to lifestyle. Therefore, it is critical to thoroughly understand these lifestyle factors and their intercorrelations if better research on the relationship of lifestyle to health is to be done. Furthermore, these issues have importance for national health policy and planning. We consider it highly important that the National Center for Health Statistics (NCHS) obtain data on the various lifestyle factors from a representative sample of Americans. These data can then be analyzed to better understand lifestyle-health associations. Several specific issues need to be investigated.

- There is a need for more detailed studies to determine the chronic effect of regular physical activity on appetite regulation.
- More studies are needed on specific population subgroups, such as various racial, ethnic, age, and geographic groups. National surveys by the National Center for Health Statistics are ideally suited to obtain this information.
- There is much clinical and lay opinion supporting the value of regular exercise as a stress management technique. Very few data are currently available to support or refute this hypothesis, and more studies are needed.
- Intervention studies should be conducted to determine the possible contribution of vigorous exercise as an adjunct to smoking cessation and alcohol intervention therapy.
- Because the foundations for the various lifestyle contributors to health are established in childhood and adolescence, more studies for this age group are needed. For example, the use of exercise in a smoking prevention program could be investigated in these age groups. This may ultimately prove to be more useful than similar studies for adults whose smoking patterns are already well established.
- The greatest barrier to research in the area under consideration in this paper is undoubtedly the lack of assessment tools to precisely determine physical activ-

ity patterns. This shortcoming also exists for several of the other health behaviors. We strongly recommend that the National Center for Health Statistics contribute to the development of better methodology for assessment of physical activity and other health behaviors.

- As mentioned previously, there are some areas in which physical activity patterns appear to be related to other health behaviors. Additional work is needed in these areas to establish a threshold or a dose response relationship.
- Although premature at this time, cohort studies and, ultimately, randomized clinical trials on physical activity and other health behaviors need to be implemented to more completely address causality issues. Studies of these types will be able to address the differences between development of healthy behaviors and the maintenance of such. This distinction may be especially important if physical activity could be hypothesized as either an initiating or promoting influence on other health behaviors. It may be of interest to study these distinctions and how each may relate to physical activity. A tenable hypothesis may be that interrelationships among health behaviors may not occur simultaneously. It is possible that changes in physical activity may initiate subsequent positive changes in other health behaviors and then in turn help to maintain those behaviors. Time series analyses of longitudinal data sets are necessary to test these and other similar hypotheses.
- There is a need for much more extensive analyses on the issues discussed in this report. Many of the existing analyses, including our own for this report, have been motivated by explanatory and descriptive intentions or have been secondary to other analytic study objectives. Primary analyses of large representative samples using multivariate techniques are needed. Very little is known about cultural and demographic characteristics in relation to the associations described in this paper.
- We strongly recommend that repeat surveys on exercise and physical fitness be planned by NCHS. It is very unfortunate that we have so little data on exercise and fitness trends over the last several years, when there have apparently been major changes in the population over that time. It is critical that we now begin to obtain such data.
- It would be helpful for future surveys to obtain more extensive psychosocial data. This may be valuable in helping to understand the determinants of exercise and in further exploring the exercise or physical fitness associations with other health behaviors.

Implications for NCHS surveys

As stated above, it is critical to future research in this area that more sophisticated physical activity assessment methodologies be developed, tested, and validated. Improved techniques should enable us to quantify activity more precisely and to identify accurately highly active individuals. It seems reasonable to conclude that associations between exercise and other health behaviors are more likely to be found in this highly active group. For example, it is somewhat incompatible for an individual to be a long-distance runner and continue to smoke cigarettes. It is plausible, however, for a person to be above the median in total activity and continue to use cigarettes. Assessment techniques for other health behaviors such as the use of stress management techniques are virtually absent.

Most of the work on the topic of this report has dealt with exercise behavior. Assessment of physical fitness on a representative sample of Americans in the National Health and Nutrition Examination Survey would provide useful and objective measures to further examine the association of health behavior to exercise or physical fitness. National estimates of physical fitness would also be useful for numerous scientific and public health purposes, and we strongly urge NCHS to obtain such data.

References

1. National Center for Health Statistics. *Health, United States, 1983.* U.S. Government Printing Office, Washington, D.C., 1984.
2. Dean, A.G., West, D.J., and Weir, W.M.: Measuring loss of life, health, and income due to disease and injury. A method for combining morbidity, mortality, and direct medical cost into a single measure of disease impact. *Pub. Health Rep.* 97:39–47, 1982.
3. Burdette, M.E., and Mohr, M.: Characteristics of Social Security Disability Insurance Beneficiaries, 1975. U.S. Department of Health, Education, and Welfare, SSA Pub No. 13–11947, 1979.
4. Cooper, B.S., and Rice, D.P.: The economic cost of illness revisited. *Soc. Sec. Bull.* 21–36, 1976.
5. Milio, N.: *Promoting Health Through Public Policy.* Philadelphia, F.A. Davis Company, 1981.
6. Paffenbarger, R.S., Hyde, R.T., Wing, A.L., and Steinmetz, C.H.: A natural history of athleticism and cardiovascular health. *JAMA* 252:491–495, 1984.
7. Kristein, M.M.: Economic issues in prevention. *Prev. Med.* 6:252–264, 1977.
8. Kottke, T.E., et al.: Projected effects of high-risk versus population-based prevention strategies in coronary heart disease. *Am. J. Epidemiol.* 121:697–704, 1985.
9. Montoye, H.J.: *Physical Activity and Health: An Epidemiologic Study of an Entire Community.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975, p. 95.
10. Folsom, A.R., et al.: Leisure time physical activity and its relationship to coronary risk factors in a population-based sample: the Minnesota Heart Survey. *Am. J. Epidemiol.* 121:570–579, 1985.
11. Cooper, K.H., et al.: Physical fitness levels vs. selected coronary risk factors. *JAMA* 236:166–169, 1976.
12. Gibbons, L.W., Blair, S.N., Cooper, K.H., and Smith, M.: Association between coronary heart disease risk factors and physical fitness in healthy adult women. *Circulation* 67:977–983, 1983.
13. Brownell, K.D.: Obesity: Understanding and treating a serious, prevalent, and refractory disorder. *J. Consult. Clin. Psychol.* 50:820–840, 1982.
14. Epstein, L.H., and Wing, R.R.: Aerobic exercise and weight. *Addict Behav.* 5:371–388, 1980.
15. Thompson, J.K., Jarvie, G.J., Lahey, B.B., and Cureton, K.J.: Exercise and obesity: Etiology, physiology, and intervention. *Psychol. Bull.* 91:55–79, 1982.

16. Wilmore, J.H.: Body composition in sport and exercise: Directions for future research. *Med. Sci. Sports Exercise* 15:21–31, 1983.
17. Blair, S.N., et al.: Comparison of nutrient intake in middle-aged men and women runners and controls. *Med. Sci. Sports Exercise* 13:310–315, 1982.
18. Garcia-Palmieri, M.R., et al.: Increased physical activity: a protective factor against heart attacks in Puerto Rico. *Am. J. Cardiol.* 50:749–755, 1982.
19. Montoye, H.J., Block, W.D., Metzner, H.L., and Keller, J.B.: Habitual physical activity and serum lipids; men, age 16–64 in a total community. *J. Chron. Dis.* 29:697–709, 1976.
20. Pomrehn, P.R., Wallace, R.B., and Burmeister, L.F.: Ischemic heart disease mortality in Iowa farmers: The influence of life-style. *JAMA* 248:1073–1076, 1982.
21. Short, S.H., and Short, W.R.: Four-year study of university athletes' dietary intake. *J. Am. Dietetic Assoc.* 82:632–645, 1983.
22. Wood, P.D., et al.: Increased exercise level and plasma lipoprotein concentrations: A one-year, randomized, controlled study in sedentary, middle-aged men. *Metabolism* 32:31–39, 1983.
23. Wood, P.D., et al: Effects of a two-year running program on plasma lipoproteins, body fat and dietary intake in initially sedentary men. (Abstract) *Med. Sci. Sports Exercise* 14:104, 1982.
24. Durrant, M.L., Royston, J.P., and Wloch, R.T.: Effect of exercise on energy intake and eating patterns in lean and obese humans. *Physiol. Behav.* 29:449–454, 1982.
25. Gorsky, R.D., and Calloway, D.H.: Activity pattern changes with decreases in food energy intake. *Hum. Biol.* 55:577–586, 1983.
26. Woo, R., Garrow, J.S., and Pi-Sunyer, F.X.: Effect of exercise on spontaneous calorie intake in obesity. *Am. J. Clin. Nutr.* 36:470–477, 1982.
27. Woo, R., Garrow, J.S., Pi-Sunyer, F.X.: Voluntary food intake during prolonged exercise in obese women. *Am. J. Clin. Nutr.* 36:478–484, 1982.
28. Hegsted, D.M.: Efficiency of utilization of various sources of energy for growth. Paper presented at Workshop on Activity Assessment Methods for Use in Epidemiologic Studies, sponsored by the National Heart, Lung and Blood Institute, Bethesda, MD. May 30–31, 1984.
29. Carney, R.M., and Goldberg, A.P.: Weight gain after cessation of cigarette smoking. *N. Engl. J. Med.* 310:614–616, 1984.
30. Kannas, L.: The dimensions of health behavior among young men in Finland: An overview of theories and findings. *Int. J. Health Educ.* 14:146–155, 1981.
31. Canada Fitness Survey. Fitness and Lifestyle in Canada, Fitness Canada, Ottawa, 1983.
32. Leon, A.S., Jacobs, D.R., DeBacker, G., and Taylor, H.L.: Relationship of physical characteristics and life habits to treadmill exercise capacity. *Am. J. Epidemiol.* 113:653–660, 1981.
33. Blair, S.N., Goodyear, N.N., Wynne, K.L., and Saunders, R.P.: Comparison of dietary and smoking habit changes in physical fitness improvers and nonimprovers. *Prev. Med.* 13:411–420, 1984.
34. Langlie, J.K: Interrelationships among preventive health behaviors: A test of competing hypotheses. *Pub. Health Rep.* 94:216–225, 1979.
35. The Perrier Study: Fitness in America. Perrier-Great Waters of France, Inc., New York, 1979.
36. Sedgwick, A.W., et al.: Long-term effects of physical training programme on risk factors for coronary heart disease in otherwise sedentary men. *Br. Med. J.* 6232:7–10, 1980.
37. Epstein, L., Miller, G.J., Stitt, F.W., and Morris, J.N.: Vigorous exercise in leisure time, coronary risk-factors, and resting electrocardiogram in middle-aged men civil servants. *Br. Heart J.* 38:403–409, 1976.
38. Bjartveit, K., Foss, O.P., and Gjervig, T.: The cardiovascular disease study in Norwegian countries: Results from first screening. *Acta Medica Scand.* (Suppl.) 675, 1983.
39. Holme, I., et al.: Physical activity at work and at leisure in relation to coronary risk factors and social class: A 4-year mortality followup. The Oslo Study. *Acta Medica Scand.* 209:277–283, 1981.
40. Hickey, N., et al.: Study of coronary risk factors related to physical activity in 15,171 men. *Br. Med. J.* 5081:507–509, 1975.
41. Blair, S.N., et al.: Physiological responses to maximal graded exercise testing in apparently healthy white women aged 18–75 years. *J. Cardiac Rehab.* 4:459–468, 1984.
42. Blair, S.N., et al.: Changes in coronary heart disease risk factors associated with increased treadmill time in 753 men. *Am. J. Epidemiol.* 118:352–359, 1983.
43. Sinyor, D., Brown, T., Rostant, L., and Seraganian, P.: The role of a physical fitness program in the treatment of alcoholism. *J. Stud. Alcohol* 43:380–386, 1982.
44. Hays, R., Stacy, A.W., and DiMatteo, M.R.: Covariation among health-related behaviors. *Add. Beh.* 9:315–318, 1984.
45. Taylor, C.B., Sallis, J., and Needle, R.: The relationship between physical activity and exercise to mental health. *Pub. Health Rep.* 100:195–202, 1985.
46. Blumenthal, J.A., Williams, R.S., Needels, T.L., and Wallace, A.G.: Psychological changes accompany aerobic exercise in healthy middle-aged adults. *Psychosom. Med.* 44:529–536, 1982.
47. Sinyor, D., et al.: Aerobic fitness level and reactivity to psychosocial stress: Physiological, biochemical, and subjective measures. *Psychosom. Med.* 45:205–217, 1983.
48. Kobasa, S.C., Maddi, S.R., and Puccetti, M.C.: Personality and exercise as buffers in the stress-illness relationship. *J. Behav. Med.* 5:391–404, 1982.
49. Blumenthal, J.A., Williams, R.S., Williams, R.B., and Wallace, A.G.: Effects of exercise on the Type A (coronary prone) behavior pattern. *Psychosom. Med.* 42:289–296, 1980.
50. Williams, A.F., and Wechsler, H.: Interrelationship of preventive actions in health and other areas. *Health Serv. Rep.* 87:969–976, 1972.
51. Belloc, N.B., and Breslow, L.: Relationship of physical health status and health practices. *Prev. Med.* 1:409–421, 1972.
52. Paxton, S.J., et al.: Effect of physical fitness and body composition on sleep and sleep-related hormone concentrations. *Sleep* 7:339–346, 1984.
53. Mechanic, D., and Cleary, P.D.: Factors associated with the maintenance of positive health behavior. *Prev. Med.* 9:805–814, 1980.
54. Criqui, M.H., et al.: Clustering of cardiovascular disease risk factors. *Prev. Med.* 9:525–533, 1980.
55. Tapp, J.T., and Goldenthal, P.: A factor analytic study of health habits. *Prev. Med.* 11:724–728, 1982.
56. Stephens, T.: Health practices and health status: evidence from the Canada Health Survey. *Am. J. Prev. Med.* 2:209–215, 1986.
57. Kok, F.J., Matross, A.W., Van den Ban, A.W., and Hautvast, G.A.J.: Characteristics of individuals with multiple behavioral risk factors for coronary heart disease: The Netherlands. *Am. J. Pub. Health* 72:986–991, 1982.
58. Berkman, L.F., and Breslow, L.: *Health and Ways of Living.* Oxford University Press, New York, 1983.
59. Public attitudes and behavior related to exercise. Conducted for American Health. The Gallup Organization. Princeton, N.J., Feb. 1985.
60. SPSS, Inc. SPSSX Users Guide. McGraw Hill, New York, 1983.
61. Blair, S.N., Goodyear, N.N., Gibbons, L.W. and Cooper, K.H.: Physical fitness and incidence of hypertension in healthy normotensive men and women. *JAMA* 252:487–490, 1984.
62. National Center for Health Statistics: Highlights from Wave I of the National Survey of Personal Health Practices and Consequences: United States, 1979. *Vital and Health Statistics.* Series 15–No. 1. DHHS Pub. No. (PHS) 81-1162. U.S. Government Printing Office, Washington. 1979.
63. National Center for Health Statistics: Basic Data From Wave I of the National Survey of Personal Health Practices and Consequences: United States, 1979. *Vital and Health Statistics.* Series 15–No. 2. DHHS Pub. No. (PHS) 81-1163. U.S. Government Printing Office, Washington. 1979.

64. Schoenborn, C.A. and Drury, T.F.; Response rates and nonresponse bias in the National Survey of Personal Health Practices and Consequences: United States, 1980. National Center for Health Statistics, Working Paper Series, 12, Oct. 1982.
65. Eisenstadt, R.K. and Schoenborn, C.A.: Basic data from Wave II of the National Survey of Personal Health Practices and Consequences: United States, 1980. National Center for Health Statistics, Working Paper Series, 13, Oct. 1982.
66. National Center for Health Statistics: Public use data tape documentation. Wave I and Wave II of the National Survey of Personal Health Practices and Consequences 1979–1980. Hyattsville, Md., 1982.
67. Blair, S.N., Collingwood, T.R., Reynolds, R., Smith, M., Hagan, R.D., and Sterling, C.L.: Health promotion for educators: Impact on health behaviors, satisfaction, and general well-being. *Amer. J. Public Health* 74:147–149, 1984.

Chapter 24

Evaluating Interrelationships Among Physical Fitness and Activity Assessments

Ronald E. LaPorte, Ph.D.
University of Pittsburgh

Physical activity appears to be an important component of health, whereby increased activity appears to be associated with a reduction of several chronic diseases, including cardiovascular disease (1–3), obesity (4), osteoporosis (5), and noninsulin-dependent diabetes mellitus (6). Moreover, several disorders associated with disease also appear to be highly associated with activity. These include insulin resistance (5), obesity (6), and hypertension (7).

There have been many articles that have reviewed the relationship of activity to health (1–4, 6). The current paper is not another review to add to the literature. Instead, the focus of the paper will be on evaluating the relationship between activity and fitness, on differentially reviewing the evidence linking activity and fitness to health, and finally on reviewing the currently available techniques for assessing physical activity.

Before beginning, it is important to define what is meant by physical activity and what is meant by fitness. We will be employing a definition of activity that has been recently suggested. Physical activity is defined as any bodily movement produced by skeletal muscle that results in energy expenditures (8). The energy expenditure can be measured in kilocalories. Physical fitness, in contrast, is typically defined in relationship to cardiovascular fitness; however, this need not be the case because fitness can also consist of strength, muscular endurance, coordination, etc. However, most people view physical fitness as synonymous with cardiovascular-aerobic fitness. Physical fitness is the capacity to perform prolonged heavy work. Individuals who are fit can maintain a higher energy expenditure than nonfit individuals. Typically, fitness is measured on a treadmill or bicycle ergometer by assessing maximal oxygen consumption ($\dot{V}O_2$max). Fitness is measured using the $\dot{V}O_2$max, and a highly fit individual has a higher maximum oxygen consumption than less fit individuals. Only certain types of physical activities are associated with increases of $\dot{V}O_2$max. The primary type of activity is the typical aerobic exercise pattern of three times per week at 70–80% of maximum heart rate for 20–30 minutes per session (9). Below this, exercise intensity increases or even maintenance of cardiovascular fitness have been inconsistent.

With these definitions of activity, it is possible to construct the model of physical activity.

The spectrum of physical activity

The spectrum of physical activity is presented in figure 1 (7). Physical activity can range from extremely low levels for individuals who are disabled (spinal cord injured, polio victims, arthritics, etc.) to sedentary able-bodied populations to high-activity groups. At the upper end of the physical activity spectrum, there is a separation, with the high activity-high fitness branch consisting of strenuous activity of sufficient intensity to achieve cardiovascular fitness. The high activity-low fitness branch represents long-duration activities not of a sufficient intensity to produce cardiovascular fitness.

In general, epidemiologic research and health surveys have focused on the upper end of the activity spectrum by comparing heart attack risk between sedentary working populations and active working populations. A problem with this approach is that only the upper strata of activity have been examined. Persons in the "inactive" control group of sedentary working individuals have a very active lifestyle when compared with the spinal cord injured or even individuals with a less drastic reduction in activity, such as that caused by arthritis or chronic back pain. The comparison in epidemiologic studies has, therefore, been between relatively high activity groups and moderately high activity groups, resulting in the evaluation of a truncated distribution of activity. It is difficult to determine a relationship between a factor and a disease if only a limited portion of the distribution of the factor is examined. As we have recently pointed out (7), it would be difficult to identify a relationship between smoking and heart disease if everyone smoked at least moderately. Therefore, an explanation for the studies that have not found a relationship between activity and heart attack may be that the groups under examination often have only a small difference in physical activity. It is striking that, even when examining only a small range in the activity spectrum, 60% of the studies have found a relationship. An extremely strong gradient should appear when the total spectrum, including activity less than that of able-bodied populations, is examined, and therefore such examination offers an opportunity for additional insight into the relationship of activity to disease.

Part of the reason that groups with impaired activity have not been included in studies to date may be the belief that they are only a small nonrepresentative segment of the population. However, surveys such as the Disability 1972 Survey (10) or the National Health Interview Survey have indicated that close to 15% of the U.S. adult population are disabled, and for some ages, over 30% have impaired activity. The criteria for entry into epidemiologic studies or national surveys have often screened out the lower strata of activity. For example, focusing on occupational groups or people who enroll in health insurance plans effectively eliminates most of the disabled. Moreover, even for community-based cohort studies, the individuals with impaired activity often do not fulfill the criteria for entry. Moreover, the disabled are much less likely to attend health screenings for which individuals are required to go to a central location. A large group of individuals with impaired activity is thus eliminated from much epidemiologic research or health surveys evaluating physical activity and health.

Another major difficulty is that researchers have not recognized that high activity consists of both high activity patterns that lead to fitness and high activity patterns that do not lead to fitness. The results of epidemiologic studies examining physical activity and

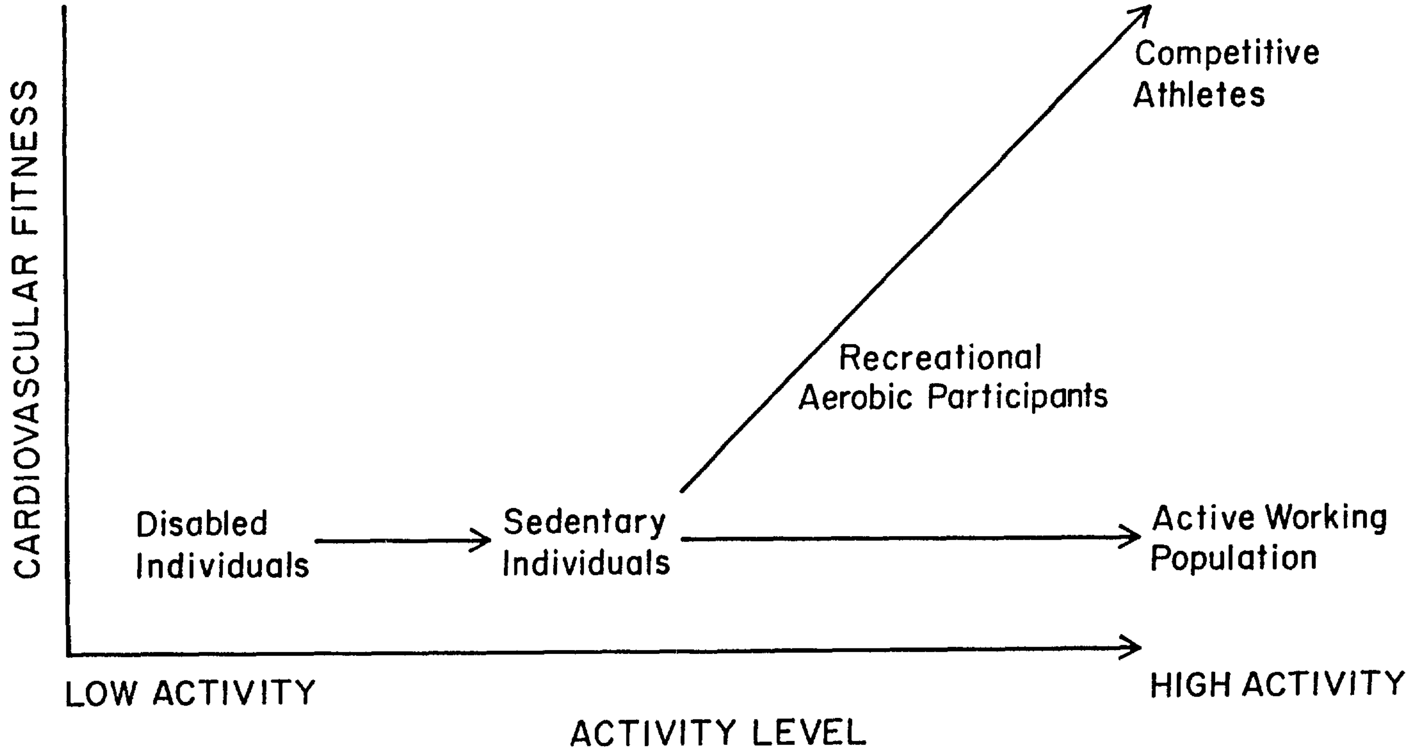


Figure 1. Spectrum of physical activity. From LaPorte et al. (7).

cardiovascular risk have been taken as evidence that individuals should become fit. Thus, the two high activity branches of the activity spectrum have been equated. Many have concluded that it is necessary for normal individuals to exercise at 80% of maximum heart rate two to three times per week for 20 minutes per session to reap the benefits of physical activity (9). The implied message is that fitness is a necessary condition for overall health and protection against cardiovascular disease. It is likely that fitness is beneficial; however, one could postulate that the high activity-low fitness branch is equally or more beneficial. An evaluation of the epidemiologic literature demonstrates that there is considerable evidence indicating that the high activity-low fitness branch is different and more important than the high activity-high fitness branch of the activity spectrum.

Relationship of fitness to cardiovascular disease

Suprisingly, very few studies have evaluated the relationship of fitness as measured by $\dot{V}O_2$max to subsequent risk of heart attack in population studies. Over 60 epidemiologic studies on "activity" have relatively consistently indicated that increased activity is associated with a reduced risk of heart attack (1–3). To our knowledge, there have been only five cohort studies directly relating $\dot{V}O_2$max to subsequent risk of heart attack in nondiseased populations (11–15).

Two studies suggested no relationship between fitness and coronary heart disease (CHD), and three indicated a positive relationship. Of interest is that the study of Gyntelberg et al. (15), which is presented in table 1, determined that the primary association was with the lowest quintile of fitness, whereby the people with fitness levels in the bottom fifth of the fitness distribution were two to three times more likely to have a heart attack than those with levels in the upper quintile. One could argue that this is not the type of pattern that one would expect if the association was an exercise-fitness-disease association because so few people within a population exercise to achieve cardiovascular fitness (probably less than 20%). The exercise-fitness-disease relationship would predict that the locus of the effect would more likely be with the upper 20% of fit individuals, whose heart attack risk is reduced, rather than the lowest 20%, whose risk is markedly increased, as demonstrated in the Gyntelberg data. One wonders whether a more likely explanation is that the lower 20% already had underlying disease that reduced their $\dot{V}O_2$max rather than that fitness was protective. Irrespective, however, the data would suggest that movement from the lowest quintile of fitness to the upper 80% of fitness would predict significantly greater benefit than movement from the upper 40% to the upper 20%. Thus, extremely low levels of fitness appear to be associated with increased CHD risk; however, once beyond this very low level of $\dot{V}O_2$max, there does not appear to be a strong gradient between activity and fitness. Moreover, as presented in table 2, randomized aerobic exercise trials with postmyocardial infarction patients have contributed little support for the concept that increased fitness is associated with reduced CHD risk (16). Thus there is minimal direct evidence that $\dot{V}O_2$max on an individual basis is associated with risk of heart attack.

Fitness levels of high activity groups in epidemiologic studies

Examination of the epidemiologic studies of physical activity and coronary heart disease indicates that the high activity groups may not achieve a sufficient intensity level to produce cardiovascular fitness. For example, postal carriers or police officers walking a beat appear to have a lower risk of heart attack than sedentary controls. Direct measurement of cardiovascular fitness ($\dot{V}O_2$max) on these groups has failed to reveal an increase in cardiovascular fitness. Similarly, in the Morris et al. study of the London transit workers (17), it is unlikely that the conductors had higher fitness levels than the drivers, even though there was a lower heart attack risk among the conductors. Careful examination of the high activity groups in epidemiologic studies relating activity to cardiovascular disease reveals few groups that would have high levels of cardiovascular fitness, none of which approaches that of marathon runners, for example.

Table 3 presents a comparison of the energy expenditure for high and low activity by groups that have been shown to have different CHD risks (18). As can be seen, the difference in caloric expenditure is quite small,

Table 1. Relationship betwen physical fitness measured by indirectly determined maximal oxygen uptake (ml $O_2 \cdot kg^{-1} \cdot min^{-1}$), and incidence of myocardial infarction among 5,249 Copenhagen males aged 40–59 in 1970–71

Physical fitness score (ml $O_2 \cdot kg^{-1} \cdot min^{-1}$)	MI	T	R
At entry		Percent	
27	37.3	22.5	27.3
28–30	20.1	17.4	20.3
31–34	18.3	22.3	20.5
35–38	11.2	17.6	14.5
39	13.0	20.3	17.3
		$p < 0.00001$ % i = 26.71	$p < 0.005$ Kendall's tau 13 0.08
At 1–a followup			
27	31.5	21.7	28.5
28–30	26.1	18.3	20.0
31–34	24.3	24.6	22.9
35–38	10.8	18.4	15.5
39	7.2	17.0	13.0
		$p < 0.01$ % i = 17.82	$p < 0.03$ Kendall's tau 13 0.07

Notes: Lethal and nonlethal myocardial infarction cases (MI) are compared with noninfarction cases in the total sample (T) and age-matched referents (R). The distribution within quintiles of physical fitness at entry into the study and at 1–a followup is presented.
Source: Gyntelberg et al. (15).

Table 2. Physical exercise: results

Trial	Control			Intervention			
	Number randomized	Number of deaths	Percent mortality	Number randomized	Number of deaths	Percent mortality	p value [1]
Sarne et al.	157	35	22.3	158	28	17.7	0.38
Kantale [2]	146 (81)	32(8)	21.9 (9.9)	152 (77)	26 (5)	17.1 (6.5)	0.37 (0.63)
Palatari [3]	200	28	14.0	180	18	10.0	0.30
Kallig et al.	187	56	29.9	188	41	21.8	0.093
National Exercise and Heart Disease Project	328	24	7.3	323	15	4.6	0.20
Southern Ontario Multicenter Exercise Heart Trial [4]	354	26	7.3	379	36	9.5	0.36

[1] p values computed for chi-square test comparing the proportion of deaths in each group.
[2] Numbers in parentheses refer to patients considered suitable for long-term followup.
[3] Primary end point is coronary death.
[4] Incomplete reporting; personal communication.
Source: May et al. (16).

probably not being associated with major differences in cardiovascular fitness.

To our knowledge, there have been no epidemiologic studies that have evaluated the specific relationship between aerobic exercise programs (for example, engaging three times per week in a relatively intense program such as jogging) and subsequent risk of heart attack. This is probably the result of there being relatively few individuals in cohort studies who have an exercise pattern such as this. However, studies have indicated that people engaged in strenuous-vigorous exercise are at lower risk of subsequent heart attack. The definition of strenuous-vigorous exercise used in epidemiologic studies typically is of a much lower intensity than prescribed exercise programs. For example, in the classic studies by Morris (17, 19) and Chave et al. (20), it was found that civil servants in England who reported "vigorous exercise" in their leisure time had a lower subsequent risk of heart attack than people who did not report vigorous exercise. Vigorous exercise was defined as peak energy expenditure greater than 7.5 kcal/minute. Of particular importance was that the primary form of vigorous exercise was gardening. Therefore, the definition of vigorous or strenuous exercise used in epidemiology that has been related to risk of heart attack typically is of a much lower magnitude than that prescribed in an aerobic exercise program. The epidemiologic research might suggest that gardening three times a week is as effective as running for the prevention of heart attack. Therefore, the high activity groups in population-based research are likely to have only a modest, if any, increase in cardiovascular fitness relative to the low activity groups, suggesting that activity is likely to be more importantly related to risk of heart attack than fitness.

It is interesting that when both activity and $\dot{V}O_2$max are measured within representative populations, not clinic populations, the association between the two has been either weak or nonexistent. For example, in Evans County the correlation between activity and fitness was only 0.13 (21), and in a large study from Belgium, all correlations between measures of activity and fitness were <0.15 (22). The weak relationship indicates that, within a population, the individuals who are most active are not necessarily the most fit. Clearly, there are extremely active individuals who are not highly fit, as previously mentioned: Postal carriers, waitresses, or mothers with several young children (23). There also are people who have very high $\dot{V}O_2$max levels but who have not exercised regularly for decades. For example, ex-Olympic athletes tend to be very fit as measured by $\dot{V}O_2$max despite their inactive lifestyles (24). Moreover, the physiologic response to the same exercise regimen is quite variable. Some individuals demonstrate major changes in $\dot{V}O_2$ and others have only modest changes. This is not what one would expect if activity is the primary determinant of cardiovascular fitness. One possible reason is that an individual's $\dot{V}O_2$max level on a population basis is primarily genetically determined rather than activity related (7). Admittedly, activity can influence fitness levels, but given the relatively low variability of activity on a population basis, a person's

Table 3. Amount of on-job activity associated with less coronary heart disease

Study	Estimated kilocalories expenditure		Type of activity
	Rate/min	Total/day [1]	
Evans County (1966)	3–7.5	400–500	Farming and laboring
North Dakota (1959)	5–8	300–600	Farming and laboring
Health Insurance Plan of Greater New York (1969)	4–8	300–500	Walking, lifting, carrying
U.S. Railroad (1966)	5–8	350–600	Walking, hanging, climbing
Longshoremen (1977)	5.2–7.5	810	Cargo handling

[1] Difference between sedentary and active groups.
Source: Haskell (18).

heredity is probably a more important determinant of $\dot{V}O_2$ max levels than activity.

A popular misconception has been developed that one needs to run 20 miles per week and thus be very physically fit in order to achieve maximum reduction in risk of heart attack. This concept has been presented and was based on the results of analyses completed by Paffenbarger in his excellent study of Harvard alumni (25). As is presented in figure 2, the research demonstrated that the risk of coronary heart disease declined as activity above that necessary for daily living increased to a level of 2,000 kcal per week. There was no additional benefit for activity above 2,000 kcal per week. The figure of 20 miles per week was determined because it takes approximately 100 kilocalories to run a mile; thus 20 miles represented 2,000 kcal per week in energy expenditure. In addition, this dose of exercise fit very nicely with the current concepts of the exercise regimen needed to maintain cardiovascular fitness.

There are major problems in using Paffenbarger's results as proof that one should run 20 miles per week. The primary difficulty is that the 2,000 kcal figure came from a series of three questions in the survey. The questions evaluated the blocks walked per day, the number of flights of stairs climbed per day, and other recreational and sports activities. We have recently found that the average baseline calorie expenditure per week for postmenopausal women was 1,200 kcal, none of which came from running. Moreover, over 70% of the calories came from walking and stair climbing (26).

Thus, interpreting the 2,000 kcal to mean 20 miles of running per week does not take into account the normal daily activities that people are doing, such as walking and stair climbing, which were assessed by Dr. Paffenbarger. It is relatively easy to achieve 2,000 kcal per week without engaging in a running program. Even if one engaged in a running program, it is likely that only 5–8 miles per week in addition to the normal activities of daily living would achieve the 2,000 kcal per week.

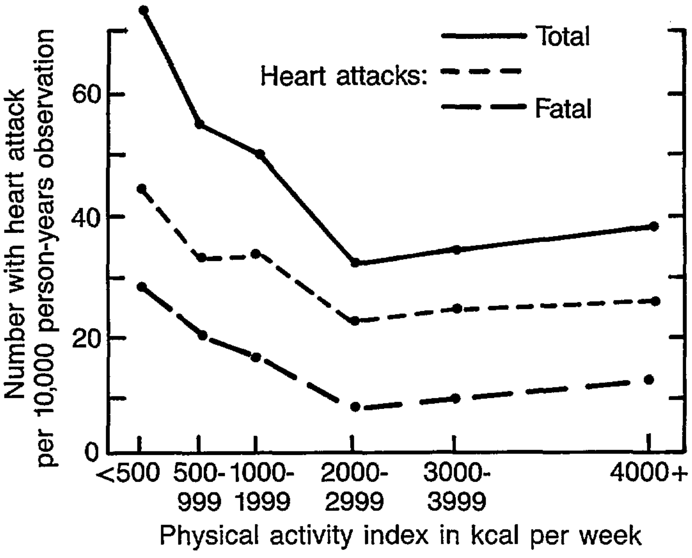


Figure 2. Age-adjusted first heart attack rates, by physical activity index, in a 6–10-year followup of male Harvard alumni. From Paffenbarger et al. (25).

It is interesting to speculate as to the mechanisms by which activity could reduce CHD risk. There is evidence that fitness improves the pumping of the heart. This has been hypothesized to be the result of a decrease in heart rate and an increase in stroke volume at rest or when exercising (27). There is, however, little evidence that the exercise-induced changes to the myocardium and heart function reduce the risk of CHD in normal populations. For example, one of the primary markers of improved fitness is a reduction of resting heart rate. However, resting heart rate does not appear to be a major predictor of heart attack (28–34).

It is likely that the effect of activity is through cardiovascular disease risk factors, the major candidate being the level of high-density lipoprotein cholesterol (HDL–C). HDL–C is a primary risk factor for CHD (35, 36) and appears to be associated with activity levels (37). Aerobic training programs have not been very successful in raising HDL–C levels; as shown in table 4, only about 50% of the reported studies have reported an increase (37). It is of interest that, even in the studies that have found an increase in HDL–C with aerobic training, to our knowledge there has been no aerobic training trial that has identified a strong significant correlation between changes in $\dot{V}O_2$max and changes in HDL–C on an individual basis. Instead, the studies report that the groups exhibited increases in $\dot{V}O_2$max and an increase in HDL–C. There thus is little evidence that the individuals who changed their fitness levels the most were also the individuals who increased their HDL–C levels the most.

Of interest is that Williams et al. (38) noted that participants with initially higher HDL–C levels were more easily persuaded to run longer distances. This may suggest that individuals who engage in aerobic exercise have higher HDL–C levels and higher fitness levels than sedentary controls and that there is a circular enhancing relationship between innate physiological makeup and the selection of a lifestyle. Yet, on a population basis, activity tends to be highly associated with HDL–C, with highly active groups demonstrating HDL–C levels 15 mg/100 ml above those of sedentary individuals (37). Furthermore, extremely inactive groups, such as spinal cord injured patients, have extremely low levels of HDL–C approaching 20 mg/100 ml below normal.

The physical activity pattern needed to raise HDL–C may not be the same as that needed to improve fitness. Evidence for this possibility comes from the Lipid Research Clinic study (39), which found no relationship between treadmill measures of fitness yet found a strong relationship to the activity question of, "Do you exercise or labor at least three times per week?" Also, there is recent evidence that suggests that anaerobic weight lifting may raise HDL–C. Thus, a different activity pattern may be needed to raise the HDL–C than to increase $\dot{V}O_2$max. If it is found that a different activity pattern is needed to raise HDL–C than is needed to achieve cardiovascular fitness, the issue would be:

Table 4. Uncontrolled longitudinal studies of the effect of exercise training on plasma and serum high-density lipoprotein (HDL) cholesterol concentration in men and women free of overt heart disease

Study	Subjects	Training level	Length of training (weeks)	Results for HDL cholesterol (mg/dl)
Ballantyne et al. (1981)	20 men, 16 women	Royal Canadian Air Force 5BX plan	24	Increased significantly for men (7.0 ± 10.3) but not for women (1.9 ± 13.9)
Brownell et al. (1982)	24 men, 37 women, 20–60 years	Aerobic exercise 1.5 hours/week	10	Not significant (men: 1.5 ± 4.1; women: −0.9 ± 7.3). However, percent change was increased significantly in men but not women
Dufaux et al. (1982) [1]	10 men, 21–25 years	Ran 40 km/week	8	Not significant
Farrell and Barboriak (1980)	7 men, 9 women, mean age 23 years	Ran 1.5–2 hours/week	8	Not significant (3.6)
Gilliam and Burke (1978) [1]	14 girls, 8–10 years	Strenuous exercise 3–3.5 hours/week	6	Significant increase of 4.6
Horby-Peterson et al. (1982) [1]	54 men, 19–28 years	Aerobic exercise 2 hours/week	12	Not significant (0.0)
Kaufman et al. (1980) [1]	16 men, 22–42 years	Ran 9.3 km/week	6	Not significant (0.0)
Leon et al. (1979)	6 obese men, 19–31 years	Walked 7.5 hours/week	16	Significant increase of 5.0
Lewis et al. (1976)	22 obese women, 30–52 years	Walked, jogged, and did calisthenics 2.5–3 hours/week	17	Not significant (4.7)
Lipson et al. (1980)	5 men, 5 women, 19–22 years	Walked or jogged 3.5 hours/week	6	Not significant (−4.0)
Moll et al. (1979)	14 women, 22–26 years	Walked or jogged 2.5–3 hours/week	6	Not significant (−5.0)
Nye et al. (1981) [1]	17 men, 31–44 years	Group calisthenics 1–1.5 hours/week	10	Not significant (0)
Rorkis et al. (1981)	22 women, 23–37 years	Increased their running from 21.8 to 74.2 km/week	16–28	Significant increase of 5.0
Sutherland and Woodhouse (1980)	21 men, 20–55 years	Trained for marathon	16	Significant increase of 14.0
Separate evaluations of treatment and control groups				
Huttumen et al. (1979) [1]	44 male exercisers, 46 male controls, 40–45 years	Exercised 3–4 hours/week	16	Exercisers: significant increase of 5.5; controls: 0.8 change not significant Exercisers significantly greater than controls at end of study
Keins et al. (1980)	24 male exercisers, 13 male controls	Exercised 1.5–3 hours/week	12	Exercisers: significant increase of 3.4; controls: 1.5 increase not significant
Moore et al. (1979)	11 male exercisers, 9 male controls, 31–44 years	Canadian Air Force 5BX exercise program	10	Exercisers: significant decrease of 14.3; controls: 5 increase not significant
Peltonen et al. (1981) [1]	20 male exercisers, 7 male controls, 23–56 years	Ran or skied at least 1.5 hours/week	15	Exercisers: significant increase of 3.1; controls: −0.8 change not significant
Weltman et al. (1980) [1]	11 male exercisers, 5 male controls, mean age 47 years	Walked 3 hours/week	10	Exercisers: not significant; controls: not significant
Direct comparison of treatment and control groups				
Allison et al. (1981) [1]	25 male, 23 female exercisers; 9 male, 9 female controls	Jogged 1.5–2.5 hours/week	8	Decreased significantly more in exercisers than controls for both men (−6.5 vs. 2.3) and women (−4.9 vs. −1.1)
Wood et al. (1983)	48 male exercisers, 33 male controls	Ran an average of 8.6 km/week	52	Not significant (1.8 ± 8.0 vs. 0.5 ± 6.1)
Wynne et al. (1980)	13 female exercisers, 6 female controls, 19–30 years, on oral contraceptives	Bicycle ergometer for 1.5 hours/week	10	Not significant (exercisers: 1; controls: −3)

[1] Used serum measurements; all others used plasma measurements.

Source: Reference 37.

Which activity pattern is more beneficial for health—activity designed to raise HDL–C or exercise to increase $\dot{V}O_2max$?

One needs also to examine other health conditions that have been associated with physical activity. Perhaps the disorder for which there is relatively strong evidence that activity reduces the risk of the disease is osteoporosis.

The population of the United States is aging rapidly. Currently, 11% of the population (25 million people—15 million men, 10 million women) are over the age of 65. By the year 2030, with the maturation of the Baby Boom generation, one in five Americans will be over the age of 65 (40). Associated with the greying of the population is an ever increasing prevalence of hip fracture, which is the extreme end product of osteoporosis.

Osteoporosis is a major health problem and cause of disability in older populations. With osteoporosis, there is an increased vulnerability of the bony skeleton, which may sustain fractures through minimal trauma such as slipping, stepping off curbs, sneezing, or even just turning over in bed. It is the second leading musculoskeletal disorder and is responsible for over 148,000 hip fractures per year. Of importance is the associated morbidity and mortality of hip fracture because the progress of victims suffering osteoporotic hip fractures is poor, with about 12–20% dying within 1 year and the survivors suffering major deteriorations in the quality of life (41).

The relationship between activity and osteoporosis is complex and has not been adequately addressed. A central problem is that there have been few studies focusing on the relationship of activity to fracture risk. As presented in figure 3, fracture risk is a function of the likelihood of trauma as well as bone density. The problem with physical activity is that it may be a double-edged sword, in that increased activity may retard bone loss, yet at the same time, it may increase the likelihood of falling, tripping, or running into objects. Most skeletal research has only evaluated the relationship of activity to bone density and has not considered the potentially detrimental effects of the increased trauma risk.

Ecologic studies

Many authors have employed ecologic types of analysis to demonstrate that with aging there is a marked decline in physical activity, as shown in figure 4 (42). Concurrently, there is a marked increase in bone loss and fracture risk, as shown in figure 5 (43). This decline of activity with aging may contribute to the reduction of bone density. Geographic differences in fracture risk indicate that less active populations have a higher fracture incidence (44). These ecologic analyses are suggestive of an association between physical activity and osteoporosis but are subject to the ecologic fallacy whereby population associations may not reflect indi-

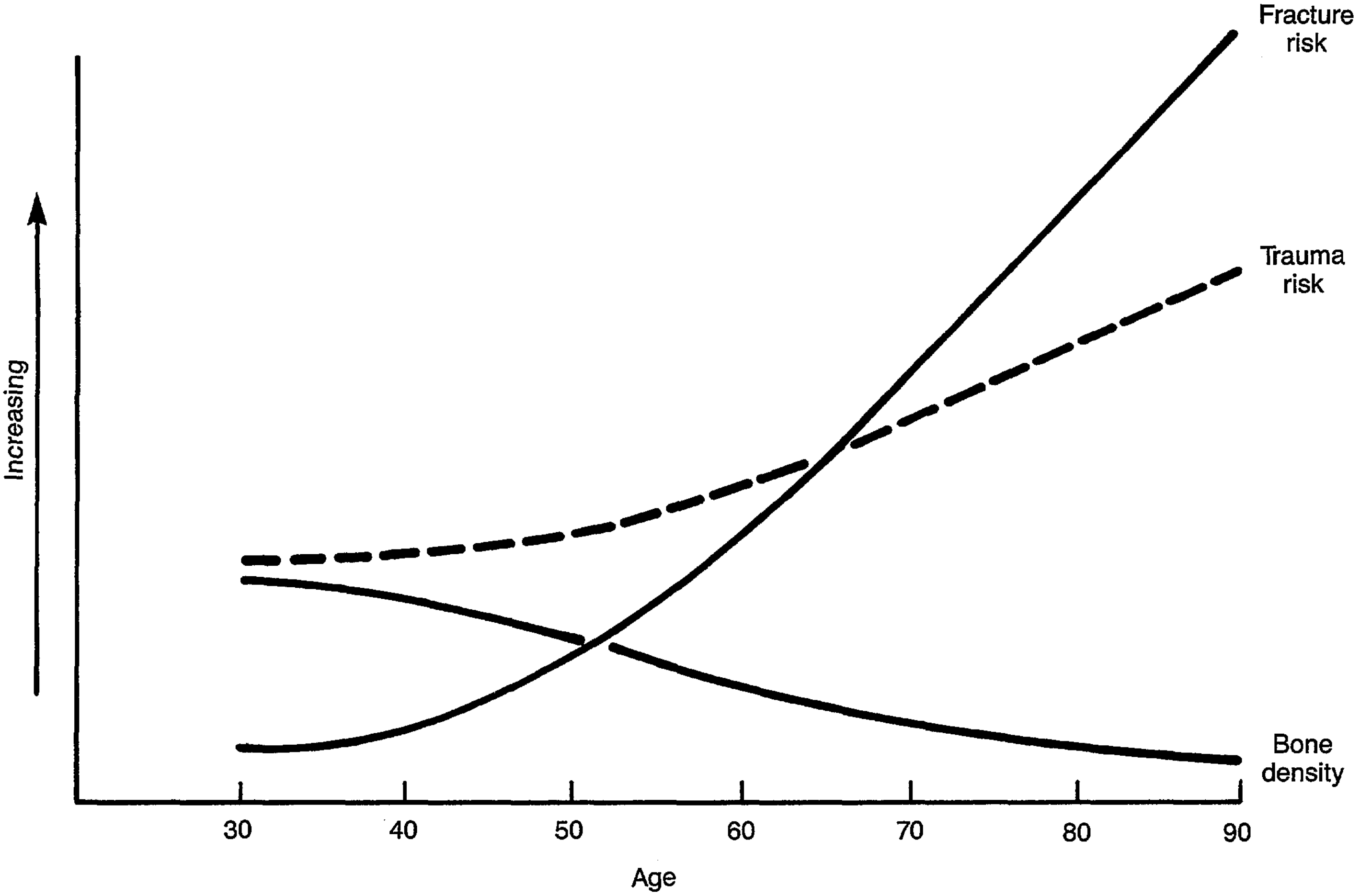


Figure 3. Relationship of bone density, trauma, and fracture risk.

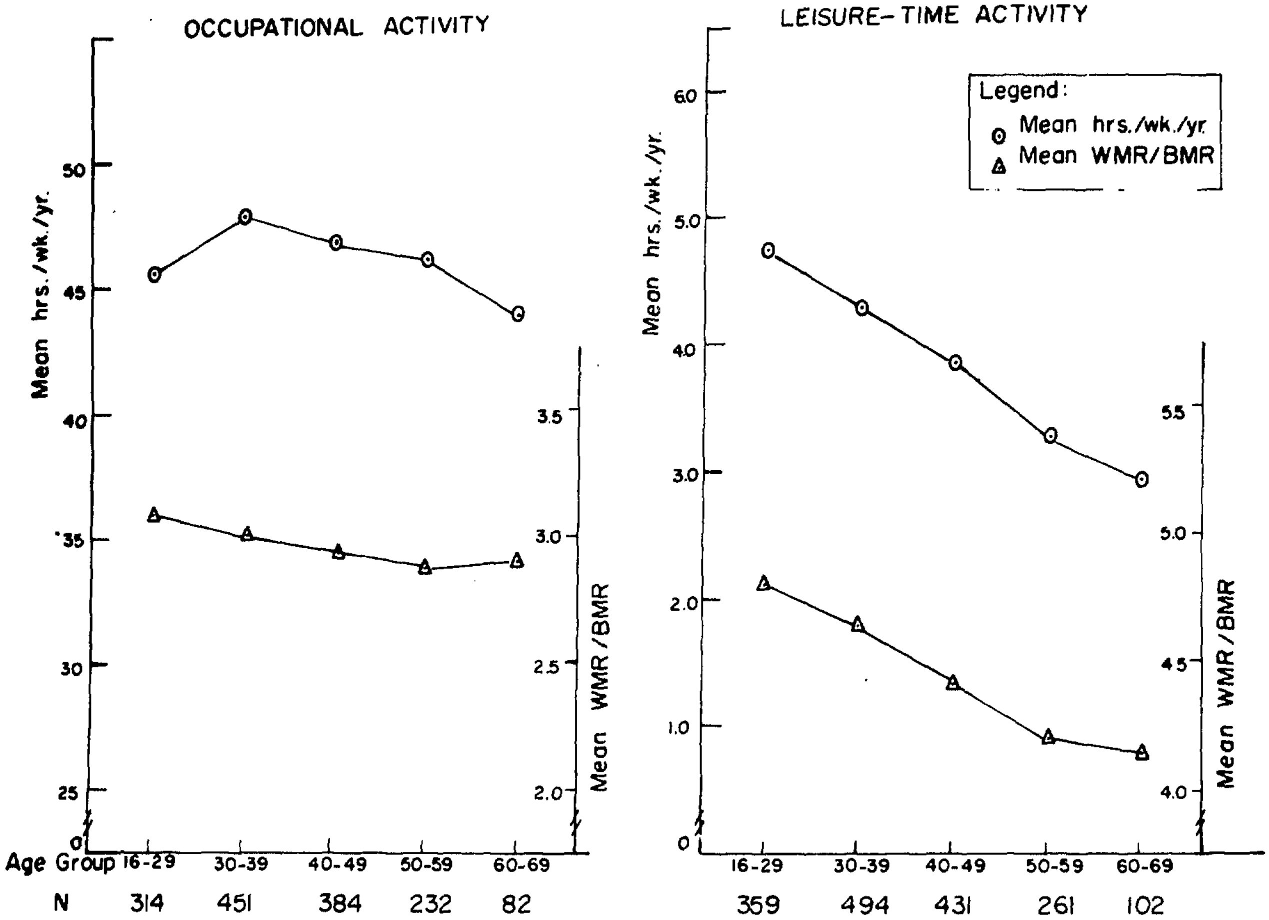


Figure 4. Left panel: mean hours worked and mean estimated energy expenditure (WMR/BMR) in occupation. Right panel: mean hours and mean estimated energy expenditure during active leisure pursuits. From Montoye (42).

vidual relationships. Studies that directly relate individuals' activity levels to bone density and fractures provide considerably more information than the ecologic studies, as is presented next.

Low activity

Considerable information has been derived from the evaluation of the effects of activity at the lower end of the activity spectrum (activity ranging from disabled individuals to sedentary able-bodied populations). This research can be broadly defined as immobilization (or disuse) osteoporosis and space osteoporosis research.

Immobilization osteoporosis is usually produced by a severe impairment of activity such as casting of a limb, chronic bed rest, and spinal cord injury (45–51). A consistent pattern emerges in which reduction of activity to levels characteristic of disabled people is accompanied by decrements in bone density. This has been clearly demonstrated in a variety of disabling conditions such as fractures (46), back pain (47), disc surgery (48), polio (49), spinal cord injury (50), and prolonged bed rest (51). It appears that a reduction in activity to the levels encountered in the disabled is accompanied by a rapid reduction in bone density.

Space osteoporosis has many of the characteristics of immobilization osteoporosis (52) because there is a rapid loss of calcium and loss of bone over short periods of time. For example, astronauts and cosmonauts have

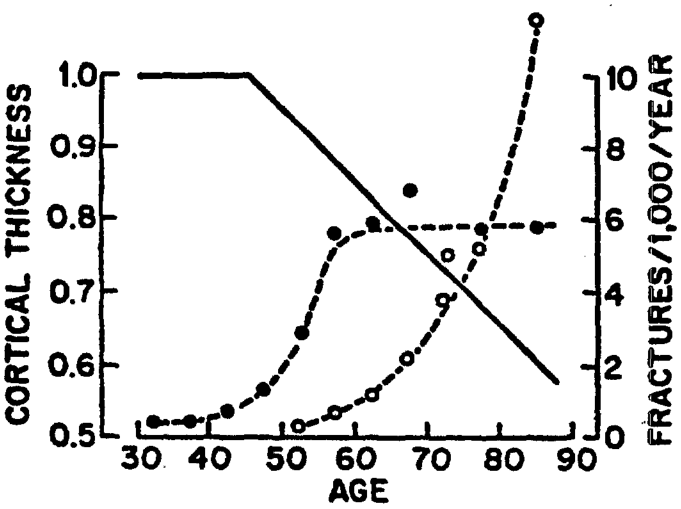


Figure 5. Decrease in thickness of metacarpal cortex in aging women (—) and age-specific frequencies of fracture of the forearm(•) and femur (○). Data from Morgan B: Osteomalacia, renal osteodystrophy and osteoporosis. Springfield IL: Charles C. Thomas, 1973. Source: Worley R.J. Age, Estrogen, and Bone Density. Clinical Obstetrics and Gynecology, Vol. 24, No. 1, March 1981. Copyright 1981 by Harper & Row, Publishers, Inc.

lost a considerable amount of calcium and bone even during relatively short flights.

Immobilization osteoporosis and space osteoporosis provide important information concerning the relationship of physical activity to bone loss. One important aspect of both immobilization osteoporosis and space osteoporosis is that when individuals regain their activity, bone density increases. Yet, the increase in bone density accompanying the return to activities of daily living generally does not result from intense aerobic exercise (53).

The second important point is that aerobic exercise does not appear to be needed, but rather activity that reflects primarily the *pull of gravity* on the bone. Thus, spinal cord injured patients lose calcium and bone very rapidly irrespective of whether they are spastic or flaccid. The spasticity often involves a considerable degree of skeletal muscle movement. It is only when weight-bearing activities are resumed that the calcium and bone loss is slowed (50). Similar results with normal individuals confined to bed indicate that an increase in weight-bearing activity resulting from tilting or rocking may reduce calcium excretion (54). Moreover, aerobic or anaerobic strength exercises by astronauts do not appear to prevent bone loss despite maintenance of cardiovascular fitness levels. It is only the return to gravity that reverses the deterioration of bone structure seen in weightless environments (55).

High activity, high fitness

Moving up the activity scale to the aerobic level, several important studies focused on the relationship of aerobic exercise to bone density. Highly conditioned athletes, such as marathon runners or cross-country skiers, considered by many to be the epitome of physical activity and fitness, have been found in several studies to have higher bone densities than nonathletes (56).

Krolner et al., from Denmark, have recently published data from a trial of 31 women ages 59–73, all with an earlier history of fracture of the distal forearm (57). Sixteen women were assigned to an exercise group and 15 to the control group. This was not a randomized trial. The exercise consisted of an aerobic training program of running. Improvements were not found in "fitness." Moreover, the exercise group actually increased bone mineralization. Similarly, Aloia demonstrated an increase in total body calcium in an intense exercise trial with eight women (58).

Smith (59), from Wisconsin, has also published results from a nonrandomized trial in which 80-year-old women were assigned to either an exercise or a non-exercise program. The exercise program consisted of a series of chair exercises that were designed to increase cardiovascular fitness. At the end of the 36 months, only 51 women (65%) continued to participate. The evidence appeared to indicate that increased activity retarded loss of bone.

It is important to point out several negative aspects of such intense aerobic activities, lest we become enamored with intense exercise designed to increase fitness. The first is the recent report on 38 young, *amenorrheic* women, all but one participating in vigorous exercise programs. These women had markedly lower bone density than age-matched controls (60). The second problem is the issue of injury. A recent article on young individuals running the Peachtree 10-kilometer road race has revealed a markedly increased prevalence of orthopedic problems (61). Over 38% of the females experienced an injury in the previous year. Thus, the aerobic activity of running is likely to expose individuals to markedly increased trauma risk. This would be even more evident in older populations.

High activity, low fitness

An important aspect of activity is the potential effect of low-intensity, long-duration activities in which there is an increase in *weight-bearing activity* with no concomitant increase in fitness. Several studies have indicated that increased muscle area is associated with higher bone densities (62, 63). Increased muscle areas are best achieved with anaerobic types of activity such as weight lifting. Of particular interest is that a comparison of various types of sports indicates that the highest bone densities are found in weight lifters, the next highest is runners. In contrast, there was no major increase in bone density in swimmers, which again suggests that weight-bearing activity may be the critical factor associated with increased bone density (64).

The results again indicate that there is little evidence or biologic plausibility that cardiovascular fitness is associated with bone density. The most likely association is with weight-bearing activities. Therefore, one needs to accurately quantify the activity patterns in order to relate this (but not aerobic fitness) to risk of osteoporosis.

Moreover, it is also very difficult to conceptualize the biologic plausibility linking aerobic exercise to other diseases or conditions, including obesity, psychiatric disorders, osteoarthritis, and low back pain, each of which has been associated with activity.

The question then arises as to why we should measure cardiovascular fitness when: (a) It is not as highly associated with risk of disease as activity, (b) the relationship between activity and physical fitness on a population basis is relatively weak, and (c) most cardiovascular fitness on a population basis is genetically determined rather than determined through physical activity. It is likely that, for surveys designed to evaluate associations with health or risk factors, assessments of physical activity are as profitable, if not more profitable, than assessment of physical fitness.

The major difficulty is that it is much more difficult to assess physical activity than it is to measure physical fitness. There are several well designed and reproducible protocols for treadmill testing for assessing fitness. With physical activity, the methods for assessment are in their infancy. A recent review, which we have condensed here, shows that the measurement techniques are crude, and typically the validity and reliability of the measurement levels are not known (65).

Assessment of activity in population research

This review considers seven major categories (table 5) of physical activity assessment procedures that have been used in various settings and evaluates their potential for use in surveys. The instruments should fulfill certain criteria.

To be *valid*, the instrument must measure what it is intended to measure. To be *reliable,* the instrument must consistently give the same results under the same circumstances. If the instrument is reliable and valid, it is also accurate. To be *practical,* the instrument must have acceptable costs to both the investigator and the participant. To be *nonreactive,* the instrument must not alter the population or the behavior it seeks to measure.

Five primary types of physical activity assessment are available for health surveys. These include classification of jobs, surveys, physiologic markers of activity, objective monitoring techniques, and dietary assessment measures.

Job classification

Job classification has been used as an index of physical activity by ranking jobs according to levels of activity and assuming that all persons in that occupational category expend similar levels of energy. This method, in use for many years, can be employed in large representative populations at minimal cost and is nonreactive. However, job classification has limitations for assessing physical activity and may not be sufficiently valid and reliable for epidemiologic use. These limitations include:

- Within-job classification variability.
- Job intensity misclassification.
- Secular changes in job requirements.
- Seasonal changes in job requirements.
- Possible selection bias.
- The exclusion of leisure and nonoccupational physical activity.

These limitations may explain why epidemiologic studies using job classification schemes have not produced a clear pattern concerning the relationship of physical activity to coronary heart disease (1–3). However, it is important to be able to assess on-job activity.

First, variability of physical activity within a job class can be large. For example, cross-country truck drivers and local delivery truck drivers spend markedly different amounts of time sitting and unloading. This may contribute to a second problem, that job intensity is often estimated differently by independent judges. Third, job requirements can change over time. For example, the longshoreman occupation, extremely active in the 1950's, now is almost completely mechanized

Table 5. Characteristics of physical activity assessment procedures

	Group		Study costs [1]		Subject costs [1]		Probability of	Acceptability		
Activity assessment procedure	Size	Age	Money	Time	Time	Effort	interfering [1]	Personal	Social	Activity specifics
Calorimetry:										
Direct	Single	Infant-elderly	VH	VH	VH	H–VH	H–VH	No	No	Yes
Indirect	Single-small	Young adult-elderly	H–VH	VH	VH	M–VH	H–VH	No	No	Yes
Job classification	Large	Employed only	L–M	L–M	L	L	L	Yes	Yes	?
Surveys:										
Indirect calorimetry diary	Single-small	Young adult-elderly	M–H	M–H	M–H	M–H	VH	No	No	Yes
Task-specific diary	Small-large	Adolescent-elderly	L–M	L–M	H–VH	VH	VH	?	Yes	Yes
Recall questionnaire	Small-large	Adolescent-elderly	L–M	L–M	M–H	M–H	L	Yes	Yes	Yes
Quantitative history	Small-large	Adolescent-elderly	L–M	L–M	L–M	L–M	L	Yes	Yes	Yes
Physiologic markers:										
Cardiorespiratory fitness	Small-large	Child-elderly	M–VH	M–H	M–H	M–VH	L	?	?	No
Doubly-labeled water	Single-small	Infant-elderly	H–VH	M–VH	M	M	L–H	Yes	Yes	No
Behavioral observation	Single-small	Infant-elderly	H–VH	H–VH	H–VH	L–H	L–VH	?	?	Yes
Mechanical and electronic monitors:										
Heart rate	Single-small	Infant-elderly	H–VH	M–VH	M–H	M–H	L–M	Yes	Yes	No
Stabilometers	Single-small	Infant	M–H	M	H–VH	L	L	Yes	Yes	No
Horizontal time monitor	Single-small	Child-elderly	M–H	M	H–VH	L–M	L–M	?	Yes	No
Pedometers	Single-large	Child-elderly	L–M	L	L	L	L–M	Yes	Yes	No
Gait assessment	Single-small	Child-elderly	H–VH	M–VH	L–M	M–H	L–M	?	Yes	No
Electronic motion sensor	Single-large	Child-elderly	M–H	L	L	L	L–M	Yes	Yes	No
Accelerometers	Single-large	Infant-elderly	L–M	L–M	L	L	L–M	Yes	Yes	No
Dietary measures	Large	Adolescent-elderly	M–H	M	M–H	M–H	L	Yes	Yes	No

[1] L = low, M = moderate, H = high, VH = very high.

and is a more sedentary occupation as the result of containerization. Fourth, the pattern of job-related physical activity may have seasonal variability. Postal carriers expend much more energy when walking through winter snow than when walking during the summer months. Fifth, a selection bias may exist as persons with illnesses migrate to less strenuous jobs. Hence, illness may be the cause of occupational activity rather than vice versa. Finally, job classification accounts for only the occupational subcategory of physical activity and therefore fails to identify the importance of leisure-time physical activity or those activities performed by the unemployed or retired (65).

Survey procedures

Survey procedures seek to acquire information from the participants about their physical activity; they have four components (table 6). One component is the *timeframe* the respondents are asked to remember. The timeframe may be as short as 5 minutes or as long as a year or more. A second component is the *nature and detail* of the physical activities. Participants may be asked to provide the frequency, duration, and intensity of specific activities or they may merely be asked if they have performed an activity or a group of activities. A third component is the *mode of data collection.* Personal interview, telephone interview, self-administration, mail surveys, or combinations of these are common methods. The *summary index* may be based on a calculated estimate of kilocalories expended or an ordinal scale to rank-order individuals according to their level of physical activity.

Based on these four characteristics, surveys may be grouped into four general types. Surveys that use short timeframes (e.g., less than 24-hour intervals) and are self-administered are commonly referred to as physical activity *diaries.* Surveys that obtain information about the past 1–7 days via personal or telephone interview or via mail questionnaire are called physical activity *recall* surveys. Methods that are similar to the recall methods yet inquire about physical activities performed over a longer time period, usually the past year, are often called *quantitative history* procedures. Surveys that solicit little specific information about the nature and detail of the physical activities are referred to as *general* surveys, regardless of the timeframe of reference.

Diaries

Although the diary technique has been seldom used in epidemiologic investigations of physical activity, it has been used in energy balance studies. For example, Edholm and colleagues used the results of a physical activity diary to estimate total daily caloric expenditure (66). The procedure they used consisted of having individuals perform common tasks with concurrent measurement of energy expenditure via indirect calorimetry. Thereafter, individuals completed an ongoing diary, entering the specific tasks performed throughout the day. By summating the product of time spent in each task by the previously measured rate of energy expenditure for each activity, an overall estimate of total daily caloric expenditure was made. A direct comparison of the results of the ongoing diary with other estimates of total daily energy expenditure, measured either through indirect calorimetry or through caloric intake, revealed

Table 6. Characteristics of various physical activity survey procedures used in epidemiologic investigations

Type of general survey	Timeframe of survey	Nature and detail of activity	Physical activity survey time unit/level of detail	Mode [1] of collection	Summary index of physical activity
Shapiro	1 day	General work and leisure	General statements of participation level	MQ, SAQ	4 classes for work and leisure
Framingham	1 day	4 intensity classes for work and leisure	Hrs/day in each class	PI	Ordinal scale, 24 or more
Salonen	1 week	4 intensity levels for work and leisure	General statements of level of intensity	SAQ, PI	Assigned 1 of 4 levels for work and leisure
Magnus	1 year	Walk, cycle, garden activity (WCG); more strenuous than	Hrs/wk, mon/yr	PI	WCG groupings—occasional seasonal, habitual; at 1–4, 4–7, or 7+ hrs/wk
	Past 2 years	WCG	Yes or no		
Baecke	1 year	Work, leisure sports and nonsports; sports participation walking and cycling	5-point scale for each question; hrs/wk, mon/yr in sports; min/day recreational	MQ, SAQ	Ordinal scale
Morrison	1 week	8 leisure activity intensity classes	0, 1, 2, 3, 4+ times/wk	PI	Assigned 1 of 8 classes of leisure activity
Lipid Research Clinics	1 week	General work and leisure	2 general questions of strenuous and regular labor or exercise	PI	Classed inactive, moderately, active, or very active

[1] PI = personal interview, MQ = mail questionnaire, SAQ = self-assessment questionnaire.

that diaries could be very accurate indexes of daily energy expenditure (67). However, although the technique is very accurate, it suffers from cost, time, and acceptability constraints. Clearly, in a national survey, it would be impossible to collect activity data using the diary method. In addition, persons may be unwilling to record every physical activity they do throughout the day, or they may alter their normal pattern of physical activity to simplify the recording process. Each of these limitations makes this procedure of little value for health surveys.

Modifications of Edholm's technique have been devised to eliminate the need for actual assessments of task-specific energy expenditure. These modified techniques make use of tabled values available in sources such as Passmore and Durnin (68) and use the method of scoring kilocalories as employed in Edholm's study (69). It should be noted, however, that using tabled values may produce variability in estimating energy expenditure, as there may be relatively large differences among individuals for certain tasks. Moreover, the tables typically have been produced through the assessment of physical activity for young college-aged males, who are often atypical of the populations generally measured in epidemiologic investigations of chronic diseases, such as women, black persons, children, and obese individuals. Although diary procedures using tabled values of activity intensity may not provide accurate estimates of caloric expenditure, they seem to be adequate to rank-order individuals according to overall activity levels (70).

Another modification of the Edholm diary technique had individuals record the physical activities they had performed during 4-hour periods ending at 12 noon, 4 p.m., and 8 p.m. rather than record continuously throughout the day (71). Such a method may make the diary procedure less arduous; however, the likelihood of forgetting important details of physical activity participation may occur.

Recall

The diary technique of recording physical activity participation throughout the day is relatively difficult to employ on a population basis owing to the high demands placed on the participant. More often epidemiologic research of physical activity has used the recall procedure. The recall procedure used by Morris (72) was derived from a questionnaire-interview procedure developed by Yasin (73, 74). The advantage of the survey is that is does appear to be prospectively related to disease.

Bouchard et al. employed another daily recall in a study of 150 men and 150 women aged 10–50 years (75). The subjects provided a 15-minute by 15-minute recall for 3 days (2 weekdays, 1 weekend day). The activity for each segment was to be scored from 1 to 9 using tabled values for oxygen consumption and was thereafter converted to a kilocalorie score. Repeat administration yielded a reliability correlation of $r = 0.96$. The highest intensity category was most reliably repeated. In addition, Bouchard indirectly validated this recall against physical work capacity on a bicycle ergometer and with body fat estimates and found a small but significant relationship (75). This has not, however, been employed in studies to relate activity to disease.

Another recall survey by Paffenbarger probed for the distance walked, stairs climbed, and sports or recreational activities during the previous week (76). Estimates of activity have been shown to directly relate to risk of disease. The Paffenbarger questionnaire is simple to implement via mail.

A somewhat different recall procedure developed at Stanford has promise in the assessment of physical activity (77). In this technique, individuals recall the time spent over the previous 7 days doing activities within several levels of intensity rather than providing detailed time estimates for specific activities. Thus, the measurements are thought to provide a more accurate assessment of intensity-related issues. This 7-day recall has been indirectly validated using results from a community health survey, a randomized clinical trial, and two worksite health promotion programs, yielding expected relationships with cross-sectional physical activity patterns of men and women; changes in maximal oxygen consumption, body fatness, HDL–C and triglycerides; as well as increased physical activity over time because of intervention efforts. A short-term repeatability of $r = 0.67$ has been established (78). However, it has not been used in studies linking activity to heart disease. Future research employing this technique is likely to prove valuable as a physical activity assessment tool.

Quantitative history

The quantitative physical activity histories have timeframes greater than 1 week and request rather detailed information on specific activities. In the Taylor and Montoye questionnaires, individuals recall their activity patterns over the previous year for a list of specific physical activities (79, 80). The Montoye questionnaire has been indirectly validated by comparing results of the objective summary kilocalorie index with subjectively ranked estimates of physical activity level made by an independent judge, yielding high correlations of 0.86 to 0.97. Further, the questionnaire was compared with rank ordering of caloric intake, which differentiated the 20 percent least and 20 percent most active. The middle 60 percent were less well discriminated from the 20 percent most active. Further, indirect validation showed relationships with body fat estimates and blood lipids (81). The Taylor questionnaire has been indirectly validated against treadmill work performance in 175 men and with estimated caloric intake in a comparison of female college swimmers and college students (79). Further, relationships have been found between results of the Taylor questionnaire and HDL–C, body mass index, and heart rate (81). Short-term (roughly 2 weeks) repeatability of the Taylor questionnaire has recently been examined in a group of middle-aged men having previous experience with the questionnaire. Results suggest that the Taylor questionnaire has a correlation between first and second report of $r = 0.85$ for total leisure and of 0.73–0.79 for activity groupings of light, moderate, and heavy intensity (unpublished observations, C. J. Caspersen, 1983). Continued breakdowns into activity class categories (sports, conditioning, lawn and garden or home repair, etc.) or the individual activities themselves showed lower repeatability, as there are a greater number of parameters to correctly remember. Overall, the quantitative history surveys can be implemented on a population basis, yielding enormous detail on the physical activity pattern. A problem with these surveys is that they take a relatively long time to complete and probably could not be done in uneducated or unmotivated populations and therefore may be inappropriate in a battery of surveys. There have been no publications linking activity as measured through these surveys to disease.

General

Another type of assessment that has been used is the general physical activity survey. Surveys that fit into this category provide less detail than the other techniques. In fact, individuals may merely provide a subjective impression of their usual physical activity or otherwise may simply select one of several descriptive classes.

One example of a general survey was developed by Shapiro and colleagues for the New York Health Insurance Plan and had separate sections for work and leisuretime physical activity (82). The questionnaire was mailed to plan members and could be easily self-administered in about 5–10 minutes. Scored results of the survey could range from 1 to 28 for work-related physical activity and from 0 to 10 for leisure-time physical activity. Cutoff points were determined such that both sets of scores could be divided into four classes. In an effort to determine the consistency of the self-report, a followup interview of 38 persons was conducted. Despite the survey's simplicity, these persons were repeatedly placed in the same work, leisure, and total work and leisure classes only 55 to 61 percent of the time, suggesting a somewhat low reliability. However, the survey has been strongly related to risk of disease. Currently, there has been no reported attempt at validation of this questionnaire.

The Framingham Study also employed a general physical activity survey for personal interview that queried the hours per day spent in sleep, work, and extracurricular activities (83). For the last two categories, the hours spent in sedentary, slight, moderate, and heavy activity were probed for and multiplied by weighting factors of 1.1, 1.5, 2.4, and 5.0, respectively. Sleep hours were multiplied by a weighting factor of 1.0. The weighting factors were chosen to parallel the increased rate of oxygen consumption associated with increasingly more intense physical activity. The lowest possible summary score of this survey was 24 (24 hours of sleep). Although this survey was capable of rank ordering individuals by physical activity level, there are no published reports regarding the repeatability or validity of this index. Despite the crudity of this method, assessments of activity were related to coronary heart disease risk.

Salonen and others of the North Karelia Study employed four alternative questions to separately index work and leisure physical activity (84). The work alternatives ranged from a category of work activity described as mostly sedentary to a category of regular, hard physical labor, such as digging, lifting, and carrying burdens daily. Similarly, the leisuretime physical activity alternatives ranged from a category of almost completely inactive to a category of regular hard physical training, such as running and soccer, several times per week. No repeatability of this survey has been reported. However, Grimby et al., using essentially the same questions, placed 641 men into groups of the two lowest and two highest categories for both work and leisure questions and found a significant relationship with predicted maximal oxygen consumption, thereby achieving a simple level of indirect validation (85). This survey has been shown to be related to risk of disease.

Another general survey, used in the Netherlands by Magnus and colleagues, asked questions to determine the total hours per week spent in walking, cycling, and gardening activities as a group (86). Further, three categories of yearly participation were queried for these combined activities: Habitual (more than 8 months), seasonal (4–8 months), and occasional (less than 4 months). Another question was used to ask about the time spent in low-level moving-about activities, ascribing sedentary status to those who performed 2 hours per day or less in such activities. Although this survey is quite useful in sorting out issues of seasonality of activity participation, there have been no published reports of the repeatability of this survey nor any reports of validation attempts, nor has it been related to end points.

Another general survey used in the Netherlands was developed by Baecke and colleagues (87). In its original form, the questionnaire was comprised of 29 items; however, after performing a principal-component analysis, 16 items representing three meaningful factors emerged: Physical activity at work, sports activity, and nonsports leisuretime physical activity. Questions relating to specific sports participation and occupation were open ended; the remaining questions were based on 5-point scales for subjective ranking. The scored questionnaire provided ordinal values ranging from a minimum of 16 for a sedentary person to approximately 70 for a person with considerable work, sports, and nonsports leisuretime physical activity. This self-administered questionnaire was mailed to 306 men and women aged 19–31 years and was followed by a debriefing at a clinic interview. Thereafter, the same questionnaire was readministered during a home interview. Reliability coefficients were acceptable: Between $r = 0.81$ and $r = 0.88$ for work and sports; $r = 0.74$ for leisure and nonsports. Indirect validation was performed by correlating results of the questionnaire with education, lean body mass, and subjective experience of workload (SEWL). SEWL was a Dutch series of 57 questions that could be answered either "yes" or "no" and queried perceived physical stress of many common tasks. It was found that, for both men and women, work activity was inversely associated to educational level; leisure nonsports activity was positively associated with educational level; and an inverse association was found between SEWL for both leisure sports and nonsports physical activity. For women, a positive association between leisure sports and educational level was found; for men an association was found between both work and leisure sports activity and lean body mass. In all, the questionnaire seemed to be both reliable and at least indirectly validated; however, it has not been evaluated in relationship to disease risk.

Morrison et al., of South Africa, recently developed a general physical activity survey in which individuals are assigned to one of eight classes of exercise training activities (88). Persons reporting no exercise training at all were assigned to class one. Persons reporting one, two, and three or more occasions per week of low-intensity training in activities such as golf, bowling, "keep-fit" activities, and yoga or 30 minutes of walking were assigned to classes two through four. Persons reporting one, two, three, and four or more times per week of high-intensity activities such as soccer and tennis or 30 minutes of walking were assigned to classes five through eight. There are no published reports regarding repeatability of this classification; however, an indirect validation was performed using estimated maximal oxygen consumption, yielding a correlation of $r = 0.67$. It has not been related to coronary heart disease.

In addition to the general surveys just described, several attempts at overall ratings of physical activity have been attempted (89). The Lipid Research Clinic's Study employed two questions regarding the regularity and perceived intensity of strenuous work or leisure activity. These simple questions appeared to accurately index individuals into three categories, which were in turn related to physiologic parameters. Similarly, the National Health Interview Survey had individuals simply rate their physical activity level with respect to persons their own age (90), also succeeding in rank-ordering populations. Still another type of overall assessment pertains to the assessment of functional ability. Surveys of disability, impaired movement, and activities of daily living are designed to assess physical activity levels that are below that of able-bodied populations. In these surveys, the ability of individuals to perform certain movements or daily activities, such as walking, climbing stairs, and bathing, is indexed. The assumption is that persons with the greatest impairment are also the least active. These surveys may be very important in indexing those persons at the lower end of the physical activity spectrum. However, details regarding the reliability, validity, and relationship to health are lacking.

The recall, general, and quantitative history surveys have been increasingly in use in epidemiologic investigations of physical activity because of their ease of implementation. Currently, they are the most practical means for measuring physical activity in large populations. Certain issues need to be considered, however, with these procedures. For example, the degree to which these instruments are accurate across populations is not known. (Plowing requires quite a different energy expenditure in Mexico than in Iowa.) Assessment of physical activity cross-culturally, such as in black populations compared with white populations or women compared with men, may not be a difference in activity but rather the perception of how to complete the surveys.

Further, each of the survey procedures has one significant limitation: The capacity of an individual to remember details of past physical activity participation. At present, few studies have addressed this issue; however, one study is particularly noteworthy. Baranowski et al. examined some of the limitations of a 24-hour recall in children by comparing results with direct observation throughout the day (91). Although the study queried only aerobic-type activity participation, the daily recall seemed to be improved when context-specific time periods (such as before school, during school, and after school) were probed for. This study clearly points to the need to study methods of improving memory when developing physical activity surveys.

It should be pointed out that, although short-term diary and recall procedures permit very detailed information regarding physical activity, there is a limitation on the representativeness of the physical activity behavior or the population itself. Generally, short-term physical activity surveys encompassing 1–7 days are more likely to miss the physical activity behavior not performed during other seasons. One method to overcome this limitation is to readminister the survey during other seasons of the year and thereafter pool the summary indexes. Unfortunately, individuals may be unwilling, or more importantly, unavailable to make repeated assessments. Further, repeated surveys require more effort by investigators. On the other hand, the quantitative physical activity history is capable of covering seasonal variation but does so at the expense of the capacity of the individual to remember over long periods of time. Further, as the demand on the subject to recount minute details increases, both ability and willingness to comply decrease.

As noted, the survey procedures each derive a summary index capable of rank ordering individuals by their level of physical activity. In other instances, it is desirable to identify participation in more specific activities or classes of activities. For example, the Stanford 7-day physical activity recall employs a series of questions relating to walking and aerobic activity participation, as each may be related to different health outcomes. Hence, this survey provides an overall kilocalorie index and at the same time specifically details involvement in common high- and low-level physical activity. Other surveys may need to employ similar questions to pinpoint more specific activity participation. In table 6, we summarize the various existing survey methods concerning physical activity.

Physiologic markers of physical activity

The fact that changes in the level of vigorous physical activity are known to influence cardiorespiratory endurance has led to the frequent use of maximum oxygen consumption to estimate physical activity. For nonpopulation samples, the correlation between reported physical activity level and physical work capacity is modest (21, 22); however, for large, population-based studies, the relationship is very weak to nonexistent. As indicated earlier, cardiovascular fitness appears to be primarily genetically determined; therefore it is unlikely that $\dot{V}O_2$max can be a good assessment of activity.

Another physiologic approach, the doubly labeled water (DLW) technique, appears to provide an integrated measure of energy expenditure over time. In this method, subjects ingest water containing isotopically labeled hydrogen and oxygen atoms. Measurement of the relative proportions of unmetabolized water and water that has been incorporated into the energy cycle can provide an overall estimate of energy expenditure (92). Originally this method was economical only in studies of small animals. However, declining costs for the isotopes along with technological advances in isotopic mass spectrometers have made it economically feasible to apply this procedure in studies involving humans. The method appears to be quite accurate, with error estimates ranging from 2 to 10 percent when compared with calories from weighed dietary intake (93–95).

In addition, the DLW technique can be done with any aged individual, does not restrict free-living physical activity, requires minimal cooperation by the subject, is generally acceptable to the subject, and takes a minimum of 2–3 days but can extend through several weeks. Although this technique has had encouraging results, the cost for the isotopes is still about $225 per person, making it prohibitive for large population studies but perhaps most useful for validation studies. In addition, it provides only an overall kilocalorie index that is incapable of identifying specific types or patterns of physical activity participation that may be very important to assess.

A number of mechanical and electronic instruments have been developed to assess body movement or heart rate (HR) responses to physical activity. Although some have been around for many years, technological advances continue to make them viable methods of objective assessment.

Heart rate monitoring

Recent advances in HR monitoring have made it feasible to obtain continuous heart rate recordings over an extended period of time. Several reports in normal populations have assessed physical activity through the use of HR monitoring approaches (96–99). HR monitoring is attractive because it provides for the direct measurement of a physiologic parameter known to be related to physiologic activity and it can provide a continuous record that may reflect both intensity and duration of daily activity.

To estimate energy expenditure, one must assume a close linear relationship between HR and oxygen consumption ($\dot{V}O_2$). Heart rate and $\dot{V}O_2$ are typically measured for each subject over a range of work rates on a cycle ergometer or treadmill, and individual HR/$\dot{V}O_2$ regression curves are developed (100).

These initial laboratory measures are time consuming and expensive and may eliminate many subjects, thereby biasing the sample; thus, this technique is impractical for most epidemiologic research. In addition, the assumption of a linear HR/$\dot{V}O_2$ relationship has also been shown to be affected by the amount of muscle mass involved in the activity (100), the type of muscular contraction (101), environmental temperature (102), state of physical training (103), fatigue (104), and emotional stress (105).

Given the potential number of confounding factors affecting HR, it is not surprising that in the few available reports in which the daily energy expenditure has been assessed using the heart rate method, the results are rather poor (106). It may be possible to avoid the problem of subject calibration and obtain information on physical activity from HR monitoring by considering the difference between resting HR and mean daily HR (107). This would make the use of HR monitoring more practical for epidemiologic research and warrants further evaluation.

Motion sensors

The assessment of physical activity by measuring "movement" is appealing because more active people typically move more than less active people. Also, movement may measure physical activity more accurately than estimates of energy expenditure. Clearly many factors influence energy expenditure and the physiologic measurement of energy expenditure, such as specific dynamic action, basal metabolic rate, and perhaps body weight, ambient temperature, and age (108, 109). Movement per se, however, may itself be very complex and may require indepth analysis evaluating the type, frequency, and intensity to determine its physiologic effects.

Research in the 1960's and 1970's developed several excellent approaches for measuring movement in infants using stabilometers on mattresses; continuous 24-hour recording could be easily done (110–112). The measures appeared to be relatively accurate assessments of activity in infants but of no utility for children or adults.

Another instrument, affixed to the subject's thigh, has been used to compare the amount of time the individual was in a seated or lying position rather than standing or moving about (113). There is no information concerning either the reliability of the instruments or the practicality of using the instruments in large populations.

The primary type of measurement of body movement has been through the use of pedometers (114). Pedometers are instruments designed specifically to evaluate walking behavior. Several different types are available commercially. Reports evaluating the accuracy of the instruments have identified interinstrument and intrainstrument variability primarily associated with the mechanical fulcrum of the instruments, each responding inconsistently to a given force (115, 116). There is little information concerning the applicability of pedometers across diverse populations and whether the instruments can index individual and group activity levels.

Instruments from gait research have not as yet been employed in population studies but may afford a major increase in accuracy for assessing walking behavior. These instruments fit into the shoe and measure not only the step frequency, as with a pedometer, but also the force applied with each step (117). Marsden and Montgomery (118) found that in-shoe step counters could easily discriminate activity between postal carriers and schoolchildren, as well as different activity patterns within individuals. There has been little followup on the use of in-shoe step counters in population studies, and their reliability and validity are not established. These instruments have not been used in population studies but may prove to be valuable in epidemiologic studies of osteoporotic women, for whom weight-bearing activity appears to be a critical dimension in bone strength (119).

The physical activity monitors that have received the most interest are *electronic motion sensors*, now commercially available (120–124). Electronic motion sensors can be classified into instruments that assess only the quantity of movement and instruments that assess both the quantity and intensity of movement (accelerometers).

The large-scale integrated motor activity monitor (LSI) is an instrument about the size of a pocket watch. It measures body movements and can be placed on various body locations. Within the monitor is a cylinder with a ball of mercury. A 3-percent inclination or declination of instruments causes the closing of a mercury switch that registers in an internal counter. The results of the counter become available by holding a magnet to the side of the instrument, activating a light-emitting diode (LED) display indicating the number of movements (124).

LSI units have been shown to have low interunit variability (124). Hence, one can safely conclude that one LSI is measuring the same as another unit, and the instruments appear to measure correctly across time. Moreover, the LSI units have been able to discriminate differences between varying population groups (125).

Accelerometers, which measure both frequency and intensity of movement, may prove even more helpful and have recently been introduced (126). Future methodologic studies need to be completed to evaluate the effectiveness of this instrument for population use.

The mechanical and electronic monitors of heart rate and movement may be very useful in assessing physical activity, at least in small groups. Their appeal is that they require very little time on the part of the investigator as well as limited effort on the part of the subject. Further, many do not interfere with or influence physical activity and seemed to be personally and socially acceptable. Currently, the cost of these devices make their use in large population studies unlikely; they provide information on specific categories or types of physical activity, but they do not provide an estimate of energy expenditure. As technology improves and the costs decrease, these monitors may be made applicable to population studies of physical activity.

Dietary measures

Food intake may be used as an estimate of energy expenditure and hence of physical activity if one assumes that energy balance has been achieved with stable body weight. Edholm and colleagues have demonstrated that long-term assessment of food intake, measuring total amount consumed with analysis of caloric content either via bomb calorimetry of equivalent samples or via tabled values, provides a very accurate assessment of caloric intake and hence of caloric expenditure for an individual (66).

However, it is well known that total caloric intake is influenced by the individual's physical activity level and also by the total body weight of the subject. For example, two persons having markedly different body weight may have the same total daily caloric expenditure, with the lighter person doing more physical activity than the heavier person. Further, using weighed dietary intake as a measure of physical activity is very costly and requires great demands on the subject.

To address these issues, Sopko et al. recently analyzed data from two carefully controlled feeding experiments, one of obese and the other of normal weight subjects, using precisely weighed dietary samples for total daily caloric intake (127). After adjusting the total caloric intake by the individual's body weight in kilograms (kcal/kg/day), they were able to show a strong negative correlation with body fat ($r = -0.79$) and an equally strong positive association with maximal oxygen consumption ($r = 0.76$), each serving as an indirect validation of physical activity level. In addition, they employed a 3-day food recall to estimate daily caloric intake for their obese subjects, and when they compared kcal/kg/day with maximal oxygen consumption and body fat, weaker yet still significant correlations of $r = -0.54$ and $r = 0.31$, respectively, were noted.

It is important to recognize that estimates of dietary intake are also known to have considerable variability, as evidenced by the smaller correlations found by Sopko et al. when the 3-day food recall was used. Further, Beaton et al. have demonstrated that total caloric intake estimated via 24-hour dietary recall has considerable intraindividual and interindividual variability (128) and may therefore be inadequate for indexing an individual's physical activity level within a group. Hence, dietary measures may have to improve considerably to be a useful and practical index of physical activity. Further, dietary measures of physical activity are unable to identify the types, frequency, intensity, or duration of physical activities performed.

Relationships among measures

Researchers have begun to evaluate the interrelationships among measurement techniques to determine the degree to which each technique measures the same thing. For example, Buskirk et al. compared results of the Health Insurance Program and Montoye questionnaires and noted significant correlations of relatively low magnitude (129). This indicated that the two instruments, each attempting to measure work and leisure-time physical activity, were less than congruent. In addition, the LSI motion sensor has been compared with various survey techniques in free-living populations, having correlations of 0.70 with diaries in several populations; however, correlations with recall procedures were much smaller (table 7).

Further, it has recently been suggested that the Paffenbarger survey appears to be related to HDL_3-cholesterol, whereas activity monitoring appears to be associated with HDL_2-cholesterol and bone density (119, 130). Thus, the component of physical activity measured by these two instruments may not be the same. Moreover, each may be associated with a different aspect of health.

Discussion

It is very important to employ accurate assessments of physical activity in the National Health and Nutrition Examination Survey (NHANES). As indicated in the review, one of the major difficulties has been the defining of what is meant by physical activity. Physical activity is not a single entity such as overall caloric expenditure. Defining activity as overall caloric expenditure is like defining diet as caloric intake. The current conceptualization is that physical activity is comprised of numerous interrelated but not perfectly correlated dimensions. Some of the possible dimensions are:

- Overall caloric expenditure.
- Weight bearing.
- Aerobic.
- Strength.
- Flexibility.

Each one of the dimensions may have different relationships to health; therefore the dimensions should be measured and analyzed separately. For example, people who leisurely swim 2 hours per day may have a high caloric output; however, the intensity might not be sufficient to produce a major increase in fitness, nor is the activity weight bearing. Alternatively, a person who jogs three times per week for 15 minutes per session may not have a major increase in overall caloric expenditure; however, the fitness levels may increase. Therefore, it is important to employ the surveys to identify the various dimensions of activity.

The question arises, therefore, as to what survey or combination of surveys is most appropriate for assessing the activity dimensions. We are, of course, constrained because of the limited time available for assessing activity as a part of NHANES; however, it is likely that the time spent on the indepth assessment of activity will be as important as, if not more important than, the fitness evaluations.

The criteria of surveys should be that they: (1) Enable researchers to index individual and group activity levels on the spectrum of activity, (2) allow evaluation of the various dimensions of activity, (3) be applicable across the total activity spectrum from disabled through extremely active groups, and (4) be practical for administration in populations differing in educational as well as socioeconomic states. In addition, it would be

Table 7. Correlations between activity monitor readings and surveys

Activity monitor	Survey	Population	Pearson correlation
Total activity	Daily logging	20 male college students	.69 $p < .01$
Total activity	Taylor	20 male college students	.05 $p > .05$
Total activity	Taylor	20 male 8th graders	.02 $p > .05$
Total activity (workday)	Taylor	42 industrial workers	.11 $p > .05$
Total activity (offday)	Taylor	42 industrial workers	.45 $p > .05$
Total activity	Paffenbarger	79 menopausal females	.23 $p < .05$
Day activity	Paffenbarger	256 menopausal females	.21 $p < .05$
Total activity	Paffenbarger	256 menopausal females	.13 $p < .05$

ideal if the surveys had been previously shown to be related to risk of disease or at least risk factors for disease. To our knowledge, there are no surveys that fulfill all these criteria.

The recommended battery of tests is:

- Health Insurance Plan of Greater New York Survey Questionnaire.
- Morris Walking Survey Questionnaire.
- Stanford Assessment of Aerobic Activity.
- Activities of daily living.

The primary survey questionnaire will be that employed in the Health Insurance Plan of Greater New York. The advantages of using this questionnaire are that it has been associated with risk of disease, it allows assessment of both leisuretime and occupational activity, it has been employed in groups having vast differences in educational and occupational status, and it can be administered very rapidly. The difficulties of the survey are that results provide only a general overall picture of activity, no attempt has been made to identify specific components of activity, and the limited evaluation of reliability indicates that there appear to be some difficulties in indexing activity levels. The survey therefore needs to be supplemented. One approach would be to include questions from the Stanford survey questionnaire that are designed to assess aerobic activities. In addition, we as well as others are finding that the major component of physical activity/weight-bearing activity is walking. Therefore, a more extensive assessment of walking behavior, such as is collected by Morris, should also be included. A problem with the battery of questions is that it is not of sufficient specificity to index activity levels for individuals who are at the low end of the activity spectrum. Therefore, in addition to these three survey questionnaires, a modified version of the activities of daily living survey that was developed for the elderly supplement of the National Health Interview Survey would be employed. The assessment of the ability of the participants to perform activities of daily living is likely to be a very accurate assessment of low levels of physical activity.

This battery of survey questionnaires will take approximately 15–20 minutes to administer. It could be administered through interview or through self-administration. At this point, objective monitoring of physical activity in NHANES is not likely to be very practical. It would be ideal if a biochemical marker of activity could be employed; however, techniques such as doubly labeled water, although quite promising, have not been sufficiently developed for use in population studies.

It is important to once again repeat that a functional assessment of cardiovascular fitness is a poor index of physical activity because, on a population basis, there is only a relatively small association between activity and fitness. Therefore, the most practical approach will be the assessment of activity through survey methodology.

Part of the charge of the current paper was to discuss the approaches for assessment of physical fitness. From an epidemiologic perspective, however, the data suggest that the focus of the assessment should be on physical activity rather than physical fitness.

As indicated in the current review of fitness, activity, and health and the assessment of activity, there is very little evidence that fitness as measured by VO_2max is associated with disease. Moreover, measurements of cardiovascular fitness are not strongly related to measures of activity because of the very strong genetic loading associated with fitness. Therefore, if the focus of a health survey is to assess the relationship of the activity-fitness parameters to health, then the primary focus should be on the assessment of activity. Clearly, if health is the primary interest, one can even question the need of assessing maximum oxygen consumption. The time needed to collect valid data on maximum oxygen consumption could perhaps be better spent in collecting factors more associated with health. However, cardiovascular fitness is strongly related to athletic performance. If one of the goals of NHANES is to assess the athletic performance of the population, then clearly factors such as maximum oxygen consumption should be collected. It is therefore recommended that, if assessment of maximum oxygen is completed, the time should be minimized and more time spent on the assessment of activity.

References

1. Froelicher VF, Brown P: Exercise and coronary heart disease. *J Cardiovasc Rehab* 1:277–288, 1981.
2. Leon AS, Blackburn H: The relationship of physical activity to coronary heart disease and life expectancy. *Ann NY Acad Sci* 301:561–578, 1977.
3. Wyndham CH: The role of physical activity in the prevention of ischaemic heart disease: A review. *S Afr Med J* 56:7–13, 1979.
4. Epstein LH, Wing RR: Aerobic exercise and weight. *Addict Behav* 5(4):371–388, 1980.
5. Editorial: Osteoporosis and activity. *Lancet* 1:1365–1366, 1983.
6. King H, Taylor RR, Zimmet P, et al.: Noninsulin-dependent diabetes (NIDDM) in a newly independent Pacific nation: The Republic of Kiribati. *Diabetes Care* 7:409–415, 1984.
7. LaPorte RE, Adams LL, Savage DD, et al.: The spectrum of physical activity, cardiovascular disease, and health: An epidemiologic perspective. *Am J Epidemiol* 120:507–517, 1984.
8. Caspersen CJ, Powell KE, Christenson GM: Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports*, 1985.
9. American College of Sports Medicine position statement: The recommended quality and quantity of exercise for developing and maintaining fitness in healthy adults. *Sports Med Bull* 13:3–4, 1978.
10. Disability Survey 1972: Department of Health and Human Services. Research Report No. 56, 1981.
11. Wilhelmsen L, Tibblin G, Aurell M, et al.: Physical activity, physical fitness, and risk of myocardial infarction. *Adv Cardiol* 18(0):217–230, 1976.
12. Wilhelmsen L, Bjure J, Ekstrom-Jodal B, et al.: Nine years' follow-up of a maximal exercise test in a random population sample of middle-aged men. *Cardiology* 68(2):1–8, 1981.
13. Schroll M: A ten-year prospective study, 1964–1974, of cardiovascular risk factors in men and women from the Golsthrup population born in 1914. *Dan Med Bull* 29:213–251, 1982.

14. Peters RK, Cady LD, Bischoff DP, et al.: Physical fitness and subsequent myocardial infarction in healthy workers. *JAMA* 249:3052–3056, 1983.
15. Gyntelberg F, Lauridsen L, Schubell K: Physical fitness and risk of myocardial infarction in Copenhagen males aged 40–59: A five- and seven-year follow-up study. *Scand J Work Environ Health* 6:170–178, 1980.
16. May GS, Eberlein LA, Furberg CD, et al.: Secondary prevention after myocardial infarction: A review of long term trials. *Prog Cardio Dis* 24:331–352, 1982.
17. Morris JN, Everitt MG, Pollard R, et al.: Vigorous exercise in leisure-time: Protection against coronary heart disease. *Lancet* 2(8206):1207–1210, 1980.
18. Haskell WL: Habitual exercise and ischemic heart disease: How much might be of value. (Unpublished manuscript.)
19. Morris JN, Everitt MG, Pollard R, Chave SPW, Semmence AM: Exercise and the heart. *Lancet* 1:267, 1981.
20. Chave SPW, Morris JN, Moss S, et al.: Vigorous exercise in leisure time and the death rate: A study of male civil servants. *J Epidemiol Community Health* 32:239–243, 1978.
21. Skinner JS, Benson H, McDonough JR: Social status and coronary proneness. *J Chronic Dis* 19:773–783, 1966.
22. DeBacker G, Korintzer M, Soboloski K: Physical activity and physical fitness levels of Belgian males aged 40–59. *Scand J Work Environ Health* 6:170–178, 1980.
23. Goldsmith R, Heile R: Relationship between physical activity and physical fitness. *Am J Clin Nutr* 24:1489–1493, 1971.
24. Astrand I, Astrand PO, Hallback I, et al.: Reduction in maximal oxygen consumption with age. *J Appl Physiol* 35:649, 1973.
25. Paffenbarger RS Jr., Wing AL, Hyde RT: Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol* 108:161–175, 1978.
26. Cauley JA, et al.: Physical activities and HDL–C subfractions in post-menopausal women. *Med Sci Sports Exerc*, 1984. (Abstracts)
27. Blomquist CG: Cardiovascular adaptations to physical training. *Ann Rev Physiol* 45:169–189, 1983.
28. Paffenbarger RS Jr., Hyde RT, Wing AL, et al.: A natural history of athletics and cardiovascular health. *JAMA* 252:491–495, 1984.
29. Dyer AR, Persky H, Stamler J, et al.: Heart rate as a prognostic factor for coronary heart disease and mortality: Findings in three Chicago epidemiologic studies. *Am J Epidemiol* 112:736–749, 1980.
30. Schroll M, Hagerup LM: Risk factors for myocardial infarction and death in men aged 50 at entry. *Dan Med Bull* 24:252–255, 1977.
31. Tibblin G, Wilhelmsen L, Werko LL: Risk factors for myocardial infarction and death due to ischemic heart disease and other causes. *Am J Cardiol* 35:514–522, 1975.
32. Keys A, Taylor HL, Blackburn H, et al.: Mortality and coronary heart disease among men studied for 23 years. *Arch Intern Med* 128:201–214, 1971.
33. Friedman GD, Klatsky AL, Siegelaub AB: Predictors of sudden cardiac death. *Circulation* 52(3):164–169, 1975.
34. Goldbourt U, Medialie JG, Neufeld HN: Clinical myocardial infarction over a five-year period—III: A multivariate analysis of incidence, the Israeli ischemic heart disease. *J Chronic Dis* 28:217–237, 1975.
35. Miller GJ, Miller NE: Plasma high-density lipoprotein concentrations and development of ischemic heart disease. *Lancet* 1(7897):16–19, 1975.
36. Gordon T, Castelli WP, Jhortland MC, et al.: High-density lipoprotein as a protective factor against coronary heart disease: The Framingham study. *Am J Med* 62:707–714, 1977.
37. Wood PD, Williams PT, Haskell WL: Physical activity and high-density lipoproteins. In, Miller NE, Miller GJ (eds). *Clinical and Metabolic Aspects of High-Density Lipoproteins.* Elsevier Press, 1984, pp. 133–164.
38. Williams PT, Wood PD, Haskell W, et al.: The effect of running mileage and duration on plasma lipoprotein levels. *JAMA* 247:2674–2679, 1982.
39. Haskell WL, Taylor HL, Wood PD, et al.: Strenuous physical activity, treadmill exercise test performance, and plasma high density lipoprotein cholesterol. *Circulation* 62(4):53–61, 1980.
40. Taeuber CM: America in transition: An aging society. Series P–23, No. 128, U.S. Bureau of the Census: *Current Population Reports* series. U.S. Government Printing Office, 1983.
41. Kelsey J: Osteoporosis: Prevalence and incidence. Presented at NIH Consensus Development Conference, April 2–4, 1984, Bethesda, MD.
42. Montoye HJ: Physical activity and health: An epidemiologic study of an entire community. *Res Mon Series in Physical Education.* Englewood Cliffs, NJ, 1975, p. 20.
43. Worley RJ: Age, estrogen, and bone density. *Clin Obstet Gynecol* Vol. 24, 1981.
44. Chalmers J, Ho KC: Geographical variations in senile osteoporosis: The association of physical activity. *J Bone Joint Surg* 52:667–675, 1970.
45. Donaldson CL, Hulley SB, Vogel JM, Hattner RS, Bayers JH, McMillan E: Effect of prolonged bed rest on bone mineral. *Metabolism* 19:1071–1084, 1970.
46. Stein H, Sabato S, Leichter I, Nir D, Weinreb A: A new method of measuring bone density in the lower tibia of normal and post-injury limbs: A quantitative and comparative study. *Clin Ortho* 174:181–187, 1983.
47. Krolner B, Toft B: Vertebral bone loss: An unheeded side effect of therapeutic bed rest. *Clin Sci* 64:537–540, 1983.
48. Hansson TH, Roos BO, Nachenson A: Development of osteoporosis in the fourth lumbar vertebra during prolonged bed rest after operation for scoliosis. *Acta Orthop Scand* 46:621–630, 1975.
49. Whedon DG, Shorr E: Metabolic studies in paralytic acute anterior poliomyelitis. *J Clin Invest* 36:966–981, 1957.
50. Freedman LW: The metabolism of calcium in patients with spinal cord injury. *Am Surg* 129:177–184, 1949.
51. Stevenson FH: The osteoporosis of immobilization in recumbency. *J Bone Joint Surg* 34B:256–265, 1952.
52. Hattner RS, McMillian DE: Influence of weightlessness upon the skeleton: A review. *Aero Med* 39:849–855, 1968.
53. Mazess RB, Whedon GD: Immobilization and bone. *Calcif Tissue Int* 35:265–267, 1983.
54. Whedon GD, Dietrick JE, Shorr E: Modifications of the effects of immobilization upon metabolic and physiologic functions of normal men by use of an oscillating bed. *Am J Med* 5:684, 1949.
55. Korcok M: Add exercise to calcium in osteoporosis presentation. *JAMA* 247:1106–1107, 1982.
56. Aloia JF, Cohn SH, Babu T, Abesauris C, Kaliei N, Ellis K: Skeletal mass and body composition in marathon runners. *Metabolism* 12:1783–1796, 1978.
57. Krolner B, Toft B, Nielsen SP, Tandewold E: Physical exercise as a prophylaxis against involuntary bone loss: Controlled trial. *Clin Sci* 64:541–546, 1983.
58. Aloia JF: Exercise skeletal health. *J Am Geriatr Soc* 29:104–107, 1981.
59. Smith EL, Reddan W, Smith PE: Physical activity and calcium moderately for bone mineral increase in aged women. *Med Sci Sports Exerc* 13:60–64, 1981.
60. Cann CE, Martin MC, Genant HK, Jaffe RB: Decreased spiral mineral content in amenorrheic women. *JAMA* 251:626–629, 1984.
61. Koplan JP, Powell KE, Sikes RK, Shuley RW, Campbell CC: An epidemiologic study of the benefits and risks of running. *JAMA* 248:3118–3121, 1982.
62. Doyle F, Brown J, Lachance C: Relation between bone mass and muscle weight. *Lancet* 1:391–393, 1970.
63. Sinaki M, Opitz JL, Wahner HW: Bone mineral content: Relationship to muscle strength in normal subjects. *Arch Phys Med Rehab* 55:508–512, 1974.
64. Nilsson BE, Westlin NE: Bone density in athletes. *Clin Orthop* 77:177–182, 1971.

65. LaPorte RE, Montoye HJ, Caspersen CJ: Assessment of physical activity in epidemiologic research: Problems and prospects. *Public Health Reports* (in press).
66. Edholm OC, Fletcher JG, Widdowson EM, McCance RA: The energy expenditure and food intake of individual man. *Br J Nutr* 9:286–300, 1955.
67. Acheson KJ, et al.: The measurement of daily energy expenditure—an evaluation of some techniques. *Am J Clin Nutr* 33:1155–1164, 1980.
68. Passmore R, Durnin JVGA: Human energy expenditure. *Physiol Rev* 35:801–840, 1955.
69. Durnin JVGA, Namyslowski L: Individual variations in energy expenditure of standardized activities. *J Physiol* 143:573–578, 1958.
70. Bouchard C: Reply to letter by Torun. *Am J Clin Nutr* 38(5):815, 1983.
71. LaPorte RE, et al.: An objective measure of physical activity for epidemiologic research. *Am J Epidemiol* 109(2):158–168, 1979.
72. Morris JN, Everitt MG, Pollard R, Chave SPW: Vigorous exercise in leisure-time: Protection against coronary heart disease. *Lancet* 2:1207–1210, 1980.
73. Yasin S: Measuring habitual leisure time physical activity by recall record questionnaire. In, Karvonen MJ, Barry AJ (eds). *Physical Activity and the Heart.* Springfield, IL: Charles C. Thomas Co., 1967, pp. 372–373.
74. Alderson M, Yasin S: Measuring habitual leisure time activity: A questionnaire method suitable for epidemiologic studies. In, Evang K, Lange-Andersen K (eds). *Physical Activity in Health and Disease.* Baltimore, MD: Williams and Wilkins Co., 1966, pp. 215–221.
75. Bouchard C, et al.: A method to assess energy expenditure in children and adults. *Am J Clin Nutr* 37:461–467, 1983.
76. Paffenbarger RS, Wing AL, Hyde RT: Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol* 108:161–175, 1978.
77. Blair SN: Validation of the Stanford physical activity recall. *Am J Epidemiol* (in press).
78. Sallis JF, et al.: Physical activity assessment methodology in the Stanford five-cities project. *Am J Epidemiol* (in press).
79. Taylor HL, et al.: A questionnaire for the assessment of leisure time physical activities. *J Chronic Dis* 31:741–755, 1978.
80. Montoye HJ: Estimation of habitual physical activity by questionnaire and interview. *Am J Clin Nutr* 24:1113–1118, 1971.
81. Folsom AR, et al.: Distribution of leisure time physical activity and its relationship to coronary risk factors in a metropolitan area: The Minnesota Heart Survey. *Am J Epidemiol* 121:570–579, 1985.
82. Shapiro S, Weinblatt E, Frank CW, Sager RV: The H.I.P. study of incidence and prognosis of coronary heart disease: Preliminary findings on incidence of myocardial infarction and angina. *J Chronic Dis* 18:527–558, 1965.
83. Kannel WB, Sorlie PS: Some health benefits of physical activity: The Framingham Study. *Arch Intern Med* 139:857–861, 1979.
84. Salonen JT, Puska P, Tuomilehto J: Physical activity and risk of myocardial infarction, cerebral stroke, and death. *Am J Epidemiol* 115(4):526–537, 1982.
85. Grimby G, et al.: Habitual physical activity: Aerobic power and blood lipids. In, Pernow ED, Saltin B (eds). *Muscle Metabolism During Exercise.* New York: Plenum Press, 1979, pp. 469–481.
86. Magnus K, Matroos A, Strackee J: Walking, cycling, or gardening, with or without seasonal interruption, in relation to acute coronary events. *Am J Epidemiol* 110:724–733, 1979.
87. Baecke JAH, Burema J, Frijters JER: A short questionnaire for the measurement of habitual physical activity in epidemiologic studies. *Am J Clin Nutr* 36:932–942, 1982.
88. Morrison JF, van Malsen S, Noakes TD: Leisure-time physical activity levels, cardiovascular fitness, and coronary risk factors in 1,015 white Zimbabweans. *S Afr Med J* 65:250–256, 1984.
89. Haskell WL, et al.: Strenuous physical activity, treadmill exercise test performance, and high density lipoprotein cholesterol. *Circulation* 62(4):53–61, 1980.
90. National Center for Health Statistics, Bloom B: Current estimates from the National Health Interview Survey, United States, 1981. *Vital and Health Statistics,* Series 10, No. 141. DHHS Publication No. (PHS) 83–1569. U.S. Government Printing Office, Washington, DC, October 1982.
91. Baranowski T, et al.: Validity of children's self-report of aerobic activity. *Fam Health Proj Res Q Exer Sport* (in press).
92. Schoeller DA, van Santen E: Measurement of energy expenditure in humans by doubly labeled water method. *J Appl Physiol* 53(4):955–959, 1982.
93. Klein PD, et al.: Calorimetric validation of the doubly-labeled water method for estimation of energy expenditure in man. *Hum Nutr Clin Nutr* 38:95–106, 1984.
94. Schoeller DA, Webb P: Five day comparison of the doubly labeled water method with respiratory gas exchange. *Am J Clin Nutr* 40:153–158, 1984.
95. Schoeller DA: Energy expenditure from doubly labeled water: Some fundamental considerations in humans. *Am J Clin Nutr* 38:999–1005, 1983.
96. Booyens J, Hervey GR: The pulse rate as a means of measuring metabolic rate in man. *Can J Biochem Physiol* 48:1301–1309, 1960.
97. Taylor CB, et al.: A new system for long-term recording and processing of heart rate and physical activity in outpatients. *Comput Biomed Res* 15:7–17, 1982.
98. Bradfield RB: A technique for determination of usual daily expenditure in the field. *Am J Clin Nutr* 24:1148–1154, 1971.
99. Warnold T, Lenner RA: Evaluation of the heart rate method to determine the daily energy expenditure in disease: A study of juvenile diabetics. *Am J Clin Nutr* 30:304–315, 1977.
100. Vokac Z, Bell H, Bautz-Holter H, Rodahl K: Oxygen uptake/heart rate relationship in leg and arm exercise, sitting, and standing. *J Appl Physiol* 39:54–59, 1975.
101. McCloskey D, Streatfield K: Muscular reflex stimuli to the cardiovascular system during isometric contraction of muscle groups of different mass. *J Physiol* 230:431–441, 1975.
102. Sengupta A, Saka D, Muklonadhyay S, Gosivamn P: Relationship between pulse rate and energy expenditure during graded work at different temperatures. *Ergonomics* 22:1207–1215, 1979.
103. Kappagoda C, Linden R, Newell J: Effects of the Canadian Air Force Training Program on submaximal exercise test. *Q J Exp Physiol* 64:185–204, 1979.
104. LeBlanc JA: Use of heart rate as an index of work output. *J Appl Physiol* 10:275–280, 1967.
105. Bateman S, Goldsmith R, Jackson K, et al.: Heart rate of training captains engaged in different activities. *Aerospace Med* 41:425–429, 1970.
106. Christensen CC, et al.: A critical evaluation of energy expenditure estimates based on individual O_2 consumption/heart rate curves and average daily heart rate. *Am J Clin Nutr* 37:468–472, 1983.
107. Washburn RA, Montoye HJ: Validity of heart rate as a measure of mean daily energy expenditure (abstract). *Med Sci Sports Exerc* 16:196–197, 1984.
108. Goldsmith R, Hale T: Relationship between habitual physical activity and physical fitness. *Am J Clin Nutr* 24:1489–1493, 1971.
109. Garrow JS: Energy balance and obesity in man. Elesvier/North Holland Publishing Biomedical Press. Amsterdam, Holland, 1974, p. 63.
110. Sander LW, Julia HL: Continuous interactional monitoring in the neonate. *Psychosom Med* 28(6):822–835, 1966.
111. Sander LW, Julia HL, Stechler G, Burns P: Continuous 24-hour interactional monitoring in infants reared in two caretaking environments. *Psychosom Med* 34(3):270–283, 1972.
112. Rose HE, Mayer J: Activity, caloric intake, fat storage, and energy balance of infants. *Pediatrics* 41:18–29, 1968.

113. Bloom WL, Eidex MF: Inactivity as a major factor in adult obesity. *Metabolism* 8:679, 1967.

114. Stunkard A: A method of studying physical activity in man. *Am J Clin Nutr* 8:595–601, 1960.

115. Gayle RH, Montoye HJ, Philpot J: Accuracy of pedometers for measuring distance walked. *Res Q* 48:632–636, 1977.

116. Washburn RA, Chin MK, Montoye HJ: Accuracy of pedometer in walking and running. *Res Q Exerc Sports* 51:695–702, 1980.

117. Saunders J, Inman V, Eberhart T: The major determinants in normal and pathological gait. *J Bone Joint Surg* 35A:543–558, 1953.

118. Marsden JP, Montgomery SR: A general survey of the walking habits of individuals. *Ergonomics* 15:429–441, 1972.

119. Black-Sandler R, et al.: Determinants of bone mass in the menopause. *Prev Med* 11:269–280, 1982.

120. Brugger W, Milner M: Computer-aided tracking of body movement using A.C.C.D. usage sensor. *Med Biol Eng Comput* 16:207–210, 1978.

121. Morris JRW: Accelerometry—A technique for measurement of human body movements. *J Biomech* 6:729–736, 1973.

122. Schulman JL, Stevens TM, Kupst MJ: The biomotometer: A new device for the measurement and remediation of hyperactivity. *Child Dev* 48:1152–1154, 1977.

123. Wong TC, Webster JG, Montoye HJ, Washburn RA: Portable accelerometer device for measuring human energy expenditure. IEEE Trans on Biomed Eng BME–28(6):467–471, 1981.

124. Foster FG, McPartland RJ, Kupfer DJ: Motion sensors in medicine: A report on reliability and validity. *J Inter-Amer Med* 3(1):4–8, 1978.

125. LaPorte RE, et al.: The epidemiology of physical activity in college students, middle-aged men, menopausal females, and monkeys. *J Chronic Dis* 35:787–795, 1982.

126. Montoye HJ, et al.: Estimation of energy expenditure by a portable accelerometer. *Med Sci Sports Exerc* 15(5):403–407, 1983.

127. Sopko G, Jacobs DR, Taylor H: Dietary measures of physical activity. *Am J Epidemiol* (in press).

128. Beaton GH, et al.: Source of variance in 24 hour dietary recall data: Implications for nutrition study design and interpretation—carbohydrate sources, vitamins, and minerals. *Am J Clin Nutr* 37:986–995, 1983.

129. Buskirk ER, Harris D, Mendez J, Skinner J: Comparison of two assessments of physical activity and a survey method for caloric intake. *Am J Clin Nutr* 24(9):1119–1125, 1971.

130. Cauley JA, LaPorte RE, Black-Sandler B: Measurement of physical activity in older populations. Paper presented at Society for Epidemiologic Research, Houston, TX, 1984.

Chapter 25

Cardiovascular Epidemiological Research Uses of Fitness Assessments

Erika S. Sivarajan, R.N., M.A., M.P.H.
University of California, Los Angeles

Victor F. Froelicher, M.D.
University of California, Irvine

Introduction

In 1980, the United States Public Health Service released a report titled, "Promoting Health/Preventing Disease: Objectives for the Nation." This document presented a series of goals and objectives formulated in each of 15 priority areas of concern for health promotion and disease prevention. These goals and objectives represent the consensus of more than 500 individuals and organizational representatives with expertise in the 15 areas. The areas addressed included: high blood pressure; family planning; pregnancy and infant health; immunization; sexually transmitted diseases; toxic agents and radiation; occupational safety and health; accidents and injuries; dental health; infectious diseases; smoking; alcohol and drugs; improved nutrition; physical fitness and exercise; and control of stress and violent behavior (see paper by Pearson).

For each of the 15 priority areas, the report briefly outlines the nature and extent of the specific problem, describes and evaluates available measures, and states national objectives. The general objectives are categorized as improved health status; reduced risk factors; increased public or professional awareness; improved services or protection; and improved surveillance or evaluation systems. This last component recognizes that interventions should be assessed for positive and negative effects by making measurements before and after their implementation. Such evaluation makes it possible to demonstrate whether the interventions are effective. Also, this last component was the directive to the Center for Health and Disease Statistics to include a means to quantify the activity status of Americans in their surveys.

In the establishment of the objectives in the areas of physical fitness and exercise, there were few data sets that provided accurate information on the existing fitness levels and exercise habits of the U.S. population. Although national estimates of the physical fitness status of children and youths have been collected for more than 25 years, there are no such data on adults. Without a representative data base, it would be impossible to assess changes in the population as a result of various intervention strategies. To fill this void, the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS) were requested to add a physical fitness and exercise component to their existing test batteries.

Fitness and exercise objectives

Specific objectives for the exercise and physical fitness priority area for 1990 included the following, listed by general objective:

- Improved health status:

  Though increased levels of physical fitness may contribute to reduced disease rates, no specific objectives could be developed. This is surprising because coronary heart disease has been closely and independently related to physical activity, and so a decrease in coronary disease mortality could be a marker of an increase in physical fitness. Physical activity has been shown to have a weak but consistent independent correlation with coronary heart disease mortality (1–3).
- Reduce cardiovascular risk factors (assuming that physical inactivity is a risk factor):

  The proportion of children and adolescents (ages 10–17) participating regularly in appropriate activities that can be carried into adulthood (sports other than baseball and football) should be greater than 90%.

  Greater than 60% should be participating in daily school physical activity programs.

  More than 60% of adults 18–65 should be participating regularly in moderate to vigorous exercise (in 1978, the estimate was 35%).

  50% of adults 65 years and older should be engaging in regular physical exercise (in 1975, about 36% took regular walks).
- Increased public awareness:

  More than 70% of adults should be able to identify the appropriate type of exercise needed.

  More than 50% of primary care doctors should include an exercise history as part of their initial examination which should include an occupational and recreational exercise history. These are described and their uses discussed by LaPorte (chapter 24).
- Improved services/protection:

  More than 25% of companies with more than 500 employees should offer sponsored physical fitness programs (in 1979, it was 2.5%).
- Improved surveillance and evaluation services:

  Methodology for assessing physical fitness of 70% of our children.

  Data should be available to evaluate the health effects (positive and negative) of exercise programs.

  Data should be available to evaluate effects on job performance and health care costs.

  Data should be available for monitoring national trends and patterns of participation in physical activity. This should include the availability, accessibility and use of parks and tennis courts. Surveys measuring the usage of exercise facilities could yield valuable information to aid in the achievement of the objectives for 1990.

Data necessary specifically to answer the objectives

If the primary goal of adding a physical fitness assessment to the NHANES is to see if the objectives for 1990 are met, then priority should be given to specifically assessing the population in regard to these objectives. Simply put, this means questions to determine whether adults are participating regularly in moderate to vigorous physical exercise and whether older adults are performing regular walking, swimming, or other aerobic exercise; whether there is an increase in public and professional awareness (questions regarding the type of exercise needed); whether physicians mention exercise to their patients; and whether companies sponsor fitness programs. If our only charge were to determine a baseline to monitor these objectives, a rather simple set of questions could be included in the National Survey.

The following are the general objectives worded as questions together with suggested means to answer them in terms of the specific aim in a direct and simple way.

Is health status improved? Do those who increase their fitness lessen the number and cost of contacts with health care providers? Do they lessen or improve other risk factors? Do they feel better? Are they emotionally and psychologically healthier? Are their social contacts increased in number and quality? Is their quality of life enhanced? Is risk of cardiac events increased or decreased in those who increase their exercise status in relation to their initial level of activity?

Are cardiovascular risk factor(s) reduced (i.e., physical inactivity)? (a) Is a higher proportion of physical activities continued into adult life—the emphasis should be on noncompetitive sports and sports that do not require large numbers of participants to be played. Such sports or activities would include tennis, racketball, running, walking, bicycling, skiing, and swimming. It would be necessary to determine what percentage of youths are involved in these types of activities. This would best be done through interviewers who would go to randomly chosen schools. (b) Percentage of youths in daily school physical activity? Again, this would be best accomplished by interviewers who could visit schools; (c) Percentage of adults participating regularly (three to five times per week) in moderate to vigorous exercise? This should be taken to mean those qualifying for the bare minimum for obtaining a training effect. This could be accomplished by a questionnaire completed by interview. (d) Percentage of elderly walking regularly? The definition of elderly in this context is age 65 and over (4). This could be accomplished by an interviewer-completed questionnaire.

Is public awareness increased? (e) Exercise prescription knowledge—can be evaluated by questionnaire; (f) physicians should take an activity history—this can be evaluated by a questionnaire sent to physicians and by questioning those who have seen a physician.

Are exercise services improved? (g) This could be evaluated by an interview of a random sample of companies meeting the criteria to have an exercise program for their employees. Assessment of community availability and accessibility of appropriate recreational facilities would also be useful.

Are surveillance or evaluation services improved? (h) For 70% of children—evaluating this objective necessitates a school physical education department directed testing program and could be a field walk-run test. Leisure time availability for summer and winter sports facilities should also be included.

Learning from technology used in past surveys

Care must be taken to see that proper technology is used, that the data gathered are correctly recorded, and that the measurements made have relevance. The problem with improper calibration of a bicycle ergometer for testing children in one of the surveys point out the difficulties that can arise when measuring exercise capacity. The inadvertent loss of the digital tapes of the spirometry data shows that even computerized data must be carefully protected. The ECG data were carefully digitized but in single channel format. Unfortunately, computer programs have analyzed four simultaneous channels in four lead groups for the past 10 years. Currently the best programs require 12 leads recorded simultaneously.

Methodology for future utilization of the data

The following questions will need to be resolved before planning data gathering. Will there be a questionnaire followup as for NHANES I? Will there just be a mortality or morbidity analysis or extensive future retesting? Will fixed sites be used for testing and followup or will mobile units be utilized? Will echocardiography and/or treadmill testing be included—to be performed at fixed sites or in vans? Will NHANES III be the next Framingham study? Decisions made at this point for uses of the data and decisions regarding technology require that the study design be determined soon.

Basic questions that have to be resolved regarding cardiovascular epidemiological uses of this survey include:

1. Will the measurements be repeated on the same sample or will another randomly selected sample be tested in the future?
2. Will there be a follow-up or longitudinal analysis of this population?
3. Will the results from this cross-sectional study be compared to other cross-sectional studies?
4. Are we concerned with comparison to other cultures and countries (i.e., the Canadians)?

What tests should be used in addition to questionnaires?

The major concern of this paper is to recommend how the information can be used. However, the recommendations are strongly dependent on the data gathered and the methodology utilized. Because medical technology changes so rapidly, what is "high tech" today may be outdated and useless tomorrow. For example, in only 10 years, M-mode echocardiography has been replaced by 2–D, and now doppler is considered necessary for complete ultrasound cardiac studies (5). Oxygen consumption analysis equipment can be difficult to keep operational, and new measurements have been proposed (6–9). Of these proposed, anaerobic threshold has had the most research performed, but it has had little clinical utility (10, 11). ECG reading should be performed by computer in the next survey, but physicians need to overread the tracings (12). Hi-frequency ECG analysis and/or portable digital ECG analysis may replace Holter monitoring. However, surveys should avoid the latest unvalidated research methodology or sophisticated technology. Furthermore, expensive equipment is needed that is too difficult to maintain, calibrate, and operate. Questionnaires remain the major tool in large surveys as reviewed in other parts of this document.

The survey plans might include a random sample of a subpopulation to have measured maximal oxygen consumption tests performed in a standardized laboratory. These data can be correlated with the results of questionnaires and other simpler tests. Because of the interaction of health and fitness and the social and economic strata in our society, it would be wise to correlate the results of treadmill testing including gas analysis with the simpler measures of fitness in subsets. Regression equations most likely will be specific for different groups and perhaps will not apply in other groups.

The questionnaires included in the baseline data gathering in this study must include questions assessing involvement in 10 kilometer and other popular races to assess their impact on health. Are these runners semi-professional runners or is it more common for people to become involved and then continue to run as a hobby? Are the races merely a way to work off aggression and a competitive spirit, or do they serve to motivate people to maintain a health style of regular exercise? Do they promote a healthy life style, or are they an analog of anorexia and result in extra physician visits? The health care patterns of such runners should be categorized. Does exercise promote other healthy habits and preventive health concerns, or does it lead to quackery? Analysis of individuals participating in these runs would be an interesting contrast to the random population sample. Are they harmful or helpful? Do they generate other health problems (e.g., orthopedic)?

Physical fitness versus physical activity

The question remains whether activity level or actual maximal oxygen uptake best predicts coronary risk. Leon has shown that in healthy men the results of resting measurements and a questionnaire correlate highly with total exercise time using a multivariate equation (13). Goldman (14, 15) has also shown a high correlation in cardiac patients between his specific activity scale and treadmill performance. Peters has found a relatively low physical work capacity on a bicycle ergometer to predict a significant risk for MI in a working population (16). This study relates measured work capacity directly to cardiac events that occurred during followup. Information about population data on normal expected values for fitness levels are needed.

For this survey, priority should be given to assay physical activity rather than markers of fitness. Numerous studies have measured maximum oxygen consumption in various groups including pilots, college students, athletes, and firemen. These studies involve selected groups. However, only a selected group of those asked to perform an exercise test will do so in this survey. There is a 75% acceptance rate for the NHANES examination, and this may decrease if an exercise test is included. An exercise test to be considered would be the Canadian Survey step test because it can be done in the home and has a considerable data base relating it to maximal oxygen consumption and fitness (9). Though an increase in maximal oxygen consumption is the best indicator of increased activity or training, it is not clearly related to health or cardiovascular disease. Only reported activity level has been related to risk for cardiovascular events; physical inactivity has been determined by job or leisure-time activity to be an independent risk factor. Numerous questionnaires have been validated, including the one used by the Canadians. (See Shephard, chapter 18.) A thorough review of this topic can be found in the paper by LaPorte (chapter 24).

Special testing procedures

In chapter 4, Wilmore has described a physical fitness test battery. In the following section, we will comment on it and other cardiac tests that may be considered for inclusion.

Physical fitness test battery

The concept of health-related physical fitness should be accepted in designing a physical fitness test battery (Wilmore, chapter 4). Health-related physical fitness is defined as the ability to perform strenuous physical activity with vigor and without excessive fatigue, and as a demonstration of physical activity traits and capacities that are consistent with minimal risk of developing diseases associated with physical inactivity.

Health-related physical fitness includes the components of cardiorespiratory endurance, muscular strength and endurance, body composition, and flexibility. Previous tests of physical fitness included the components of motor fitness, i.e., agility, power, speed, and balance. Although these are important considerations for the growing child and the athlete, they should not be considered part of a health-related physical fitness test battery for the general population. The limitations inherent in each are discussed below.

Cardiorespiratory endurance

The gold standard for this is measured maximal oxygen uptake. This can be estimated by field tests of running and walking, step tests, workload achieved on bicycle or treadmill, or time in fixed treadmill and bicycle protocols. However, it is affected by age, nutritional status, and genetic and other factors. Perhaps a recovery or submaximal heart rate measurement might better indicate fitness. However, this concept requires evaluation. Estimation of maximal oxygen uptake from submaximal exercise tests and from recovery heart rates has been used but the correlation is poor. Though an increase in $\dot{V}O_2$ is the best marker of a training effect (i.e., an increase in aerobic fitness), there are problems in using it as a single measurement of activity level at a given point in time. Nutritional status and genetic and other factors significantly affect $\dot{V}O_2$. Thus, it does not represent activity levels in different populations. For example, a given population with a lower nutritional status could have a lower max $\dot{V}O_2$ even though they are as physically active as another population with better nutrition.

Strength and muscular endurance

Strength and muscular endurance can be assessed directly, either by simple field tests such as the one-repetition maximum or by using sophisticated equipment such as a Cybex II isokinetic testing device. A hand grip test was used in the Framingham Study and was considered an indirect marker of fitness (1). However, there is little to suggest that isometric hand strength relates to health or to cardiovascular fitness. There is little to support the inclusion of any such test in the national survey; however, the Department of Defense has a need for such information in the younger age groups.

Flexibility

There is no universal agreement on what constitutes an adequate flexibility test battery. Flexibility is highly specific to the joints tested. Testing should involve activities of normal daily life that create limitations when flexibility is limited, i.e., tying one's shoes. Its assessment is confounded by the effects of age, arthritis, obesity, injury, and congenital abnormalities. Its inclusion could lead to a definition of normal, and it could have an important impact on ergonomics in many areas including health planning and architecture. Because its relationship to cardiovascular health and fitness is uncertain, pilot studies are needed to understand it better before its use in surveys.

Body composition

Hydrostatic or underwater weighing is the gold standard for estimating percent body fat, but skinfold measurement is a reasonable alternative that is usually used. A minimum of six sites should be chosen and equations not used; instead the absolute measurements should be indicated. Equations to estimate percent body fat from these measurements have a great deal of variability and poor correlation. It is unlikely that these measurements can add to the understanding of fitness;

thus they should not be included in the survey. Training decreases the percent of body fat; skinfold measurement thus may be justified as part of a nutritional evaluation. The recent finding that the percent body fat in children is directly related to the number of hours they watch television daily suggests that such an analysis may have value to answer selected questions (if the assumption that they are sedentary while watching TV is true). Also, the recent association with body habitus with the probability of becoming obese may support the use of body composition measurements.

Comments on cardiac techniques

Echocardiography has been a major advance in the noninvasive evaluation of cardiac disease. The different forms of cardiomyopathy can be recognized, including the congestive and hypertrophic types which may carry a poor prognosis (17). The hypertrophic forms are associated with sudden death in young athletes, and the congestive forms appear to be more common in middle-aged individuals. Training in young people can result in changes in cardiac volumes and wall thickness. It appears that cardiac changes only occur in younger individuals just as they only occur in younger animals (18). Because of the great variability in echocardiographic measurements and because changes with training are relatively small, they may be difficult to demonstrate (19).

There are technologic problems with echocardiography. The M-mode echo, which consists of an "icepick" view or usually views of the heart, simply is largely dependent on the skill of the technician, the resolution of the equipment, and the patient. Individuals with lung disease, obesity, or chest deformities often cannot be studied well. The M-mode measurements have poor reproducibility. Whether one measures from the leading or the trailing edge of structures has been argued by investigators. Many experts suggest that the interpretation of M-mode echocardiography should be qualitative rather than quantative. The two-dimensional approach, in which the icepick of ultrasound sweeps across a band through the heart showing a planar view, is most commonly used now. Results are dependent upon both technician and patient, and the equipment is rapidly changing. Quantitative measurements have not been resolved with this technique. However, this technique has been much more qualitative than quantitative even from its inception. There have not been longitudinal exercise training studies reported using two dimensional echocardiography. Most recently doppler studies have been added. The Doppler has the ability to measure flow; cardiac output and regurgitant volumes can be estimated. No exercise training studies have been reported using this technique. An interesting observation has been tricuspid regurgitation in female runners.

In summary, echocardiography has many factors needing strict control to make it suitable for epidemiological studies. The mode of recording and the instrumentation must be standardized, yet the technology continues to change. Quantitative measurements remain a problem. Images depend much on the equipment, the operator, and the subject. There would be a selection process because of image quality, possibly excluding patients who smoke or who have illnesses that make it difficult for them to be imaged. All of these problems would be compounded in a large population study that would require multiple machines and technicians.

Treadmill testing with oxygen consumption and other gas exchange measurements is the gold standard for measuring fitness and exercise capacity (20). However, the instruments for gas analysis are relatively expensive and difficult to keep calibrated and repaired. These procedures are done only in specialized research centers. There is great variability in the measurements made. The major problem in instrumentation is in flow measurement because respiration is phasic. The gas analyzers are usually accurate; but they are subject to frequent breakdown and have to be calibrated often. The new gas exchange variables including anaerobic threshold are of uncertain significance. They have poor reproducibility, and a suitable automated analysis technique is not available. Gas exchange anaerobic threshold may not relate to buildup of blood lactate but may be due to a sympathetic or catecholamine discharge. Gas anaerobic threshold is not always increased by exercise training nor is it an independent marker of training (11). For these reasons, it seems unwise to include the gas exchange variables in a population study. It may be of value to obtain these data in a small subset of subjects in a selected research laboratory in order to correlate the measurements with questionnaires or to ambulatory monitoring. The problems with using the exercise ECG as a screening technique are discussed later.

Ambulatory monitoring can accurately follow heart rate over long periods of time. The computer analysis of arrhythmias is far from perfected (21), but blood pressure and heart rate can be accurately measured. How this relates to activity depends on the individual and his or her response to other stresses (22). Such a device coupled with a device to measure lower extremity movement may serve to grade activity. These techniques require validation before application.

Cross-sectional epidemiological studies

Assuming that this will be primarily a cross-sectional study, numerous questions can be asked about the results of an activity questionnaire or other measures made that would be informative from a cardiovascular epidemiological point of view (table 1). For instance, what is the relationship of activity status to other cardiovascular risk factors? Do those who exercise regularly have better nutritional habits, ideal weight, smoke less, have lower blood pressures, and have lower lipid levels? How are exercise patterns related to alcohol ingestion and sleeping disorders? What knowledge, background, or factors motivate individuals to maintain a healthy life style? Is activity related to educational and/or socioeconomic level? Does peer pressure, to exercise or to not exercise, operate in the different socioeconomic groups? Are there gender differences in fitness levels? Is mental health better or worse in those who exercise? Are those who exercise less depressed? What is the relationship of childhood to adult exercise? What is the reaction of work to leisure exercise? Is productivity greater in those who exercise regularly? Do they have less sick days? Do those who exercise at the job have a greater loyalty to their employers? Is usage of the health care system greater or less in those who exercise? Is the quality of life better for those who exercise? Are their interpersonal relationships and sexual activity better? Does a fear of exercise secondary to experiences or misconceptions affect exercise habits? Do parents encourage children to maintain an exercise program? Do those who exercise use the health care system more or less? Are they hypochondriacal? Do they turn more to quackery and "health foods"? All these questions could provide data about the exercise hypothesis and could help find out why and how people comply with personal exercise programs and whether there are adverse effects.

It is important in a cross–sectional analysis to relate disease to activity status and to ask questions about temporal relationships. Caution is needed since a causative relationship cannot be determined.

Research questions potentially answered by questionnaire data

Numerous hypotheses regarding the relationship of cardiovascular disease and physical activity could be tested by using the results of a survey with an activity questionnaire. Some relationships that could be evaluated in regard to physical activity include:

1. What factors motivate an individual to exercise?
2. Does regular exercise improve the quality of life?
3. Is productivity enhanced?
4. How does socioeconomic status affect exercise habits?
5. Does exercise affect other risk factors?
6. Does exercise affect health care utilization?
7. Does exercise affect psychological status?
8. Does exercise affect health consciousness?
9. Type and amount of availability, accessibility of facilities.

Table 1. Cardiovascular epidemiologic uses of the data

Relationship of physical activity levels to:
  Other risk factors
  Health knowledge or preconceptions
  Quality of life
  Motivation
  Eating habits and nutrition
  Alcohol consumption
  Sleeping
  Psychological health
  Economic status
  Dyadic adjustment and socialization
  Work attendance and productivity
  Health care utilization
  Socioeconomic status
  Health consciousness and personal responsibility for health care or maintenance

Motivation to exercise

Knowledge of what motivates different groups to maintain an active lifestyle or what preconceived ideas keep them from exercising could be helpful in effecting changes in behavior. Marketing techniques could be targeted if we knew what got different individuals started and what kept them in exercise regimens. Sports or sports figures might be helpful in advertisement efforts in certain groups, and scientific discussions of potential health benefits would be more effective in other groups. Teenagers certainly try to emulate their peer leaders and rock stars more than they do their parents. Strategies to cause the greatest changes in physical activity could be developed by such findings.

Quality of life

There are numerous questionnaires assessing quality of life. The Rand and McMasters questionnaires are examples that most closely apply to the NHANES objectives. These questionnaires deal with issues of mental health, socialization, life satisfaction, functionality, health perceptions and other areas of personal adjustment. It has been assumed that quality of life is better for those who exercise, but this has not been tested (23). Does participation in regular exercise lead to greater social contacts and improve dyadic adjustments, or is an exercise-habit a marker for isolation and problems with interpersonal relations?

Does regular exercise promote confidence in self, or is it a marker of feelings of inadequacy? Recent correlation of distance running and anorexia nervosa may extend to less severe forms of exercise addiction. Validated questionnaires now make it possible to relate activity levels to these very vital aspects of life.

Productivity

It has been suggested that individuals who exercise regularly are more productive at their jobs and in their personal lives, presumably because of a higher level of energy and stamina. It has also been suggested that active individuals have less absenteeism from work because of sickness or physicians visits. These hypotheses could be tested using data gathered during NHANES III. It would be important to establish age and gender differences. Occupational differences in relation to occupational activity are also of interest.

Socioeconomic status

The relationship of socioeconomic status to physical and mental health is well documented, i.e., the higher socioeconomic groups get better care, have less illness, and are less depressed. However, there are no longer simple associations between blue collar, unskilled labor and increased physical activity and physical inactivity and professional status. Both activity at work and pay status have changed over time. Blue collar labor, such as carpentry and equipment repair, is now a relatively high paying job and is relatively inactive. Trends toward increased leisure physical activity has particularly affected the upper socioeconomic classes. Activity at work and leisure-time activity are not inversely related. In fact, the small gradient of activities at work makes it unlikely that anyone is too fatigued from work activity to maintain a personal exercise regimen. Nonetheless, most of these relationships have not been established in a scientific manner; this could be done with survey techniques used in this study.

Cardiovascular risk factors

It has been hypothesized that those who maintain an active lifestyle are also more likely to modify other cardiovascular risk factors. Do those who exercise regularly have better nutritional habits, weight, smoke less, have lower blood pressures, and optimal blood lipid levels? Do they more often turn to "quackery" and pseudo health foods? Another important relationship is obesity and exercise status. Do those who exercise regularly more easily control their weight? Is exercise an effective means of weight control for many individuals? If exercise has "spinovers" into other areas of healthy behavior modification, it could be more strongly supported. Many individuals in an exercise program greatly alter other risk factors; however, in the context of a cardiac rehabilitation program, it is difficult to see other risk factors modified as exercise capacity improves.

Health care utilization

It has been hypothesized that individuals who maintain a more active lifestyle utilize health care services much less. On the other hand, the increase in exercise-induced orthopedic injuries may result in an "army of walking wounded" to the offices of physicians. The neurotic weekend athlete is most likely an extreme; he or she may be a hyperchondriacal individual who would frequently consult a number of physicians even if he or she did not exercise. These important questions could be answered by this study.

Psychological status

There are many reported psychological benefits of exercise (24–26), but few have been well validated. Are those who exercise regularly less depressed or less likely to have a drinking problem, drug dependency, or sleeping disturbances? Is the personality addicted to exercise equally likely to be addicted to other less healthy behavior? Many individuals exercise as a means of relieving stress, and this may be a healthier means than drinking or drugs. Can exercise be used to control antisocial behavior? Are aggressive and angry people made more tolerant of society by a regular exercise program? Can sports and exercise be promoted as a means to avoid gang behavior in lower socioeconomic groups?

Health consciousness

The next major advances in health care will probably be evident as individuals become responsible for their own health. Is a regular exercise program a marker for health-related social responsibility, or is it a means of developing these qualities?

Longitudinal epidemiological studies

Followup of the study population could evaluate what types of exercise are preventive of cardiovascular disease. The types of activity or even a threshold activity level has not been established. Physical activity could alter risk in many ways. These include thrombolysis, lowering lipid levels, lessening catecholamine induced damage, decreasing stress and aggressive behavior, and strengthening the arterial wall (20). These actions might be influenced by different types of activity.

Followup for disease end-points

One value in quantifying activity levels would be their relationship to the future development of coronary heart disease and other cardiovascular illnesses. However, the foremost reason for data gathering is to evaluate strategies to effect health objectives. In other words, an intervention is to be initiated, and we only aim to quantitate the effect of different strategies for this intervention. In contrast to a valid intervention study, here there is always a strong selection bias in the choice of whether to exercise. It would be better to specifically question people regarding which interventions were most effective in regard to their making changes. Why did they alter activity status? Use of the National Death Registry would permit the assessment of mortality in the survey sample.

Screening for silent coronary heart disease

If exercise testing is used to measure physical fitness in this survey, its use as a screening technique must also be considered. Its varying predictive value is the result of its use in populations with different prevalences of coronary disease. Various techniques have been recommended to improve the sensitivity and specificity of exercise testing, such as new computerized and noncomputerized electrocardiographic criteria, nonelectrocardiographic exercise test responses, cardiac radionuclide procedures, systolic time intervals, cardiokymography, cardiac fluoroscopy, and the computerized application of Bayesian statistics using risk factors and risk markers (27). Limited data suggest that angiographically documented asymptomatic coronary disease has a relatively good prognosis compared with symptomatic disease and rarely should require coronary artery bypass surgery. Individuals so identified should be prime targets for behavior modification with the hope of avoiding the usual course of this disease. Problems arise regarding the impact an abnormal test result might have on an individual. It may ruin his or her occupation, increase the cost of health insurance, and cause emotional trauma. If followup is to study the predictive value of the test for screening, there is moral and ethical concern as to whether the individual should be informed of the abnormal result.

Screening can be defined as the presumptive identification of unrecognized disease by the utilization of procedures that can be applied readily. The relative value of techniques for identifying individuals who have asymptomatic or latent coronary heart disease (CHD) should be assessed to optimally and cost effectively direct secondary preventive efforts towards those with disease. Eight criteria (28) have been proposed for the selection of a screening procedure: the procedure is acceptable, appropriate, and safe; the quantity and/or quality of life can be favorably altered; the results of intervention outweigh any adverse effects; the target disease has an asymptomatic period during which its outcome can be altered; acceptable treatments are available; the prevalence and seriousness of the disease justify the costs of intervention; the procedure is relatively easy and inexpensive; and sufficient resources are available. In addition, seven guides (29) have been recommended for deciding whether a community screening program does more good than harm: Has the program's effectiveness been demonstrated in a randomized trial? If so, are efficacious treatments available? Does the current burden of suffering warrant screening? Is there a good screening test? Does the program reach those who could benefit from it? Can the health care system cope with the screening program? Will those who had a positive screening comply with subsequent advice and interventions?

Sensitivity and specificity are inversely related. That is, when sensitivity is the highest, specificity is the lowest and vice versa. Any test has a range of inversely related sensitivities and specificities that can be chosen by selecting a certain discriminant or diagnostic value. Attempts have been made to use a series of tests to improve diagnostic power, but test interaction is complex.

Economic factors must be considered when planning a large survey. Hartley and colleagues reported an exercise testing program designed to examine large numbers of people effectively, conveniently, and inexpensively (30). More than 1,800 subjects were examined in 3 years. As many as 55 tests per day were performed at a cost of $60 to $70 each. Abnormalities uncovered were similar to those observed in other studies. The program was considered successful for rendering services conveniently, at low cost, and with accuracy.

Followup studies that have used exercise testing

Table 2 summarizes 11 followup studies (31–41) that used maximal or near-maximal exercise testing to screen asymptomatic individuals for latent coronary heart disease, and one that evaluated men and women with atypical chest pain (summarized in reference 20). The populations in the top eight studies were tested and followed for the coronary heart disease endpoints of angina, acute myocardial infarction, and sudden death. The bottom three studies considered only hard endpoints (i.e., other than angina). Table 3 shows the endpoints that occurred in each study. There has been controversy as to whether, in the absence of conventional risk factors, exercise testing provides additional prognostic information in asymptomatic men. Another concern is whether the knowledge of having an abnormal exercise test makes an individual more likely to report angina during followup.

McHenry et al. (38) reported the results of an 8- to 15-year followup of 916 apparently healthy men between the ages of 27 and 55 (mean 37 years) who underwent serial medical and exercise test evaluations. In 1968, the Indiana University School of Medicine entered into an agreement with the Indiana State Police Department to provide employees with periodic medical evaluations including treadmill tests. The report of their experience with the first male employees who underwent initial medical evaluations between July 1968 and June 1975 includes a followup for all subjects through June 1983. A CC5 lead was monitored, and 1 mm or more horizontal or downsloping ST segment depression during or after exercise was considered abnormal. A modified Balke protocol was used for all treadmill tests, and most were symptom limited. Serial evaluations were planned at 2- to 5-year intervals; however, about 15% of subjects elected not to return after their initial evaluation. During the initial evaluation, there were 23 subjects with an abnormal ST segment response. During followup, there were 9 coronary events in this group: 8 cases of angina and 1 of sudden death. With serial testing, an additional 38 subjects experienced conversion to abnormal ST segment response. During followup, there were 12 coronary events in this group: 10 cases of angina, 1 MI, and 1 "other." There were 833 subjects with normal ST segment responses to exercise with all tests. In this group, there were 44 coronary events, 25 MI, 7 sudden deaths, and 12 angina. They concluded that an abnormal ST segment response to exercise predicted angina pectoris but not other coronary events.

McHenry et al. did not present sensitivity or specificity calculations, but the data they reported enabled the calculations shown in table 2. The surprising low sensitivity from initial testing is probably the result of the long followup period. An abnormal test indicates obstructive coronary disease that was most likely not present initially in most subjects who developed endpoints but developed later during the 12-year followup. An analysis of the treadmill test performance at 5 years, a time frame similar to the prior studies reporting a higher sensitivity, would be most informative. Otherwise, there is the possibility that the treadmill test is less sensitive in asymptomatic men than previously demonstrated.

They found that serial testing did not improve the predictive value of the test and that angina was the main cardiac event predicted. Sudden death was more common in individuals with normal test results. The USAFSAM study also had angina as its most common endpoint, supporting the concept that the knowledge of an abnormal exercise test makes an individual more likely to report angina and/or to be diagnosed as having CAD. The MRFIT, LRC, and Seattle Heart Watch studies only used hard cardiac endpoints (death or MI) and abnormal ST depression had a much lower predictive value in these three studies (33, 40, 41).

Studies using coronary angiography can be thought of as "instant epidemiologic" studies because endpoints need not be delayed. In an angiographic study at USAFSAM, of 111 asymptomatic men with an abnormal ECG response to a treadmill test, only one-third had at least one lesion equal to or greater than 50% luminal narrowing of a major coronary artery (42). Erikssen and colleagues reported their angiographic findings in 105 males from a working population of 2,014 men with either a positive questionnaire for angina pectoris, typical angina during a near maximal bicycle test, or an abnormal exercise electrocardiogram (43). The exercise test had a predictive value of 84% if a slowly ascending ST segment was included. The higher predictive value in this study might be because of the older age of their population and the inclusion of men with angina. Of the 36 men who had normal coronary arteries, a 7-year followup revealed that 3 died suddenly, 4 received a diagnosis of cardiomyopathy, and 1 developed aortic valve disease. They had a relative decline in their physical performance over the followup period. Thallium studies were normal, but the radionuclide ventriculogram revealed a subnormal increase in ejection fraction during exercise in half. Eriksson's findings have been disproven in a larger data set as part of CASS (44). Kemp et al. demonstrated that individuals with a normal coronary angiogram have a good prognosis regardless of the exercise ECG response.

In 255 asymptomatic men who underwent coronary angiography for an abnormal ECG response to exercise testing over a 7-year period at the USAFSAM, none of the clinical or resting ECG variables were able to detect those with significant disease (29). The three exercise test responses with high likelihood ratio were at least 0.3

Table 2. Studies using exercise testing as a screening test

Study	*N*	Years followed	Incidence CHD (%)	Sens (%)	Spec (%)	Predictive value + %	Risk ratio
			Angina endpoint				
Bruce (31)	221	5	2.3	60	91	14	14×
Aronow (34)	100	5	9.0	67	92	46	14×
Cumming (35)	510	3	4.7	58	90	25	10×
Froelicher (32)	1,390	6	3.3	61	92	20	14×
Allen (36)	356	5	9.6	41	79	17	2.4×
Manca (37)	947	5	5.0	67	84	18	10×
	508(w)	5	1.6	88	73	5	15×
MacIntyre (39)	578	8	6.9	16	97	26	4×
McHenry (38)	916	13	7.1	14	98	39	6×
Averages[1]				48	90	26	9×
			Nonangina endpoint				
Seattle Heart Watch (33)	2,365	6	2.0	30	91	5	3.5×
MRFIT (40)	6,217 (SI)	6–8	1.7	17	88	2.2	1.4×
	6,205 (UC)		1.9	34	88	5.2	3.7×
LRC (41)	3,630	8	2.2	28	96	12.0	6×
Averages				27	91	6	4×

[1] Averages do not include women's study.
w = Women.
SI = Special intervention group.
UC = Usual care group.

Table 3. Events used as endpoints for followup studies

Study	*N*	Number of events	Total deaths	CV deaths	MI	CABS	AP
Aronow	100	9	3	3	4	1	1
Bruce	221	5	NR	1	1		3
Cumming	510	26	5	3	8		13
McHenry	916	65	8	8	26		30
MacIntyre	548	38	NR	10	16	6	6
Allen	888	48	NR	?	?	NR	?
Froelicher	1,390	46	17	11	6	4	19
Seattle Heart Watch	2,365	47	25	8	23	5	11
MRFIT (SI)	6,427	265	115	NR	NR	NR	NR
(UC)	6,438	260	124	NR	NR	NR	NR
LRC	3,630	NR	151	75	NR	NR	NR

MI = myocardial infarction; CABS = coronary bypass surgery; AP = angina pectoris; NR = not reported; ? = used as endpoint; SI = special intervention group; UC = Usual care group.

mv ST depression, persistence of ST depression 6 minutes post-exercise, and less than 9 METS exercise capacity. However, because of their low sensitivity and predictive value, it was necessary to combine them with risk factors. A combination of any risk factor and two of these exercise responses was highly predictive (89%), but insensitive (39%), for any coronary disease. However, this combination had a sensitivity of 55% and a predictive value of 84% for two- or three-vessel diseases. These angiographic studies confirm the low predictive value of an abnormal exercise test response.

Techniques to improve screening

Numerous techniques have been recommended to improve the sensitivity and specificity of exercise testing. Various computerized criteria for ischemia have been proposed, as well as new standard visual ST criteria. In addition, there are ancillary techniques that might improve the discriminating power of the exercise test. These methods are listed in table 4. As described, of these methods, digital fluoroscopic imaging of coronary artery calcification has the most promise (27).

Computer probability estimates

Diamond and Forrester (45) reviewed the literature to estimate pretest likelihood of disease by age, sex, symptoms, and the Framingham risk equation (based on blood pressure, smoking, glucose intolerance, resting electrocardiogram, and cholesterol level). In addition, they have considered the sensitivity and specificity of four diagnostic tests (the exercise test, cardiokymography, thallium imaging, and cardiac fluoroscopy) and applied Bayes's theorem. They derived a system of decision analysis that prescreens individual patients before they undergo more expensive tests. This enhances the predictive value of these noninvasive tests by selecting a subgroup with a greater pretest likelihood of disease (perhaps with a 15–40% prevalence), so that the post test probability of an abnormal test will be raised to 60–80%. The major weakness of this approach is that

Table 4. Ancillary techniques used to screen for asymptomatic CHD

Thallium perfusion imaging

Radionuclide ventriculography during bicycle exercise and after treadmill exercise

Cardiac fluoroscopy for coronary artery calcification (27) (enhanced with digital subtraction angiography)

Cardiokymography

Total cholesterol/HDL ratio, conventional coronary risk factors

ECG gated chest x ray before and after exercise

Computerized multifactorial risk prediction using Bayesian statistics

Systolic time intervals during and after exercise

Digital subtraction angiography with intravenous injection of contrast to visualize the coronary arteries

Echocardiography (or Doppler) during and/or after exercise (even after treadmill) the sensitivities and specificities of the secondary tests are not yet established, and it is uncertain how they interact (46).

In addition, a step approach that uses risk markers to identify a high risk group excludes the majority of individuals who will eventually develop coronary disease. This approach concentrates the preventive impact on the small, high risk group while ignoring the majority of individuals in the moderate risk range who will contribute larger numbers but at a lesser rate to disease endpoints.

Bruce and colleagues examined the motivational effects of maximal exercise testing for modifying risk factors and health habits (47). A questionnaire was sent to nearly 3,000 men who had undergone symptom-limited treadmill testing at least 1 year earlier. Sixty-three percent of the responders indicated that they had modified one or more risk factors and health habits and that they attributed this change to the exercise test.

Secondary prevention and testing

Hypothetically, if a method of secondary prevention of coronary artery disease were proven and available today, the following three-step approach to screening for asymptomatic coronary heart disease in men over 35 years old appears reasonable. First, angina history, risk factor analysis, and a resting electrocardiogram should be obtained. Digital fluoroscopy to look for coronary artery calcification has the greatest promise for providing cost-effective, sensitive screening at this step (27). If any data collected place the individual at risk, the second step should be a maximal exercise test. If this test is interpreted as abnormal based on ST-segment or other abnormal responses, the third step should be Thallium exercise scintigraphy, cardiac fluoroscopy, or cardiokymography. The lack of data on the diagnostic value of these techniques in asymptomatic individuals prevents precise recommendations at this time. Good clinical judgment must be exercised to avoid causing professional, psychological, and/or financial harm by mislabeling healthy people. The severity of the abnormal response must be considered, and often it is appropriate merely to follow an individual with an abnormal test response. Exercise testing well people has been judged to be inappropriate by the AHA/ACC Task Force (48).

Exercise testing for exercise programs

There are multiple reasons for performing an exercise test prior to initiating an exercise regimen. The optimal exercise prescription, based on a percentage of an individual's maximal oxygen consumption (50–80%), can only be written based on an exercise test. The best way to assess the risk of an adverse reaction during exercise is to observe the individual during exercise. The level of exercise training then can be set at a level below that at which adverse responses or symptoms occur. Some individuals motivated by popular misconceptions about the benefits of exercise may disregard their natural "warning systems" and push themselves to dangerous levels of ischemia.

An individual with a good exercise capacity and only 0.1 mV ST segment depression at maximal exercise has a relatively low risk of cardiovascular events in the next several years compared with an individual with marked ST segment depression at a low double product (heart rate and systolic blood pressure). Most individuals with an abnormal exercise test can safely exercise if the level or intensity of the exercise at which the response occurs is considered. Such patients can be followed with risk factor modification rather than being excluded from exercise or their livelihood.

Exercise testing is indicated prior to entering an exercise program for individuals with a strong family history of coronary disease (i.e., family members aged less than 60 years with coronary event), the presence of significant coronary risk factors (particularly serum cholesterol), or any symptoms suggestive of myocardial ischemia currently or in the past. In addition, a group of patients self-select themselves for exercise testing; they may request the test even though they deny having symptoms.

High risk selection

A problem with using exercise testing only in patients with abnormal risk factors is that a large number of patients with coronary artery disease would be excluded. Thus, although this approach increases the pretest probability of coronary artery disease and improves the predictive value of an abnormal response, it leaves a large number of patients with potential coronary artery disease without the potential benefit of screening. It has been hypothesized that this approach concentrates the preventive impact on the small, high risk group but ignores the majority of individuals in the moderate risk range. This hypothesis could be tested in this survey.

Conclusions: Use of tests for screening

Because it will be some time before the primary prevention of coronary artery disease is a reality, it is advisable to evaluate screening methods for diagnosing the earliest signs or symptoms of myocardial ischemia. Because risk factor screening and techniques with the patient at rest have limited sensitivity, exercise testing to induce ischemia and elicit abnormalities not present at rest deserves consideration. However, digital fluoroscopy appears to be a noninvasive procedure performed at rest that may be better than exercise testing (27). The iatrogenic problems resulting from screening must be considered, and the results of testing must be applied using the predictive model and Bayesian statistics. Test results must be thought of as probability statements, not absolutes. The problems not answered and the limitations of previous screening studies could be resolved by a followup study of subjects properly tested and screened as part of NHANES. Because of the questionable results from prior studies, it would be ethically possible to not inform individuals with abnormal results of them unless they were markedly abnormal. This would avoid the potential bias of reporting angina if one knows of an abnormal result. Thus, it could be determined if the exercise test adds information to risk factor screening and if abnormal responses only "predict" angina or just bias individuals toward reporting angina. A strategy for evaluation of abnormal responders using other procedures could also be tested. Although 10 screening studies used exercise testing, more questions remain than are answered. In addition, in the studies that only used hard endpoints, the predictive value of the test was much lower than in studies that included angina as an endpoint. If careful planning was done, NHANES could resolve some of these issues. Digital fluoroscopy to identify coronary artery calcification should be considered for inclusion in NHANES (27).

Dissemination of results

The NHANES and NHIS are much less known among health professionals than they should be. They represent a national information resource that should be utilized more. Relatively fewer publications have come from these surveys than from, for example, the Framingham Study. This situation could be improved by contracting with members of the academic community for analysis and presentation of portions of the data. There can be positive advantages to the "publish or perish" philosophy that prevails in the academic setting. Perhaps the current wide availability of adequate computer and statistical power to handle large data base and sophisticated statistical techniques at universities will lead to more use of these data. Though available to schools of public health, future efforts should include dissemination to medical schools and the medical profession.

References

1. Kannel, W.B., Wilson, P., and Blair, S.N.: Epidemiological assessment of the role of physical activity and fitness in development of cardiovascular disease. *Am. Heart J.* 109:876–885, 1985.
2. Paffenbarger, R.S., Hyde, R.T.: Exercise in the prevention of coronary heart disease. *Preventive Medicine.* 13:3–22, 1984.
3. Paffenbarger, R.S., Hyde, R.T., Wing, A.L., et al.: A natural history of athleticism and cardiovascular health. *JAMA* 252:491–495, 1984.
4. Wenger, N.K., Furberg, G.D., and Pitt, E. (eds). *Coronary Heart Disease in the Elderly.* New York. Elsevier, 1986.
5. Nishimura, R.A., Miller, F.A., Callahan, M.J., Benassi, R.C., et al.: Doppler echocardiography: Theory, instrumentation, technique, and application. *Mayo Clin. Proc.* 60:321–343, 1985.
6. Hammond, H.K., and Froelicher, V.F.: Exercise testing for cardiorespiratory fitness. *Sports Medicine.* 1:234–239, 1984.
7. Shephard, R.J.: Tests of maximum oxygen intake. *Sports Medicine* 1:99–124, 1984.
8. Weber, K.T., and Janicki, J.S.: *Cardiopulmonary Exercise Testing.* New York. W.B. Saunders, 1986.
9. Sullivan, M., and Froelicher, V.F.: Maximal oxygen uptake and gas exchange in coronary heart disease. *J. Cardiac Rehabil.* 3:549–560, 1983.
10. Yeh, M.P., Gardner, R.M., Adams, T.D., Yanowitz, F.G., and Crapo, R.O.: Anaerobic threshold: Problems of determination and validation. *J. Appl. Physiol: Respirat. Environ. Exercise Physiol.* 55:1178–1186, 1983.
11. Sullivan, M., Ahnve, S., Froelicher, V.F., and Meyers, J.: The influence of exercise training on the ventilatory threshold of patients with coronary heart disease. *Am. Heart J.* 109:458–462, 1985.
12. Willems, J.L., Arnaud, P., Van Bemmel, J.H., Bourdillon, P.J., et al.: Assessment of the performance of electrocardiographic computer programs with the use of a reference data base. *Circulation* 71:523–534, 1985.
13. Leon, A.S., Jacobs, D.R., DeBacker, G., and Taylor, H.L.: Relationship of physical characteristics and life habits to treadmill exercise capacity. *Am. J. Epidemiol.* 113:653–660, 1981.
14. Goldman, L., Cook, E.F., Mitchell, N., Flatley, M., Sherman, H., and Cohn, P.F.: Pitfalls in the serial assessment of cardiac functional status. *J. Chron. Dis.* 35:763–771, 1982.
15. Goldman, L., Hashimoto, B., Cook, E.F., and Loscalzo, A.: Comparative reproducibility and validity of systems for assessing cardiovascular functional class: Advantages of a new specific activity scale. *Circulation.* 64:1227–1234, 1981.
16. Peters, R.K., Cady, L.D., Bischoff, D.P., Bernstein, L., and Pike, M.C.: Physical fitness and subsequent myocardial infarction in healthy workers. *JAMA* 249:3052–3056, 1983.
17. Miller, D.H., and Borer, J.S.: The cardiomyopathies. *Arch. Intern. Med.* 143:2157–2162, 1983.
18. Schaible, T.F., and Scheuer, J.: Cardiac adaptations to chronic exercise. *Prog. Cardiovasc. Dis.* 27:297–324, 1985.
19. Ditchey, R.V., Watkins, J., Mc Kirnan, M.D., and Froelicher, V.: Effects of exercise training on left ventricular mass in patients with ischemic heart disease. *Am. Heart J.* 101:701–706, 1981.
20. Froelicher, V.F.: Exercise and the Heart. Clinical Concepts. Year Book Medical Publishers, Inc. Chicago 1987.
21. Kennedy, H.L., and Ratcliff, J.W.: Ambulatory electrocardiography and computer technology—practical advantages. *Am. Heart J.* 113:186–193, 1987.
22. Rauscha, F., Glogar, D., Weber, H., Niederberger, M., and Kaindl, F.: Diagnostic value of exercise testing versus long-term ECG in evaluation of arrhythmias in old age. *Europ. Heart J.* 5:79–83, 1984.
23. Wenger, N.K., Mattson, M.E., Furberg, C.D., and Elinson, J.: Assessment of quality of life in clinical trials of cardiovascular therapies. *Am. J. Card.* 54:908–912, 1984.
24. Hughes, J.R.: Psychological effects of habitual aerobic exercise: a critical review. *Preventive Medicine* 13:66–78, 1984.
25. Montgomery, W.A., Jones, G.E., and Hollandsworth, J.G.: The effects of physical fitness and exercise on cardiac awareness. *Bio. Psych.* 18:11–22, 1984.
26. Perri, S.: The effects of an aerobic exercise program on psychological variables in older adults. *Int'l. J. Aging and Human Development* 20:167–172, 1984–85.
27. Detrano, R., and Froelicher, V.F.: A logical approach to screening for coronary artery disease. *Annals Int. Med.* 106:846–852, 1987.
28. Froom, J., Boisseau, V., and Sherman, A.: Selective screening for lead poisoning in an urban teaching practice. *J. Fam. Pract.* 65:9, 1979.
29. Breslow, L., and Sommers, A.R.: The lifetime health monitoring program: A practical approach to preventive medicine. *N. Engl. J. Med.* 296:601, 1977.
30. Hartley, L.H., Herd, J.A., Day, W.C., Abusamra, J., and Howes, B.: An exercise testing program for large populations. *JAMA* 241:269, 1979.
31. Bruce, R.A., and McDonough, J.R.: Stress testing in screening for cardiovascular disease. *Bull. NY Acad. Med.* 45:1288–1305, 1969.

32. Froelicher, V.F., Thomas, M.M., Pillow, C., and Lancaster, M.C.: Epidemiologic study of asymptomatic men screened by maximal treadmill testing for latent coronary artery disease. *Am. J. Coll. Cardiol.* 34:770–776, 1974.
33. Bruce, R.A., DeRouen, T.A., and Hossack, K.F.: Value of maximal exercise tests in risk assessment of primary coronary heart disease events in healthy men. *Am. J. Cardiol.* 46:371, 1980.
34. Aronow, W.S., and Cassidy, J.: Five year follow-up of double Master's test, maximal treadmill stress test, and resting and post exercise apexcardiogram in asymptomatic persons. *Circulation.* 52:616–618, 1975.
35. Cumming, G.R., Samm, J., Borysyk, L., and Kich, L.: Electrocardiographic changes during exercise in asymptomatic men: 3-year follow-up. *Can. Med. Assoc. J.* 112:578–581, 1975.
36. Allen, W.H., Aronow, W.S., Goodman, P., and Stinson, P.: Five-year follow-up of maximal treadmill stress test in asymptomatic men and women. *Circulation* 62:522–527, 1980.
37. Manca, C., Dei Cass, L., Albertini, D., Baldi, G., and Visioli, O.: Different prognostic value of exercise electrocardiogram in men and women. *Cardiology* 63:312–319, 1978.
38. McHenry, P.L., O'Donnell, J., Morris, S.N., and Jordan, J.J.: The abnormal exercise electrocardiogram in apparently healthy men: a predictor of angina pectoris as an initial coronary event during long-term follow-up. *Circulation* 70:547–551, 1984.
39. MacIntyre, N.R., Kunkler, J.R., Mitchell, R.E., Oberman, A., and Graybiel, A.: Eight-year follow-up exercise electrocardiograms in healthy middle-aged aviators. *Aviat. Space Environ. Med.* 52:256–259, 1981.
40. Multiple Risk Factor Intervention Trial Research Group: Exercise electrocardiogram and coronary heart disease mortality in the Multiple Risk Factor Intervention Trial. *Am. J. Cardiol.* 55:16–24, 1985.
41. Gordon, D.J., Ekelund, L.G., Karon, J.M., et al.: Predictive value of the exercise tolerance test for mortality in North American men: The Lipid Research Clinics Mortality Follow-up Study. *Circulation* 74:252–261, 1986.
42. Froelicher, V.F., Thompson, A.J., Wolthuis, R., Fuchs, R., Balusek, R., et al.: Angiographic findings in asymptomatic aircrewmen with electrocardiographic abnormalities. *Am. J. Coll. Cardiol.* 39:32–38, 1977.
43. Erikssen, K., Enge, I., Forfang, K., and Storstein, O.: False positive diagnostic tests and coronary angiographic findings in 105 presumably healthy males. *Circulation* 54:371–376, 1976.
44. Kemp, H.G., Kronmal, R.A., Vlietstra, R.E., and Frye, F.L.: Seven year survival of patients with normal or near normal coronary arteriograms: A CASS registry study. *J. Am. Coll. Cardiol.* 7:479–483, 1986.
45. Diamond, G.A., and Forrester, J.S.: Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. *N. Engl. J. Med.* 300:1350–1358, 1979.
46. Hlatky, M.A., Mark, D.B., Harrell, F.E., Lee, K.L., Califf, R.M., and Pryor, D.B.: Rethinking sensitivity and specificity. *Am. J. Cardiol.* 59:1195–1198, 1987.
47. Bruce, R.A., DeRouen, T.A., and Hossack, K.F.: Pilot study examining the motivational effects of maximal exercise testing to modify risk factors and health habits. *Cardiology* 66:111–22, 1980.
48. Blomqvist, C.G., et al.: Guidelines for Exercise Testing. A report of the American College of Cardiology/American Heart Association Task Force on assessment of cardiovascular procedures (Subcommittee on Exercise Testing). *JACC* 8:(3) 725–738, 1986.

Bibliography

Arntzenius, A.C., Kromhout, D., Barth, J.D., et al.: Diet, lipoproteins, and the progression of coronary atherosclerosis. *N. Engl. J. Med.* 312:805–811, 1985.

Blair, S.N., Goodyear, N.N., Gibbons, L.W., and Cooper, K.H.: Physical fitness and incidence of hypertension in healthy normotensive men and women. *JAMA* 252:487–490, 1984.

Campbell, M.J., Browne, D., and Waters, W.E.: Can general practitioners influence exercise habits? *Brit. Med. J.* 290:1044–1046, 1985.

DeBacker, G., Kornitzer, M., Solbolski, M., Dramaix, M., et al.: Physical activity and physical fitness levels of belgian males aged 40–55 years. *Cardiol.* 67:110–128, 1981.

Eichner, E.R.: Exercise and heart disease. *Am. J. Med.* 75:1008–1023, 1983.

Froelicher, V.F., and Brown, P.: Exercise and coronary heart disease. *J. Card. Rehab.* 4:277–288, 1981.

Gilli, P., Vitali, D.P., Tataranni, G., and Farinelli, A.: Exercise-induced urinary abnormalities in long-distance runners. *Int. J. Sports Med.* 5:237–240, 1984.

Kittel, F., Kornitzer, M., DeBacker, G., Dramaix, M., et al.: Type A personality in relation to job-stress, social and bioclinical variables: The Belgian physical fitness study. *J. Human Stress* 37–45, 1983.

Kushi, L.H., Lew, R.A., Stare, F.J., Ellison, C.R., et al.: Diet and 20 year mortality from coronary heart disease. *N. Engl. J. Med.* 312:811–818, 1985.

Lindskog, B.D., Sivarajan, E.S.: A method of evaluation of activity and exercise in a controlled study of early cardiac rehabilitation. *J. Cardiac. Rehab.* 2:156–165, 1982.

Morrison, J.F., Van Malsen, S., and Noakes, T.D.: Leisure-time physical activity levels, cardiovascular fitness and coronary risk factors in 1015 White Zimbabweans. *S. Afr. Med. J.* 65:250–256, 1984.

Oldridge, N.B.: Efficacy and effectiveness: Critical issues in exercise and compliance. *J. Cardic. Rehabil.* 4:119–123, 1984.

Patch, L.D., and Brooks, G.A.: Effects of training on VO_2 max and VO_2 during two running intensities in rats. *Pflugers Arch.* 386:215–219, 1980.

Ribeiro, J.P., Hartley, L.H., Sherwood, J., and Herd, J.A.: The effectiveness of a low lipid diet and exercise in the management of coronary artery disease. *Am. Heart J.* 108:1183–1189, 1984.

Shephard, R.J.: Adaptation to exercise in the cold. *Sports Medicine* 2:59–71, 1985.

Siscovick, D.S., Weiss, N.S., Fletcher, R.H., and Lasky, T.: The incidence of primary cardiac arrest during vigorous exercise. *N. Engl. J. Med.* 311:874–877.

Waller, B.F.: Sudden death in midlife. *Cardiovascular Med.* 1:55–59, 1985.

William, D.H., and Williams, C.: Cardiovascular and metabolic responses of trained and untrained middle-aged men to a graded treadmill walking test. *Brit. J. Sports Med.* 17:110-116, 1983.

Morgan, K., Hughes, A.O., and Philipp, R.: Reliability of a test of cardiovascular fitness. *Intn'l. J. Epidemiol.* 13:32–37, 1984.

Smolander, J., Louhevaara, V., and Oja, P.: Policemen's physical fitness in relation to the frequency of leisure-time physical exercise. *Int. Arch. Occup. Environ. Health* 54:295–302, 1984.

Cooper, K.H.: A means of assessing maximal oxygen intake. *J. Am. Med. Assoc.* 203:201–204.

Uhl, G.S., and Froelicher, V.: Screening for asymptomatic coronary artery disease. *J. Am. Coll. Cardiol.* 3:946–955, 1983.

Chapter **26**

Epidemiologic Uses of General Population Assessments of Physical Activity Patterns

Robert T. Hyde, M.A.
Ralph S. Paffenbarger, Jr., M.D.
Stanford University

Introduction

This report sketches briefly the development of epidemiological studies of physical activity patterns in relation to health, showing that in recent decades the field has taken on new importance. Much interest has centered on trends of physical activity and chronic disease, particularly relationships of sedentary lifestyle to coronary heart disease. Findings as to the latter have led toward broadening of investigations in many directions, such as studying effects of aging, athleticism, sociocultural influences, and interrelationships with various aspects of lifestyle and environment.

To illustrate these developments, a number of continuing studies are reviewed in some depth in a section entitled Epidemiologic Studies of Physical Activity Levels and Chronic Disease, with discussion of both occupational and leisure-time physical activities. Next follows a section entitled Definitions and Characteristics of Physical Activity, where a schema of cause and effect in the pathogenesis of coronary heart disease and a tabulation of benefits and hazards of exercise lead directly to discussion of current proposals for practical definitions and assessments of physical activity in relation to survey planning.

Mutual interests of epidemiologic studies and national surveys are considered, so as to develop recommendations for future efforts. These have been emphasized throughout the report and are listed at the end. Meanwhile an effort was made to avoid offering various details that should be the province of other reports on special areas of interest.

Development of epidemiologic studies of physical activity

Today we often think of the epidemiologic study of exercise and health as a relatively new field, but the importance of physical activity to health is by no means a recent discovery. Hippocrates and other writers of antiquity observed that lack of exercise was detrimental to health, and overexertion also unwise.

In 1700, Bernadino Ramazzini (1) described the occupational health hazards of sedentary tailors as contrasted to the relative well-being of fleet-footed messengers. Because tailors could not expect to get enough exercise at work, said he, "they should be advised to take physical exercise at any rate on holidays. Let them

---

Note: Supported by Grant HL37174 from the National Heart, Lung, and Blood Institute of the U.S. Public Health Service; and by the Marathon Oil Foundation, Inc.; the E.I. du Pont de Nemours Company; and the G. Unger Vetlesen Foundation.

make the best use they can of some one day, and so to counteract the harm done by many days of sedentary life." Today we prescribe leisure-time sports play to our millions of workers whose sedentary jobs with high technology deny them much opportunity for occupational exercise.

In 1843, the Royal Statistical Society published W. A. Guy's tabulations of morbidity rates of various diseases differing between sedentary and physically active workers (2). He too recommended leisure-time exercise to promote general health and to counteract the adverse effects of confining jobs. But in 1863, Dr. Edward Smith found London tailors no more healthy than Ramazzini had described them 160 years earlier (3). Smith's tables of age-specific mortality rates showed tailors had 50–75 percent greater mortality in midlife than physically active farmers.

By 1939, considerable attention was being directed to coronary heart disease (CHD), which had been clinically defined by J.B. Herrick in 1912. *Public Health Reports* published an analysis by O.F. Hedley of 5,116 cardiovascular deaths in Philadelphia during 1933–37 (4). Hedley found death rates higher among professional and managerial men than among laborers, but he did not infer that differences in physical activity level might affect cardiovascular health.

When President Dwight D. Eisenhower suffered a myocardial infarction in the 1950's, his physician Paul Dudley White of Boston was already noted as an advocate of physical exercise for health maintenance. White promoted biking and hiking for young and old, but advised Eisenhower to return to golf. In 1958 Drs. W. C. Pomeroy of Los Angeles and White published a report on CHD in former football players, which stated: "The most interesting finding concerned the amount of exercise taken habitually during the lifetime of these men. Those in the coronary group engaged in less vigorous exercise than did the others, and no individual in this study who maintained a heavy exercise program happened to develop coronary heart disease." (5)

Meanwhile, in 1953, Professor J. N. Morris and associates had published a report pointing out the inverse relationship of occupational physical activity levels and CHD, contrasting London bus drivers and conductors (6). This finding led to a series of studies of both occupational and leisure-time exercise patterns and their implications for CHD. Similar investigations were soon underway in the United States and other countries, because of great concern over the sharp rise in CHD incidence, especially in developed nations where sedentary occupations had become predominant through mechanization and other advances in technology.

Some of these continuing studies will be described below, as well as their relationships to the development of still further interests in the epidemiology of exercise. We now realize that this area of medical research has important applications to the extensive contemporary problems of health maintenance, aging, health costs containment, facilities development and utilization, preventive medicine, habit-pattern intervention, sports medicine, exercise technology, and other matters of daily concern in the lifestyles of modern society.

The current proposal to undertake broad population surveys of physical activity patterns is a timely move arriving none too soon. Physical exercise appears to have become a major interest throughout the U.S. population, for health, for enjoyment of leisure, as a popular lifestyle, as high fashion, and as big business. So much is involved that we urgently need a sharper picture of what is going on and what it means to our central interest, the public's health. Few suppose any longer that we are dealing with a mere fad. Instead the so-called physical exercise boom is considered one of the most striking social developments of the 20th century. Thoughtful observers believe this trend and its effects will extend on into the 21st century as a health legacy to our children's children.

Epidemiologic studies of physical activity levels and chronic disease

Occupational physical activity

Epidemiologic work contrasting physical activity and incidence of CHD originated in studies of London transport workers, postal employees, and British civil servants (6). Morris's 1953 report pointed out that London double-decker bus conductors expended more energy at work than the bus drivers, and they had a lower rate of first clinical episodes of CHD, a reduced case-fatality ratio, and a lower rate of early mortality from this disease. Likewise, mail carriers had lower rates of CHD than sedentary postal supervisors and telephonists. A gradient of risk appeared for workers intermediate in physical activity levels. Because the civil service personnel held sedentary desk jobs, their CHD incidence rates were influenced by their habits of vigorous sports play, yardwork, do-it-yourself projects, and other leisure-time exercise. These British study populations afforded excellent availability, stability, and access to records, but some study categories and baselines were yet to be well defined. Eventually a provisional hypothesis was formulated:

> Physical activity of work is a protection against ischaemic heart disease. Men in physically active jobs have less ischaemic heart disease during middle age, what disease they have is less severe, and they tend to develop it later than similar men in physically inactive jobs.

In 1957, H. J. Montoye and associates at Michigan State University published a monograph on the longevity and morbidity of college athletes who had earned varsity letters there prior to 1938 (7). Although dealing with a limited study population, this report was a remarkable example of careful planning and analysis, and a forthright degree of caution was displayed in its conclusions.

Because of small numbers, few distinctions could be established between the lettermen and their nonvarsity classmate control subjects as to longevity, causes of death, or incidence of various diseases, including CHD, but a number of interesting tendencies were noted. As would be expected, both athletes and nonathletes reduced their postcollege activity in vigorous sports with advancing age, but they increased their levels of moderate or light exercise. A larger percentage of athletes remained active in one or more sports up to age 45; but, surprisingly, thereafter the nonlettermen predominated. At this same point the ex-athletes began to surpass the nonathletes in weight gain. Athletes were more likely to die violent deaths, but at 10 years younger than nonathletes, perhaps because their military service tended to increase in wartime. The Michigan State University report is of further interest for its comparative summaries of earlier studies of the health of former athletes, including a sequence dating back to 1873 that dealt with the fates of former oarsmen of British and American university crews (8). The results of these investigations of physical activity levels and health of athletes suggest that a general population survey on physical activity might well include identification of athletes versus nonathletes in the survey population. It might also attempt to assess what proportions of physical activity actually qualify as exercise in the clinical sense, and how the various types of physical activity relate to health (coronary heart disease risk or other concerns).

A study of occupational physical activity among San Francisco longshoremen (9) utilized records of a wide range of energy expenditures measured in actual on-the-job situations. This was an attempt to quantify the specific levels of energy output (in kilocalories per work week) encountered in various job assignments, as a refinement over assessment of physical activity merely by job title. Selection of a longshoring career might or might not imply inheritance of a sturdy physique, but union rules required all longshoremen to work at heavy cargo-handling jobs for at least their first 5 years of service. Many men continued at that level much longer, as the average was 13 years, before shifting to less strenuous duties.

Energy output calculations for the dockworkers included allowances for mandatory rest periods and other slack times between heavy work episodes. Annual adjustments for transfers altering energy demand produced little change in findings for the 22-year followup period. Workers expending 8,500 or more kilocalories per week had less risk of fatal CHD at any age than those less active. The difference was strongest among the younger men but important at all ages studied (35–75 years). Longshoring was for most their life-long occupation, and so demanding that their leisure-time exercise was found to be minimal.

In some studies of less energetic work situations, failure to assess leisure-time activities has appeared to mask or confound effects of job differences (10, 11). Or, conversely, lack of job data has affected studies of leisure activity (12). Thus it would seem important that any general population survey of physical activity include inquiry into both occupational and leisure-time pursuits of all individuals of working age who are tallied, including retirees.

The association of vigorous job physical activity with lower risk of fatal CHD among the longshoremen is at least partly independent of other predictors. The relationship holds whether or not high energy output was accompanied by heavy cigarette smoking, obesity, hypertensive tendency, abnormal glucose metabolism, elevated blood cholesterol, prior CHD or combinations of these. Although selective influences such as severity of CHD may have been involved, energetic cargohandlers with a prior attack had only one-half the risk of subsequent fatal onset found for less active workers with diagnosed heart disease. Studies of exercise programs for rehabilitation of CHD patients have often been hampered by problems of severity, dropout, noncompliance, small numbers, and the like (13). A general population survey of physical activity should include attention to health status reasons for choices or rejections of types and amounts of exercise, whether on a job or at leisure. The findings could have value for motivational and prescriptive improvement projects.

Sports play, exercise, and other leisure-time physical activity

Technological changes have largely abolished strenuous jobs such as cargohandling and heavy construction labor, and henceforth perhaps high energy output will be found mostly in certain sports or some spare-time activities like digging in the yard. Therefore, studies of leisure-time physical activity are more important than ever, although care must be taken not to ignore possible influences of occupational circumstances in even the most modern of settings. For example, proposals are already extant recommending that work situations be designed to promote increases in incidental exercise, as by planning plant and office layouts that involve moderate use of walks, cloisters, ramps, and stairs instead of almost total use of elevators, escalators, motorized walkways, and vehicles (14). The presence and usage of such devices may already be lifestyle elements that should be surveyed and assessed. Most individuals employed in sedentary occupations will have to obtain adequate regular exercise by devoting some of their leisure time to it, as many are believed to be doing. This is both an assumption and a conclusion of current studies of physical activity, because so many jobs do not require or even permit strenuous exertions during working hours.

In 1968–70, leisure-time exercise levels were surveyed among 17,944 sedentary British civil servants, using a 2-day diary report form (15). On a Monday morning, these employees were requested to log their activities of the preceding Friday and Saturday (constituting 1 work day and 1 free day). In 8.5 years of followup, there were 1,138 first clinical episodes of

CHD, 475 fatal, in the total population. Among men active in VE (vigorous exercise) sports, as compared with rates for less active men, relative risks were 0.45 for first clinical episode, 0.38 for fatal first attack, and 0.41 for sudden death. Similar trends were found among retired civil servants.

In this study population, CHD mortality at older ages was 30 percent greater among men who reported no VE activities than among those continuing VE (16). The pattern was even more divergent for CHD morbidity. Overall CHD incidence showed age-specific rates contrasting between men who continued VE sports activity and men who did not. This finding leads to interest in relationships of exercise to the aging process. By maintaining the vitality of the cardiovascular, respiratory, and other body systems, and perhaps thus fending off disease, adequate exercise may tend to preserve homeostasis and improve the quality of life in old age (17). There are suggestions that avoidance or postponement of rest-home expenses could result in extensive savings and other benefits. A general population survey of physical activity should include careful attention to subgroups of elderly and aged persons as well as others, keeping in mind that particular types and levels of exercise are likely to have different influence and importance for older than for younger individuals (18).

The salutary effect of VE sports activity among the British civil servants is partly independent of contrary influences such as heredity, somatotype, cigarette smoking, cardiovascular disease, hypertension, and diabetes mellitus. Men with or without any of these predisposing characteristics had lower risk of CHD if they played VE sports than if less active (16).

Leisure exercise patterns and health status of U.S. college alumni have shown how their past and contemporary exercise levels relate to CHD risk (19, 20). Men who entered college during 1916–50 were studied for personal and lifestyle characteristics of their college and postcollege days, their continuing exercise habits, and any physician-diagnosed CHD, with appropriate followup. Records of first attack of CHD among 16,936 Harvard University alumni during 10 or 6 years (1962 to 1972 or 1966 to 1972) permitted calculation of age-adjusted rates per 10,000 man-years and relative risks dichotomized by levels of assessed physical activity.

Men who reported climbing at least 50 steps per day had 20 percent lower risk of CHD than men who climbed less. Those walking at least 5 blocks daily had 21 percent lower risk than men walking less. Alumni listing only light sports had as much risk as those reporting no sports at all, but the 40 percent who said they engaged in vigorous sports activity had 27 percent lower risk of CHD than men who did not.

A physical activity index expressed in kilocalories per week was totaled from standard output equivalents (in kilocalories per time unit) for stair climbing, city block walking, and participation in light or vigorous sports. Men whose index was at least 2,000 kcal/week (who represented 40 percent of the man-years of observation) had 39 percent less risk of CHD than men with lower index.

Age-specific rates of CHD declined consistently with increase in energy expenditure as assessed by each activity measure—stair climbing, block walking, sports play, and the composite physical activity index. Findings were similar for both nonfatal and fatal clinical events (angina pectoris, myocardial infarction, and sudden death). Overall CHD risk patterns were comparable in all age classes studied, 35–74 years. Adequate habitual physical activity in adult years was influential rather than hereditary constitution or vigorous youth alone.

Investigators routinely expect an increased selective influence to confound assessments of physical activity effects among the elderly and in studies of secondary prevention among CHD patients, who may reduce or eschew exercise because of illness or fear of illness. Nevertheless, after making reasonable allowance for such self-selections, there is still persuasive evidence that the moderately active have lower risks of disease or death than the inactive. Going to higher levels or more strenuous forms of exercise, however, does not add further benefit in these segments of the population. Determination of optimum levels and types of physical activity or exercise for various categories of people has become a primary objective of many contemporary studies. The findings from a general population survey of typical exercise patterns could be very helpful to the planning of epidemiological investigations of that kind and others.

A followup of Harvard alumni with reported CHD showed that those who took up or continued physical activity levels of 2,000 or more kcal/week combined in their habitual walking, stair climbing, and sports play had a CHD mortality rate only 71 percent as high as those who were less active (20). Among men initially free of CHD and classified active or inactive by the 2,000 kcal/week breakpoint, the actives aged 55–64 and 65–74 had, respectively, 51 percent and 60 percent as high a risk as the inactives in their age groups of having onset of nonfatal or fatal CHD in the 6–10 years of followup. In the younger age groups, 35–44 and 45–54 years, the actives had about 25 percent less risk than inactives. The differing percentages for the younger and older age groups appear to imply an increased selective influence among the latter, which might be expected. Also, rates were typically low among the younger men. The interrelationships of disease, aging, and physical activity are of prime concern to epidemiologists studying appropriate subgroups of the general population. The proposed NHANES III survey of physical activity should be designed to obtain data that will facilitate those investigations.

Because hypertension is a notorious strong predictor of CHD, a longitudinal study was conducted among Harvard alumni to assess any relationships of physical activity to risk of developing hypertension (21). Mild exercise was unrelated to age-adjusted rates of hypertension, but normotensives reporting vigorous sports activity were at 27 percent lower risk than men who did not. This association persisted in age groups 35–74 and with adjustment for differences in age and body mass index, but a multiple-logistic regression analysis showed the reduction was significant only for men who were at least 25-percent overweight for their height. Among the latter, hypertension risk was dose dependent on vigorous sports-play, being cut a quarter or in half, respectively, by 2 and 4 hours per week of such exercise. Very likely the actives tended to be muscular and the inactives fat, though rated the same by body mass index. This is another example of a specialized epidemiological study with findings that should be considered in planning a general population survey of physical activity. Data on kinds of exercise in relation to height and weight, etc., may help identify segments of the population most at risk of hypertension or assist in the development of intervention programs. Results of proposed skinfold and muscularity tests might be of interest here.

The broad emphasis on relationships of physical activity to alteration of CHD risk has tended to obscure and delay, though at times to encourage, attention to any associations with other diseases or physical conditions. There is considerable interest in regard to diabetes, arthritis, osteoporosis, cancer, aging, pediatrics, and mental illnesses, but few investigations of physical activity aspects of these have employed the approach of longitudinal epidemiological studies centered on exercise and CHD risk (22, 23).

Definitions and characteristics of physical activity

Recently some of this work has brought a new focus on studying the values of different kinds and intensities of physical activity, such as the effects of weight-bearing exercise on bone structures, stretching movements on flexibility, or reflex actions on balance. The health benefits of physical activity are being newly regarded as not simply a matter of something called fitness or $\dot{V}O_2$max but a complex process taking effect through all mechanisms and systems of the body. Although some relationships of various aspects of physical activity to specific health entities are yet to be defined, there is increasing evidence that adequate exercise may reduce or at least postpone risk of all-cause mortality, thus extending longevity (25). This itself is testimony that physical activity levels and characteristics must have many subtle and salutary influences on resistance to diseases and aging, whether or not we yet know what all of them are.

Mechanisms of exercise benefit

In the past, there has been much discussion of a now classic riddle as to whether the salutary relationships of exercise or physical activity to CHD risk are truly protective or merely the impressions from a selective process in which unhealthy persons, from inability or despair, elect not to be physically active (26, 27). Similar issues often addressed are questions whether the beneficial effects observed are caused by exercise or by some coincidental process such as cholesterol modification. The choices of answers to these problems are not as arbitrary as some would suppose. If blood lipid profile can be altered by either diet or exercise or both, the likelihood remains that cholesterol change is one of the mechanisms or processes through which exercise lowers risk of CHD (28). Exercise is known to alter other consequences of diet as well, not a few of them similarly related to CHD.

Once it is accepted that physical activity may be a valid health regimen against cardiovascular and other chronic diseases, attention begins to focus on specific benefits and hazards of exercise. This leads to renewed efforts to define the elements of physical activity and classify the various kinds of exercise according to their mechanical and physiological characteristics (29, 30). Much of the information thus far has come from clinical data, but at least the epidemiological studies of exercise to date have shown that physical activity is a highly complex proposition involving at the population or community level many variables of importance that may not be fully appreciated at the individual or clinical level.

In planning a general population survey of physical activity such as NHANES, there should be an effort to incorporate interests from both levels, community and clinical, but the overall approach of the survey seems likely to be patterned after epidemiological rather than clinical models, if only because of the scale of the operation.

Fitness considerations

Physiological effects exist whether or not the physical activity or exercise is deliberate. Epidemiologic studies have shown that physical activity is influential, but they have not yet fully assessed what kinds of physical activity have the most influence (either beneficial or harmful). Perhaps they cannot get at some specific as well as clinical studies, or need not attempt to.

Usefulness of a particular concept of fitness may be a case in point, because that often has been sought as a kind of yardstick for measuring the presumed benefits of exercise, particularly with regard to CHD risk but lately as to status of aging and other matters. At the epidemiological level it is not always necessary or productive to rely on a numerical measure or even a set of them. Instead, a broad empirically rated concept such as *hardiness* or *vitality* may be serviceable if it is defined in terms that general population respondents can use to evaluate it. Epidemiological analyses often combine sets of specific questionnaire responses or test data into a composite index value, such as estimated kilocalories per week of physical activity energy output. Once a study has been appropriately designed, it may be reasonable to assume that those study subjects who exercise habitually to above-median levels of energy output per week are probably more fit than those who do not. If that is true, then for purposes of the epidemiological study or its conclusions, the need of a specific definition of fitness, as such, might be academic.

Nevertheless, the NHANES should assess several kinds of fitness. The methods to be employed will depend on the definitions of fitness involved and the objectives to be sought, as well as on matters of convenience and economy. Heightened interest in sports medicine may tend to overemphasize attributes of athleticism, and conservative interests in health may limit the purview in this and other directions. The NHANES III may provide a timely opportunity to avoid aberrations and simplify test procedures. For some years, there has been much attention to cardiorespiratory endurance fitness and body composition estimates, but those still have not achieved ease of measurement in large populations. Aside from studies of athletes and school children, little seems to have been done with general testing for muscularity, musculoskeletal flexibility, and neuromuscular requirements for reflex and balance, yet all of these would seem important to the physical fitness of anyone today. Muscular development as achieved by weight training or isometric workouts is not an index to cardiovascular fitness or cardiorespiratory endurance, but tests of muscularity might provide better estimates of body composition than skinfold or circumference measurements, or supplement the latter to advantage. Fitness also should be defined and assessed in terms of the reflexes and other split-second physiological demands that are being made on more and more people who live in a world of quick-change situations, fast cars and aircraft, high-tech communications, ultramodern appliances, and heightened sports activities. These new lifestyles and their stresses must be met with appropriate standards of individual and community health. New appraisals are also needed of the health requirements of the elderly and aged whose numbers and longevity are increasing. Tests of reflexes and balance should be quick and easy to conduct, yet might reveal much.

Information on fitness testing must be ample in the clinical literature, but pilot trials could be useful too. Then, in the general survey, fitness estimates might be developed by using selected threshold values or breakpoints between which highly accurate measurements would be unnecessary. Applications of these standards would vary for different age groups and subsets of the population. The extensive fitness testing contemplated for the NHANES may make important contributions to general health monitoring. The results will be of special interest in relation to survey findings on patterns of physical activity and disease.

Development of standards

Both observed physical activity levels and specific chronic disease data influence practical definitions and evaluations of exercise, and obviously they should. This becomes evident in considering the implications of table 1, a schema of causes and effects of the pathogenesis of coronary heart disease. Such a tabulation is a convenient way of depicting a complex set of relationships. Like a flow chart, it displays the steps or stages in a process or sequence of events that could be almost any disease. This particular schema represents CHD from the epidemiological point of view, and it might be somewhat different if intended to set forth only clinical aspects of CHD. The present schema of CHD pathogenesis involves considerations alterable by physical activity or exercise sufficiently to reduce their contributions to CHD risk.

The schema begins by pointing out that the family background of a person may affect his risk of developing CHD, either directly by inherited physical characteristics or through influences of location, family lifestyle elements, and sociocultural patterns (column 1). City life differs from country life. Parents and siblings may or may not smoke cigarettes or dine on fatty foods. Neighbors and relatives may promote tennis or video games. Such opportunities and inducements alter any youngster's tendencies in his formative years, whether toward healthful exercise or toward increased likelihood of developing overt CHD by middle life. Column 2 of the schema lists several adverse elements of individual lifestyle.

Thus the schema can be read from left to right, column to column as the pathogenesis of CHD proceeds, and vertically within columns. So in column 2 imprudent diet and sedentary living lead to inadequate weight control, but also patterns of influence may spread in more than one direction. Obesity often tends to discourage vigorous exercise, and not a few heavy individuals cite their condition as an excuse to continue cigarette smoking lest they gain additional weight. Dyspnea, fatigue, hypertension, and vague feelings of discomfort or pain are likely to reduce physical activity, with or without the advice of a physician. There are many such pathways of causation within this schema, but its general trend is a progression from early background influences, to subtle changes, to overt disease and its endpoint consequences. All the elements listed in the schema are adverse developments that contribute to a worsening situation. Though not all are to be found in a single individual, they will be common in a susceptible population, some more prevalent and hazardous than others.

However, when the many salutary effects of adequate physical activity are considered, this schema reveals a two-way street full of opportunities to turn fate around. Most of the developments listed in columns 3, 4, and 5 can be avoided, reversed, or minimized by adjusting any harmful patterns capable of change, particularly those in column 2 and especially as to physical activity. All conditions in column 3 are alterable by adequate exercise, perhaps enough to forestall those in columns 4 and 5.

Unfortunately, too many individuals are likely to arrive at column 6 before recognizing that they may have been developing CHD. Even at that point, however, they may hope to reduce their symptoms and slow the pathogenetic process by adjusting adverse habits and characteristics of lifestyle. Such persons might be prescribed regimens of therapy and behavior modification (including exercise) similar to those used for secondary prevention and the rehabilitation of CHD patients. Experiences of the latter show that, even at column 7 of the schema, there is hope for some deterrent actions, better late than never.

Benefits and hazards of exercise

Although table 1 shows a downward progress toward fatal CHD, a corresponding schema could be designed to present a remedial or preventive pattern. This will be evident from table 2, which lists briefly some of the known or likely benefits and hazards of physical activity, some processes by which it works, and the body systems it influences. Considering physical activity only in the broadest sense and without specifying any types or levels of exertion, table 2 is presented in lieu of a schema. It may be interpreted as dealing with what might be termed adequate exercise, an expression that should always be qualified by its objectives and how well they are attained. The term *beneficial* itself needs definition, as some exercises such as jogging may benefit the cardiorespiratory systems but be hazardous to skeletal joints, perhaps for lack of proper equipment and instruction. Certainly an endurance regimen that required a trade-off between cardiorespiratory benefits and musculoskeletal or neurologic hazards would have to be assessed accordingly. But insofar as physical activity is a natural function of the human body, it should not have to involve making a Hobson's choice.

In terms of endurance exercise needed for maintenance of cardiovascular health, calculations from the College Alumni Health Study (19–21) arrived at the following estimates, on the basis that a normal man should expend at least 4 kilocalories per day per kilogram of body weight. Men weighing 60 kilograms would expend 1,680 kilocalories per week, and a man weighing 100 kilograms would expend 2,800:

Body weight, kilograms	(pounds)	Energy output, kilocalories per week
60	110	1,680
70	155	1,980
80	176	2,240
90	198	2,520
100	220	2,800

Age and other considerations might alter this prescription for any individual, and the definition of endurance

Table 1. Pathogenesis of coronary heart disease: An epidemiologic schema of cause and effect

1 Familial background	2 Lifestyle elements	3 Physiological changes	4 Pathophysiological manifestations
Heredity	Imprudent diet	Metabolic abnormalities	Increased plasma
Physical environment	Sedentary living	Dyslipidemia	LDL cholesterol
Air & water exposures	Occupational patterns	Hyperglycemia	Decreased plasma
Climate	Leisure-time habits	Cardiovascular unfitness	HDL cholesterol
Terrain	Cigarette smoking	Resting tachycardia	Hypertension
Community	Inadequate weight control	Poor left ventricular function	Obesity
Sociocultural exposures	Insufficient coping with psychological stress	Low stroke volume	Electrocardiographic abnormalities
Lifestyle influences		Elevated blood pressure, etc.	Myocardial oxygen insufficiency
Dietary behavior		Increased weight for height	Increased platelet stickiness
Exercise patterns		Unfavorable body composition	Decreased fibrinolytic activity
Social habits		Systemic abnormalities	Altered endocrine receptor activity
Health knowledge		Hematologic	
Health attitudes		Endocrine	
		Neurologic	

5 Pathological conditions	6 Premonitory symptoms	7 Clinical diseases	8 Endpoint categories
Vascular	Dyspnea	Angina pectoris	Disability
Fatty streaks	Fatigue	Coronary insufficiency	Death
Fibrous plaques	Pain	Myocardial infarction	Coronary heart disease
Complicated lesions	Substernal	Ancillary diseases	Sudden
Stenosis	Referred	Cerebrovascular	Delayed
Hemorrhage	Vague discomfort	Peripheral vascular	Underlying
Thrombosis	Palpitations	Aortic aneurysm	Contributing
Occlusion	Dizziness	Diabetes mellitus	Other diseases
Myocardial scarring	Syncope		Brain
Left ventricular hypertrophy	Claudication		Aorta
Neurological	Peripheral edema		Lung
Dysrhythmia			Kidney
Arrhythmia			

exercise would need to be specified as to types and duration, linked usually to some goal of cardiovascular fitness also subject to individual modification.

Renewed attention is being turned to comparative evaluations of different kinds of exercise. More data are needed to ascertain whether longer or more frequent sessions of moderate exercise are as beneficial as shorter sessions of vigorous exercise, and so on. Both clinicians and epidemiologists would like to have further information on the fine points of walking, because that is one of the most common forms of habitual exercise. For reasons like these, a general population survey of physical activity should attempt to find out what kinds and intensities of walking and other exercise are being engaged in by various segments of the population, here interpreting *intensity* to refer to both *style* and *duration* and perhaps also to *frequencies* and *percentages*, all of which affect rates and totals of energy output, subject to individual needs and capacities. One of the prime objectives of the proposed NHANES III general population survey of physical activity should be to obtain key information that is needed to develop better capabilities to prescribe and assess the effects of optimum levels and types of exercise in all segments of the population.

The NHIS 1985 questionnaire touches on these questions in part. Even broader samplings of fitness, according to several definitions of that term, should be included in future data sets, certainly cardiorespiratory and neuromusculoskeletal fitness testing in age-specific and other categories of the population. Moreover, if avoidance or reduction of chronic diseases is taken to be one of the prime reasons for interest in physical activity patterns, data on incidence of such diseases in the same population categories should be obtained for epidemiological study in relation to both physical activity patterns and fitness observations. Fortunately, that effort has already been initiated in the NHIS program.

Implications for health maintenance

Tables 1 and 2 may be interpreted also as expressions of the value and importance of health maintenance motivation programs, by which many of the evils of chronic disease such as those of CHD might be averted or at least delayed. Perhaps a national preventive effort developing through mass persuasion and education is really the ultimate objective or likely result of the general population survey of physical activity being proposed for NHANES III. By its nature, this survey will be dealing chiefly in the most influential mainstream patterns of popular lifestyle. Once these have been surveyed, identified, defined, and evaluated, the neces-

Table 2. Benefits and hazards of physical activity as a defense against cardiovascular and other chronic diseases

Body systems involved	Beneficial effects	Potential hazards
Cardiovascular-respiratory	Enhances physical work capacity, hemodynamic function, hematologic action, and cardiovascular fitness Reduces risks of hypertensive-atherosclerotic Increases maximum breathing capacity	Increases risk of sudden death in susceptible individuals Initiates asthmatic attacks Contraindicated for patients with acute myocardial diseases, myocardial infarction, myocarditis, dissecting aneurysm, severe aortic stenosis, and uncontrolled hypertension
Endocrine-metabolic	Stimulates metabolic processes Influences hormone production Decreases fat body mass and increases lean body mass Improves plasma lipid and lipoprotein fraction profiles Enhances fibrinolytic activity Reduces risk of adult-onset diabetes	Induces temperature imbalance (effects of hyperthermia)
Musculoskeletal	Enlarges and strengthens muscle fibers Increases oxygen utilization Strengthens connective tissues Eases low back pain syndrome Retards osteoporosis Reduces rheumatoid- and osteoarthritic trends (?)	Produces acute and chronic injuries of cartilage, muscle, and bone Falling injuries Overuse syndrome Induces traumatic arthritis (?)
Gastrointestinal	Promotes peristaltic and mixing action of intestine Shortens enteric passage time Modifies appetite	Induces electrolyte imbalance
Neurologic	Influences neuromyocardial action Enhances mood, thought, and psychological behavior	Initiates dysrhythmia Induces compulsive-reactive syndrome (?)

Note: Question marks indicate effects and hazards still open to debate.

sary modes for any recommended adjustments can be studied. A recent sweeping upsurge in leisure-time exercise and sports play appears to have come about largely as a matter of fashion, i.e., an episode of self-education among the people or, perhaps, an instinctive compensating response in a generation where physical activity levels were becoming reduced after eons of experience in which vigorous exertions always had been required for survival. The reasons for this social development of leisure-time liveliness, if it has occurred, as many believe, may offer clues to what more needs to be done to achieve optimum effectiveness in a national health maintenance motivation program.

Physical activity surveys

Table 3 is a brief classification of methods that have been used or proposed for assessing habitual physical activity, many of them in epidemiological studies. The list is not intended to include fitness evaluations as such, so any clinical measurements of that nature are represented as indirect means of assessing physical activity. Most of the entries are well-known general population survey techniques that may have been used by NCHS.

Table 3. Methods of assessing habitual physical activity

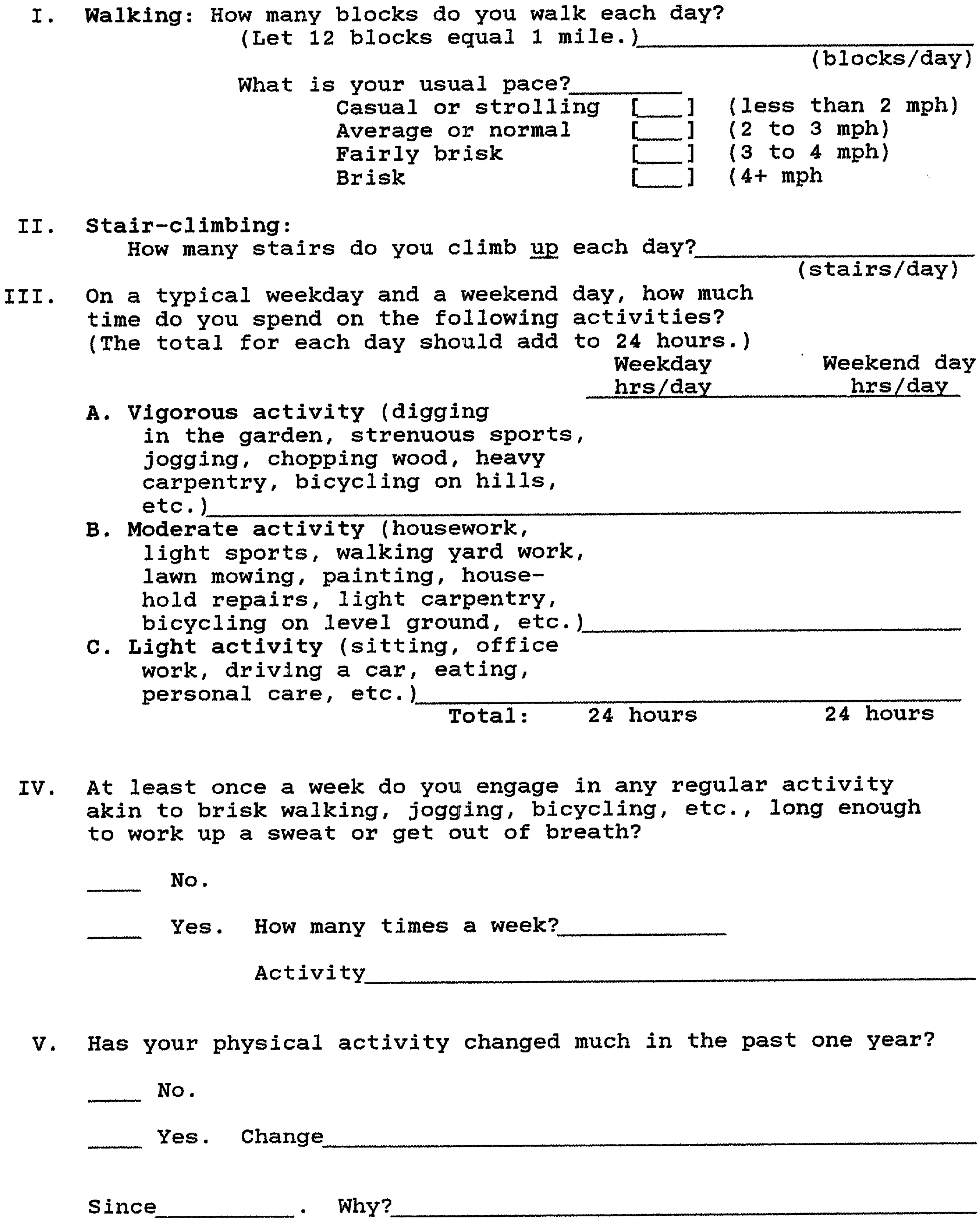
- Direct
  - Questionnaire
    - Self-administered
    - Interviewer-administered
  - Diary annotation
    - Self-recorded
    - Observer-recorded
    - Interviewer-recorded
  - Contact
    - Telephone
    - Mailed invitation or form
    - Door-to-door random survey
    - Institutional
    - Other personal approach
    - Followup techniques
  - Mechanical/electronic motion sensors
    - External (areal detectors, counters, recorders, etc.)
    - Body-carried devices
      - Self-recording
      - Remote-recording (telemetry)
- Indirect
  - Energy-intake (dietary) measurements
    - Questionnaire (A 1 a,b)
    - Diary annotation (A 2 a,b)
    - Food industry statistics
      (production, sales, distribution, consumption)
  - Body weight-for-height measurements
    - Community screening data (school, clinic records)
    - Prior and current questionnaire data
    - Body composition aspects (skin-fold test results, etc.)
    - Disease-related aspects (e.g., cancer effects)
  - Occupational categories (job titles, descriptions, ratings, etc.)
  - Sports participation (community events, clubs, marathons, etc.)
  - Facilities development and use (pools, trails, courts, etc.)
  - Recreational tourism data
    (State, community, commercial; resorts, ski-slopes, etc.)
  - Sports equipment production and sales data
  - Sports medicine clinical data

The data so gathered might become directly useful for epidemiological investigations designed to followup on findings of the survey.

The list is not exhaustive but displays the extensive range and variety of sampling methods available and applicable to physical activity surveying. Those that amass great quantities of minutiae run a risk of being less satisfactory than others that are simpler and more economical, assuming that good design is achieved. Presented as figure 1 is an example of a brief physical activity recall history form such as has been used in a number of epidemiological studies. It might be part of an interview report or a self-administered questionnaire. Other brief sections of the latter would obtain data on age, physique, health history, family, habits, etc.

Choice of sources of information may be highly important when a large study population is involved. Instead of approaching individual subjects, which would have been costly, time consuming, and perhaps impossible, the epidemiological study of San Francisco longshoremen used trade association records of job assignments, which were probably more accurate and complete than recall answers might have been. The study of Harvard alumni telescoped time and experience by drawing on university archives for health and physical activity records of their student days, which were later updated by a series of relatively simple questionnaires mailed to survivors with the assistance of the university alumni office. In a similar fashion the NHANES might include sampling of some indirect sources as a check on the results of direct approach sampling. If it is found from pilot testing that, for some data gathering, a convenient and economical indirect method will be as informative as a cumbersome and costly direct one, then a wise choice can be made.

Whatever sampling method is used, elements of selective bias or partiality are likely to be present. Questionnaire respondents tend to overstate their height and understate their weight (31). The young may exaggerate their self-assessed sense of fitness, and the elderly may depreciate theirs; yet low estimates by the latter reportedly have correlated more strongly than physicians' estimates with subsequent mortality rates (32). Cigarette use and alcohol consumption are often understated, athletic prowess and activity exaggerated, specifics of diet and income unreliable. Devices and protocols for physiological measurements have their own sources of error. Availability of study subjects may alter the representativeness of subsets or samples in a population. Data are confused by confounding influences, e.g., obesity encourages smoking and discourages exercise. Problems of this sort are commonly encountered in epidemiological studies limited by time, staff, budget, facilities, goals, and data availability. Use of two or more independent approaches to the same objective is basic technique for minimizing effects of bias, the rest of which must be dealt with by cogent analysis and interpretation. If the design and scope of the proposed NHANES III are sufficiently broad, its findings may provide new and invaluable help to continuing and future epidemiological investigations of physical activity and health by making available the kinds and quantities of data that are necessary to develop better understanding and control of elements of bias in various study populations. Also such a data base may serve literally as a springboard for launching new epidemiological follow-up studies designed to take full advantage of NHANES III



I. **Walking:** How many blocks do you walk each day?
(Let 12 blocks equal 1 mile.)________________ (blocks/day)

What is your usual pace?________

Casual or strolling	[___]	(less than 2 mph)
Average or normal	[___]	(2 to 3 mph)
Fairly brisk	[___]	(3 to 4 mph)
Brisk	[___]	(4+ mph

II. **Stair-climbing:**
How many stairs do you climb up each day?________________ (stairs/day)

III. On a typical weekday and a weekend day, how much time do you spend on the following activities? (The total for each day should add to 24 hours.)

	Weekday hrs/day	Weekend day hrs/day
A. Vigorous activity (digging in the garden, strenuous sports, jogging, chopping wood, heavy carpentry, bicycling on hills, etc.)		
B. Moderate activity (housework, light sports, walking yard work, lawn mowing, painting, household repairs, light carpentry, bicycling on level ground, etc.)		
C. Light activity (sitting, office work, driving a car, eating, personal care, etc.)		
Total:	24 hours	24 hours

IV. At least once a week do you engage in any regular activity akin to brisk walking, jogging, bicycling, etc., long enough to work up a sweat or get out of breath?

____ No.

____ Yes. How many times a week?__________

Activity____________________

V. Has your physical activity changed much in the past one year?

____ No.

____ Yes. Change____________________

Since__________. Why?____________________

Figure 1. Physical activity recall history.

by extracting further knowledge from extensions of the work it has begun. These might constitute pilot projects for the planning of sequential surveys to be anticipated as health and physical activity continue to be monitored in the future.

Trends and patterns of data

The plans for NHANES and NHIS should continue to take into account the desirability of provisions for a repeat or followup survey some years later, which should attempt to assess the same population subgroups and possibly many of the same persons so as to obtain longitudinal data for study of consistencies and trends. In the interim, links should be maintained with related information from other sources, e.g., census and vital statistics reports. It may be that State agencies could help maintain continuity in these efforts over time, which must be done without infringing on the rights and privacy of any study subjects in the survey population.

Although recall data may not have the same qualities as followup survey data, they do have the timesaving advantage of drawing on retrospective experience instead of waiting for future experience. Sources of error have already been mentioned, but the processes of memory may also include helpful screening and classification activities that naturally identify and emphasize important points. Even the memories of the elderly are known to be remarkably keen at long range though fuzzy at short range. In a large general population survey, the combined effectiveness of recall assessments might be quite sharp and meaningful. Whereas the Nation is considered to be in the midst of sweeping changes in its lifestyles, the NHANES and NHIS should be designed to obtain some historical impressions from respondents, i.e., any notes on how their physical activity patterns may have changed since 5 or 10 years ago or any particular age. Reasons for any changes, up or down, should be sought. Trend information on other topics of interest to epidemiological studies likewise should be obtained, such as on smoking and dietary habits, preferably from the same persons queried on their exercise patterns. In particular, the extent and timing of the recent so-called exercise revolution ought to be carefully documented, especially in terms of its impact on the general population.

A review of the 1985 NHIS questionnaire shows that a number of the suggestions and recommendations presented in this report are already in process of being implemented to some extent by NCHS in collaboration with the U.S. Bureau of the Census. Reports on the genesis and use of the questionnaire in recent years show that NCHS has gained experience from many of the same problems that have been encountered by various epidemiological studies as described in this report. The presence of a firm protocol followed through by a national staff of trained observers is commendable and heartening, as are the overall sampling design and data collection procedures, which are clearly intended to establish and maintain a progressive sequential followup or continuity that would be required for most epidemiological studies contemplating use of the survey findings. Coordination of sampling design adjustments with the decennial census may be somewhat artificial, yet logical if the strategic and logistic advantages outweigh any philosophical misgivings. Of further interest, this same linkage is being extended from NHIS to the other general population surveys conducted by NCHS.

From the standpoint of anticipated epidemiological studies of physical activity and chronic diseases it is most helpful that the general population questionnaire is already organized in terms of specific body systems, each of which is assessed annually in depth in a representative subset of the population. The keying of appropriate questions on physical activity or exercise to these special health interests should not be too difficult or cumbersome. The need for more information about specific kinds and effects of exercise in relation to various chronic diseases has already been mentioned. The two-page section on exercise in the 1985 interview questionnaire is quite detailed in some respects, but might be rather complicated to analyze in relation to data in other sections of the questionnaire such as those on physique, diet, and diseases other than CHD. Questions about motivation could be improved along lines that have been discussed earlier.

Recommendations for NCHS surveys of physical activity

Identify athletes vs. nonathletes in the population.

Include inquiry into both occupational and leisure-time pursuits of all individuals of working age, including retirees.

Check for personal and health status reasons for choices or rejections of types and amounts of exercise both on the job and at leisure.

Check on types of occupational settings and their influences on physical activity opportunity.

Identify exercise patterns of elderly, aged, and other subgroups (e.g., handicapped, disadvantaged, socioeconomic class, adolescent, young adult, sex). Review appropriateness of using subgroups already being surveyed for other reasons.

Survey findings can help establish optimum levels and types of physical activity for health, rehabilitation, planning, and other purposes.

Assess interrelationships of physical activity and lifestyle, personal, and familial characteristics.

Assess physical activity in relation to other chronic diseases as well as CHD. Data on hypertension, emphysema, etc. should be obtained. (See also discussions of tables 1 and 2.)

Attempt to serve both clinical and community health interests.

Consider aspects of preventive medicine, health maintenance, and health education. Broad aspects of general surveys are important and tend to be epidemiological in nature. They can supply both overview and baseline data.

Consider utility of various definitions and types of fitness and their relationships to physical activity and interpretation of physical activity assessments.

Data on kinds and amounts of walking are of special interest.

Obtain survey data on both benefits and hazards of exercise.

Consider both direct and indirect survey methods and data sources.

Continue to obtain data that establish patterns and trends of physical activity, including levels and types, distributions, etc. Historical information (e.g., on the exercise boom) may be of special interest for many purposes. As noted, the so-called exercise boom has been considered one of the most significant social developments of this century, with health implications likely to extend well into the next century or beyond. The proposed survey can make important contributions to knowledge concerning this event.

Survey protocols already in place have important applications to assessments of physical activity and some might be further refined for that purpose.

References

1. Ramazzini B. *Diseases of Workers.* (1700) (Latin) Translated by WC Wright. Hafner, New York. 1964.
2. Guy WA. Contributions to a knowledge of the influence of employments upon health. *J. Roy. Stat. Soc.* 6:197–211, 1843.
3. Smith E. Report on the Sanitary Circumstances of Tailors in London. Rep. Med. Officer Privy Council, with Appendix, pp. 416–430. London, HM Stat Office. 1864.
4. Hedley OF. Analysis of 5,116 deaths reported as due to acute coronary occlusion in Philadelphia, 1933–1936. *U.S. Weekly Public Health Reports* 54, 972 ff, 1939.
5. Pomeroy WC, White PD. Coronary heart disease in former football players. *JAMA* 167:711–714, 1958.
6. Morris JN, Heady JA, Raffle PAB, et al. Coronary heart disease and physical activity of work. *Lancet* 2:1053–1057, 1111–1120, 1953.
7. Montoye HJ, Van Huss WD, Olson HW, et al. *Phi Epsilon Kappa* (Michigan State University, Lansing, MI). 1957.
8. Morgan JE. *University Oars.* London, MacMillan & Co Ltd. 1873.
9. Paffenbarger RS, Hale WE. Work activity and coronary heart mortality. *N Engl J Med* 292:545–550, 1975.
10. Keys A. *Seven Countries: A Multivariate Analysis of Death and Coronary Heart Disease.* Cambridge, Harvard University; pp 196–217, 275–278, 1980.
11. Salonen JT, Puska P, Tuomilehto J. Physical activity and risk of myocardial infarction, cerebral stroke and death; a longitudinal study in Eastern Finland. *Am J Epidemiol* 115:526–537, 1982.
12. Chapman JM, Massey FS. The inter-relationship of serum cholesterol, hypertension, body weight and risk of coronary heart disease: Results of the first ten years follow-up in the Los Angeles Heart Study. *J Chron Dis* 17:933–947, 1964.
13. Shaw LW. Effects of a prescribed supervised exercise program on mortality and cardiovascular morbidity in patients after a myocardial infarction. The National Exercise and Heart Disease Project. *Am J Cardiol* 48:39–46, 1981.
14. Thomas GS, Lee PR, Franks P, et al. *Exercise and Health: The Evidence and the Implications.* Cambridge MA, Oegelschlager, Gunn & Hain. 1981.
15. Morris JN, Chave SPW, Adam C, et al. Vigorous exercise in leisure-time and the incidence of coronary heart-disease. *Lancet* 1:333–339, 1973.
16. Morris JN, Everitt MG, Pollard R, et al. Vigorous exercise in leisure time: Protection against coronary heart-disease. *Lancet* 2:1207–1210, 1980.
17. Fries J, Crapo LM. *Vitality and Aging.* WH Freeman & Co, San Francisco CA. 1981.
18. Hyde RT, Paffenbarger RS. Exercise and Aging. A review for the National Institute on Aging. Stanford University School of Medicine, Stanford CA. 1981.
19. Paffenbarger RS, Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol* 108:161–175, 1978.
20. Paffenbarger RS, Hyde RT, Wing AL, et al. A natural history of athleticism and cardiovascular health. *JAMA* 252:441–495, 1984.
21. Paffenbarger RS, Wing AL, Hyde RT, et al. Physical activity and incidence of hypertension in college alumni. *Am J Epidemiol* 117:245–257, 1983.
22. Siscovick DS, Weiss NS, Fletcher RH, et al. The incidence of primary cardiac arrest during vigorous exercise. *N Engl J Med* 311:874–877, 1984.
23. Siscovick DS, La Porte RE, Newman JM. The disease-specific benefits and risks of physical activity and exercise. *Public Health Reports* 100:180–188, 1985.
24. Taylor CB, Sallis JF, Needle R. The relation of physical activity and exercise to mental health. *Public Health Reports* 100:195–202, 1985.
25. Paffenbarger RS, Hyde RT, Wing AL. Physical activity, all-cause mortality, and longevity of college alumni. *New Engl J Med* 314:605–613, 1986. Letters to the Editor, *New Engl J Med* 315:399–401, 1986.
26. Leon AS, Blackburn H. Physical activity in the prevention of coronary heart disease: An update, 1981. In Arnold CB, Kuller LM, Greenlick MR (eds.), *Advances in Disease Prevention.* New York, Springer Verlag. 1981.
27. Solomon HA. *The Exercise Myth.* San Diego, Harcourt Brace Jovanovich. 1984.
28. Wood PD, Haskell WL, Blair SN. Increased exercise level and plasma lipoprotein concentrations: A one-year, randomized, controlled study in sedentary middle-aged men. *Metabolism* 32:31–39, 1983.
29. Haskell WL. Physical activity and health: Need to define the required stimulus. *Am J Cardiol* 55:4D–9D, 1985.
30. Haskell WL, Superko R. Designing an exercise plan for optimal health. *Fam & Com Hlth,* 72–86, May 1984.
31. Paffenbarger RS, Wolf PA, Notkin J, et al. Chronic disease in former college students: I. Early precursors of fatal coronary heart disease. *Am J Epidemiol* 83:314–328, 1966.
32. Davies AM. Epidemiology and the challenge of aging. *Internat J Epidemiol* 14:9–19, 1985.

Part VII

Measurement and Analysis Strategies

Chapter 27

Use of Latent Variable Models in Measuring Physical Fitness and Physical Activity

George W. Bohrnstedt, Ph.D.
Indiana University

Joseph Lucke, Ph.D.
Lehigh Valley Hospital Center

Measurement is a sine qua non of any science. In general one may discern two broad approaches to the definition and measurement of scientific constructs. The first, which we call the *operational* approach, is to *define* the construct by the method for measuring it. An example would be to define aerobic fitness as $\dot{V}O_2$ max or to define body fat content by the hydrostatic procedure, which is considered the "gold standard." We find this approach to defining and measuring theoretical constucts too restrictive. The concept of fitness or body fat content contains more implications than those contained in the $\dot{V}O_2$ max or hydrostatic measurement procedures. If other methods, such as skin caliper methods and ultrasound methods, are proposed for measuring body fat content, how does one adjudicate the claims that all three methods measure body fat, when the hydrostatic method defined body fat content? The hydrostatic method is easily preferred to the skin caliper method, as the former is the more theoretically and physiologically sound method. But between the hydrostatic and ultrasound methods, the preference is less clear as the ultrasound method may be preferred on theoretical and physiological grounds, but the hydrostatic method has prior claim to being the defining standard. Which of these measurement methods is the true gold standard?

The second approach, which we call the *latent variable* approach and the approach we wish to present here, is to define the construct in terms of other theoretical constructs and empirical relations. Measures are designed according to the hypothesized causal linkages between the latent variable and observable phenomena. Of necessity, measures are indirect and imperfect relations between the latent construct and observations. The latent variable approach views the skin caliper, hydrostatic, and ultrasound procedures as three ways of measuring body fat content, each being an indirect and error-laden measure and neither laying claim to defining body fat content. Likewise, the latent variable approach views $\dot{V}O_2$ max not as defining aerobic fitness but as one measure of it, with improved measures of aerobic fitness to be anticipated in the future.

Note: Partial support for writing this chapter came from NIMH Training Grant T32MH15789–08.

The latent variable approach logically separates the problem of defining a construct from the problem of measuring it. The definition of a construct embeds it in a network of theoretical concepts and accepted empirical laws. The measurement of a latent construct is achieved through its causal manifestations. Thus, one can maintain a construct while developing and improving its measurement procedures, or one can modify or refine a construct without first worrying how it is going to be measured. Body fat content, continuing our example, is a latent physiological concept, one which is well defined independently of any methods for measuring it. Its measurement is based on the observable causal manifestations of variations in body fat content: skinfold thickness in the case of the caliper method, the specific gravity of body fat in the hydrostatic method, and ultrasound conductivity in the case of the ultrasound method.

For a construct to be of scientific utility, it must account for the observed correlations among its observed measures. If a set of measures are tapping a single construct, then they must have reasonably high correlations among each other. On the other hand, if two measures are uncorrelated, then they cannot be said to be measuring the same thing. The degrees of correlations among measures supposedly measuring the same latent variable provide the basis for hypothesizing and testing latent variables and their measures.

Our review of the physical fitness literature coupled with discussions with academic colleagues in physical fitness has given us the impression that these two approaches are often confounded in the development of fitness and related measurements. On one hand, the measures are implicitly conceptualized as latent constructs. On the other, observed measurement procedures are restrictively treated as operational definitions of constructs. In our exposition we hope to demonstrate the advantages of the latent variable approach.

Fundamental measurement

From a formal point of view, measurement is the assignment of numbers to attributes of observed phenomena according to a definite scheme. At one extreme, naive concepts or simple empirical relations may govern the scheme for assigning the numbers. Thus, one might assign the number 0 to those people of normal body weight and 1 to those who are over- or underweight. At the other extreme, measurement rules would be embedded in an elaborate network of theoretical concepts and established empirical laws. Thus, in stress testing one would use the detailed measures derived from physiological theory.

Fundamental measurement theory is concerned with the formal structure of these schemes for assigning numbers to attributes (1–6). We discuss three of the four problems addressed by fundamental measurement: representation, uniqueness, and meaningfulness. We do not discuss the fourth problem: scaling (1). The representation problem is concerned with the existence of measurement. Given a set of attributes to be measured, does there exist a scheme for assigning numbers that appropriately represents the attributes being measured? The goal is to show that there is a homomorphism between the attributes and the real numbers that preserves the relations among the attributes. The homomorphism that achieves this is called a measurement *scale*. Then the attributes are said to be represented by the real numbers according to the scale. Any changes in the attributes are reflected by corresponding changes in the numbers that represent the attributes. In practice, the representation problem for a proposed measurement scale is rarely, if at all, addressed. Appropriate representation is assumed to be valid.

The second problem is concerned with the uniqueness of measurement. Given a measurement scale, to what degree is it determined by the measurement process and to what degree is it arbitrary? The goal here is to find the class of transformations of the scale that do not alter the scale's representativeness. That is, given a measurement scale, what other scales would do equally as well and what are the mathematical connections between them and the original scale? This problem is often discussed in textbooks on measurement and statistics, and will only be summarized here.

A *nominal* scale is one in which numbers are used as substitutes for names only to distinguish different attributes, such as in the numbering of colors. Hence, any transformation of this scale that preserved the uniqueness between the attribute and the number, i.e., a bijection, would serve equally as well. A *classification* scale is one in which numbers are used to identify equivalence classes among the attributes, such as using one number for males and a different one for females. Any bijection that preserves the equivalence classes would serve as well. The nominal and classification scales are often confounded in the literature, and consequently the incorrect claim is often made that classification scales allow arbitrary transformations.

An *ordinal* scale is one in which attributes are ordered according to an attribute, as in the case of preferences. This scale is unique up to a monotonic bijection. An *interval* scale is an ordinal scale such that equal differences between attributes are represented in the scale, as in the case of temperature scales. This scale is unique up to a positive linear transformation. A *log-interval* scale is similar to an interval scale except that equal ratios, rather than differences, between attributes are represented in the scale, as in some psychophysical functions. This scale is unique up to a power transformation. A *difference* scale is an interval scale in which the intervals are of fixed length determined by the measurement process, as in the case of some psychological functions. It is unique up to an additive constant. The *ratio* scale is an interval scale with a true origin, as in weight and height scales. This scale is unique up to a positive multiplicative constant. And lastly, an *absolute* scale is one with a true origin and a fixed interval length, as in the case of counting. Absolute scales permit no transformation.

Of these scale types, the interval scale is considered the most useful. Most proposed measures are designed to be at least an interval scale, although whether this is ever achieved is often disputed. The ratio and absolute scales are primarily found in the physical sciences. The log-interval and difference scales are rarely used, except in certain specialized areas of psychology. The nominal and classification scales are considered by many not to be measurement at all. Other classifications of measurement scales exist as well. Coombs (7) includes partial order as well as order in classifying scales. Torgerson (8) includes the concept of distance and excludes nominal and classification scale types in his typography. Luce and Narens (9) briefly present a new, more formal classification.

The meaningfulness problem, which is closely related to the uniqueness problem, is concerned with the interpretation of measurement. Valid interpretation of a measurement is constrained by its scale type, for any interpretation must remain invariant with respect to the admissible transformations within that scale type. Thus, to claim that object A is twice as hot as object B on the fahrenheit scale, which is an interval scale, is formally meaningless because object A cannot also be twice as hot as object B on the centigrade scale, which is an admissible transformation. To claim that object A weighs twice that of object B in pounds, which is a ratio scale, is formally meaningful because no matter what measure of weight is used the claim remains true.

The problems of uniqueness and meaningfulness spill over into a dispute over the appropriate statistical methods for the various scale types, over which there exists a rather confusing literature (10–14). Resolution of this dispute is severely hampered by a lack of a theory of measurement error within the formal measurement theory framework.

There is a large, unfilled gap between fundamental measurement theory and the remaining theory of measurement we wish to present. Indeed, some authors argue that fundamental measurement theory does not capture the essence of the idea of measurement (15, 16). Most approaches to measurement, including the one taken here, straddle this gap by making a useful mathematical assumption: We assume that *the measurement process can be modelled as a random variable*. That is, together with some very general mathematical assumptions, we can treat observed measures as a probabilistic process arising from true scores and errors generated by probability distributions (17). We further restrict ourselves in the paper to *classical measurement theory*, in which the measurement process can be modelled as a *continuous* random variable (18).

Classical measurement theory

When a researcher in the process of measuring assigns a number, it is always to an *observable* attribute. An important question is the relationship between the random variable representing the observed attribute and the *latent* random variable that generated the attribute. The difference between the observed random variable and the latent random variable is defined to be the *measurement error*. The classical measurement theory stipulates that the observed random variable x can be linearly decomposed into a latent true score random variable τ and an error random variable ϵ:

$$x = \tau + \epsilon \tag{1}$$

At this point the distributions of τ and ϵ need not be specified other than the trivial specifications that their expectations are finite and their variances are positive and finite. The classical theory continues with two more assumptions. The first is that the expected value of the errors of measurement is zero:

$$E(\epsilon) = 0 \tag{2}$$

The second is that the true scores are uncorrelated with the measurement error:

$$\operatorname{Cov}(\tau, \epsilon) = 0 \tag{3}$$

From these assumptions it follows that the expected value of the observed measurements is the expected value of the true scores, and that the variance of the observed measurements is the sum of the true score variance and error variance:

$$E[x] = E[\tau] = \mu_\tau \tag{4}$$

and $$\operatorname{Var}[x] = \sigma_x^2 = \operatorname{Var}[\tau] + \operatorname{Var}[\epsilon] = \sigma_\tau^2 + \sigma_\epsilon^2 \tag{5}$$

Often we will need to compare two measures, say x_1 and x_2, in terms of their respective true scores, τ_1 and τ_2 and their respective errors ϵ_1 and ϵ_2. To do so, we make the additional assumption that each error random variable is uncorrelated with the other true score random variable and other error random variable:

$$\operatorname{Cov}[\epsilon_1, \epsilon_2] = \operatorname{Cov}[\tau_1, \epsilon_2] = \operatorname{Cov}[\tau_2, \epsilon_1] = 0 \tag{6}$$

Given two measures with the above assumptions we can identify three further models relating measures to one another. The most restrictive model is that of *parallel* measures. The parallel model assumes, in addition to the above-listed assumptions, that the two observed measures x_1 and x_2 have but one true score and homoscedastic errors:

$$x_i = \tau + \epsilon_i \qquad i = 1, 2 \tag{7}$$

and $$\operatorname{Var}[\epsilon_1] = \operatorname{Var}[\epsilon_2] = \sigma_\epsilon^2 \tag{8}$$

The variance of each x_i is constant:

$$\operatorname{Var}[x_i] = \sigma_\tau^2 + \sigma_\epsilon^2 \qquad i = 1, 2 \tag{9}$$

The covariance between x_1 and x_2 is

$$\mathrm{Cov}[x_1, x_2] = \sigma_\tau^2 \tag{10}$$

A less restrictive model is that of *tau-equivalent* measures, in which x_1 and x_2 have the same true score equation as in parallel measures, but no restriction is placed on the error variances. The variances of x_i are not constant, as

$$\mathrm{Var}[x_i] = \sigma_\tau^2 + \sigma_{\epsilon,\, i}^2 \qquad i = 1, 2 \tag{11}$$

But the covariance between x_1 and x_2 is the same as in the parallel model.

The least restrictive and most useful model is the *congeneric* model in which the two measures have different true scores and errors, but the individual true scores themselves are linearly related to a common true score:

$$\begin{aligned} x_i &= \tau_i + \epsilon_i \\ \tau_i &= \mu_i + \beta_i \tau \end{aligned} \qquad i = 1, 2 \tag{12}$$

where τ is the common true score random variable with $E[\tau] = 0$ and $\mathrm{Var}[\tau] = 1$. The variance of x_i is

$$\mathrm{Var}[x_i] = \beta_i^2 + \sigma_{\epsilon,\, i}^2 \qquad i = 1, 2 \tag{13}$$

The covariance of x_1 and x_2 is

$$\mathrm{Cov}[x_1, x_2] = \beta_1\beta_2 \tag{14}$$

The properties of parallel, tau-equivalent, and congeneric measures will be further elaborated throughout this paper as we need them.

Reliability and validity defined

Reliability refers to the extent to which the variance of an observed x is due to random sources or to "noise." Therefore, reliability is defined as the ratio of variance in x from nonrandom sources to the total variance.

The reliability of x, labeled ρ_x, is simply the ratio of true score variance to the observed variance,

$$\rho_x = \sigma_\tau^2/\sigma_x^2 \tag{15}$$

Although perhaps not obvious from casual inspection, the square root of equation 15 equals the correlation between the true and observed score. To see this, note that

$$\rho_{\tau x} = C(\tau, x)/\sigma_\tau\sigma_x \tag{16a}$$

$$= \frac{C(\tau, \tau + \epsilon)}{(\sigma_x\sqrt{\rho_x})\sigma_x} \quad \text{from equations 1 and 15} \tag{16b}$$

$$= \frac{\sigma_\tau^2}{\sigma_x^2\sqrt{\rho_x}} \quad \text{from equation 3 and the fact that } C(\tau, \tau) = \sigma_\tau^2 \tag{16c}$$

$$= \sqrt{\rho_x} \quad \text{from equation 15} \tag{16d}$$

From equation 15 and equation 5 it immediately follows that

$$\rho_x = 1 - \sigma_\epsilon^2/\sigma_x^2 \tag{17}$$

as well. Hence equation 17 is an alternative expression for the reliability of a measure. Because a variance is a nonnegative quantity, it is clear from equation 17 that the reliability of a measure is between 0 and 1. The greater the error variance, relative to the observed variance, the closer the reliability is to zero. And when the error approaches zero, the closer the reliability is to unity.

Note that if one knows ρ_x it follows immediately that

$$\sigma_\tau^2 = \sigma_x^2\rho_x \tag{18}$$

That is, the true score variance of x equals the observed variance multiplied by the reliability of the measure.

The correlation between τ and $\rho_{\tau x}$ is called the theoretical validity of a measure x because it is a measure of how well an observed item correlates with some latent, theoretical construct of interest. Following Lord and Novick (18), we differentiate theoretical validity from empirical validity, where the latter is the correlation between x and a second observed variable y. That is, empirical validity can only be assessed in relation to an observed measure. Hence, it makes no sense to speak of the empirical validity of x in the singular. Given these definitions and the assumptions of true score theory, Lord and Novick (18) link the concepts of reliability and validity by proving that

$$\rho_{xy} \leq \rho_{\tau x} = \sqrt{\rho_x} \tag{19}$$

that is, the empirical validity of a measure x in relation to a second measure y cannot exceed the theoretical validity or the square root of its reliability. This means that the square root of the reliability of a measure is an upper bound to its validity with respect to any criterion measure. It also makes clear what is intuitively obvious: No measure can be valid without also being reliable, but a reliable measure is not necessarily a valid one.

Although it may not seem intuitively obvious, a biased measure can be a perfectly reliable one. Think of a scale that weighs everything 5 pounds heavier than it really is. Despite this problem, the scale's reliability is unaffected. To see this we simply note that the variance of $x + 5$ is equal to the variance of x. And since the reliability coefficient is the ratio of σ_τ^2 to σ_x^2 it is clear that constant biases do not affect the reliability of measurement.

Effect of unreliability on statistical estimates

Until this point we have not made explicit how one's estimates of statistics such as means, standard deviations, and correlation coefficients are affected by unreliability of measurement. Assuming that measurement errors in the long run equal zero (i.e., the assumption made in equation 2), it follows immediately from equation 1 that

$$\mu_x = E(x) = E(\tau) = \mu_\tau \tag{20}$$

In the long run the mean of the observed scores equals the mean of the true scores as long as the observed measures are unbiased. The variance of the true scores can be estimated by multiplying the variance of an observed score by its reliability estimate. Assuming measurement error is random, the true score variance is less than or equal to the observed score variance. (This point is obvious from equation 5 as well.) Note that equation 18 makes it clear that the reliability of one's measures approaches zero as the observed true score variance approaches zero.

Although counterintuitive, the covariance between two variables x and y is unaffected by errors of measurement in the two variables. This is proven by expressing $C(\tau_x, \tau_y)$ in terms of observables and measurement errors:

$$\begin{aligned}
C(\tau_x, \tau_y) &= C(x - \epsilon_x, y - \epsilon_y) \\
&= C(x, y) - C(x, \epsilon_y) - C(\epsilon_x, y) + C(\epsilon_x, \epsilon_y) \\
&= C(x, y) - C(\tau_x + \epsilon_x, \epsilon_y) - C(\epsilon_x, \tau_y + \epsilon_y) \\
&\quad + C(\epsilon_x, \epsilon_y) \\
&= C(x, y) - C(\tau_x, \epsilon_y) - C(\epsilon_x, \epsilon_y) - C(\epsilon_x, \tau_y) \\
&\quad - C(\epsilon_x, \epsilon_y) + C(\epsilon_x, \epsilon_y) \\
&= C(x, y)
\end{aligned} \tag{21}$$

since it follows from equation 6 that $C(\epsilon_x, \tau_y) = C(\epsilon_y, \tau_x) = C(\epsilon_x, \epsilon_y) = 0$.

Although the covariance is not affected by measurement error, the correlation coefficient is. Because $\rho_{\tau_x\tau_y} = C(\tau_x, \tau_y)/\sigma_{\tau_x}\sigma_{\tau_y}$ it follows from equations 18 and 19 that

$$\rho_{\tau_x\tau_y} = \rho_{xy}/\sqrt{\rho_x \rho_y} \tag{22}$$

Notice that the correlation between x and y approaches $\rho_{\tau_x\tau_y}$ as ρ_x and ρ_y approach unity. Equation 22 is the best known of the attenuation formulas. It expresses the fact that the true correlation between x and y is attenuated by measurement error (18).

While errors in both x and y affect the correlation between the two variables, the regression coefficient with y dependent on x is affected only by errors in x, the independent variable. In general, $\beta_{\tau_y\tau_x} = C(\tau_y, \tau_x)/V(\tau_x)$. It follows from equations 18 and 19 after some algebra that

$$\beta_{\tau_y\tau_x} = \beta_{yx}/\rho_x \tag{23}$$

Although not proven here, the intercept is also only affected by measurement error in the independent variable x.

In the three-variable case, the effects of measurement error are more complicated. The true partial correlation coefficient is given by

$$\rho_{\tau_x\tau_y\cdot\tau_z} = \frac{\rho_z\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{\rho_x\rho_z - \rho_{xz}^2}\sqrt{\rho_y\rho_z - \rho_{yz}^2}} \tag{24}$$

A proof can be found in equation 19. To show the effects of measurement errors on the partial correlation coefficient, assume that the three disattenuated (observed) coefficients ρ_{xy}, ρ_{xz} and ρ_{yz} all equal .5. Assume further that the reliability of x, y, and z is .8 in each case. Then the observed partial correlation $\rho_{xy\cdot z}$ equals .333. Since $\rho_{xy\cdot z}$ does not take measurement error into account, one might expect the coefficient corrected for measurement error to be larger. And indeed this is the case because $\rho_{\tau_x\tau_y\cdot\tau_z}$ equals .385. But now consider the case where ρ_{xy}, ρ_{xz}, and ρ_{yz} again all equal .5, but where ρ_x and ρ_y equal .8 and ρ_z equals .6. In this case it would again seem likely that $\rho_{\tau_x\tau_y\cdot\tau_z}$ should be larger than .333, the observed $\rho_{xy\cdot z}$. But it is not. Simple calculation using equation 24 shows that it equals .217. Thus although errors in measurement usually attenuate higher order partial correlations they do not always do so, as the last example illustrates. It would be nice to be able to provide the reader with some guidelines indicating when partial correlations will and will not be attenuated, but frankly we know of none.

The true regression coefficient in the three-variable case for the regression of τ_y and τ_x controlling for τ_z is given by

$$\beta_{\tau_y\tau_x\cdot\tau_z} = \frac{\sigma_y}{\sigma_x}\,\frac{\rho_z\rho_{xy} - \rho_{yz}\rho_{xz}}{\rho_x\rho_z - \rho_{xz}^2} \tag{25}$$

This result has similarities to those discussed in connection with both equations 23 and 24. It is similar to equation 23 in that careful inspection of equation 25 makes clear that errors of measurement in the dependent variable y do not affect the true regression coefficient. This result generalizes to the k-variable case as well: Errors in the dependent variable do not affect the true regression coefficient. Equation 25 is also similar to equation 24 in that errors in measurement generally attenuate the estimate of the true regression coefficient; but it is possible to find examples in which the observed coefficient is actually larger than the true one.

The reader interested in more technical treatments of regression estimation in the multivariate case in which measurement error is taken into account should read Warren, White, and Fuller (20) and Fuller and Hidiroglou (21).

To summarize, assume the assumptions of classical measurement theory hold and measurement error exists:

1. The observed mean equals the true mean.
2. The variance of the observed scores is larger than that of the true scores.

3. In two-variable cases, the observed correlation and regression coefficient are smaller than the true coefficients.
4. In the k-variable case the observed partial correlation coefficients and regression coefficients are usually, but not always, smaller than the true coefficients.
5. Errors of measurement in the dependent variable do not affect the estimate of the true regression coefficient.

Estimating the reliability of a measure with parallel tests

Thus far, we have not discussed how to actually estimate the reliability of a measure. Although far better ways exist to make estimates, and several of them will be discussed below, one way is simply to correlate two parallel measures. (Parallel measures were defined by equations 7–9.) It is very easy to show that the correlation between two parallel measures estimates the reliability of the measure. Let two measures x and x' be parallel, then

$$\begin{aligned} \rho_{xx'} &= C(x, x')/\sigma_x\sigma_{x'} & \\ &= C(\tau + \epsilon, \tau + \epsilon')/\sigma_x^2 & \text{from equation 1} \\ &= C(\tau, \tau)/\sigma_x^2 & \text{from equation 6} \\ &= \sigma_\tau^2/\sigma_x^2 & \\ &= \rho_x & \text{from equation 15} \quad (26) \end{aligned}$$

The parallel measure approach to the estimate of reliability is largely of historical interest in view of the restrictive definition of parallel measures. More useful approaches are discussed in the following sections.

Reliability as a function of the number of independent measures

It makes no intuitive sense that one's confidence in the actual or true value of a particular phenomenon should increase as the number of independent measurements of a phenomenon increase for a given level of reliability of measurement. One would be more confident with a pair of independent measures of aerobic fitness each with a reliability of .6 than with a single measure with a reliability of .6. And we would be even more confident if we had four such independent measure, and so on. This intuitive reasoning forms the rationale for using multiple items rather than a single one to measure a construct (22, 23). The assumption is that errors in measurement will be random, that is, will cancel each other out, and the average of all the measurements is a better estimate of the true value than any single one.

It is also clear that one needs fewer independent measures to achieve a given level of confidence when the reliability of each individual measurement is high. If one knows a measure has a reliability of .9, one might accept a single measurement, whereas one might need eight or nine independent assessments for a measure with a reliability of .5 to have the same degree of confidence in what the true value is.

The relationship between the reliability of a composite x_n (where $x_n = \Sigma_{i=1}^{n} x_i$) and n parallel measures of known reliability was derived independently by Spearman (24) and Brown (25). Both showed that the reliability of a measure that is n times longer than the original is

$$\rho_{x_n} = \frac{n\rho_x}{1 + (n - 1)\rho_x} \tag{27}$$

where ρ_{x_n} is the reliability of the composite.

Equation 27 shows reliability increases as a function of the number of independent measures for various initial reliabilities. The implication for research on physical fitness is obvious. When one has relatively reliable items, only a few are needed to achieve a relatively high degree of reliability for a composite score based on them; more are needed the lower the reliability of each individual measurement. Thus if one demands a reliability of .8 for ρ_{x_n}, it can be achieved with 2 items with individual reliabilities of .7, 4 of .5 or 10 of .3.

Types of reliability

To this point, our discussion of reliability has been largely theoretical and historical. We now discuss specific ways to assess reliability. Generally, reliability measures can be divided into two major classes: measures of stability and measures of equivalence.

Measures of stability

A person's response to a particular test or other measure may vary from occasion to occasion. The respondents may not be certain how they feel about being tested, or they may be distracted by other matters, or they may be tired or ill, and so on. All of these will contribute to errors of measurement and therefore depress the reliability of the items. Given that error exists, the problem is how to assess it in a way that satisfies the definition of reliability given in equation 15. Historically, the most popular way to evaluate it has been to correlate respondents' responses to a measure at one point in time with their responses to the same measure at some later point in time. Reliability evaluated by correlating a measure across time is called a *measure of stability* or, more commonly, *test-retest reliability*. The assumption being made is that shown in figure 1. A measure is assumed to correlate with itself across time because of an underlying unobserved true variable τ. The equations linking the observed responses at Time 1 and Time 2 to τ are

$$x_1 = \tau + \epsilon_1 \quad \text{and} \quad x_2 = \tau + \epsilon_2$$

But it is clear to see that if it is assumed that $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2$ and that $C(\epsilon_1, \epsilon_2) = 0$, it follows that

$$\rho_{x_1x_2} = \rho_x \tag{28}$$

That is, if these assumptions hold, the test-retest correlation is the reliability of the measure.

There are some problems with test-retest reliability estimates, however. One is likely to obtain different estimates depending on the amount of time between measurement and remeasurement. Generally it is found that the shorter the time interval the higher the estimate of reliability. Why should this be true?

First, for verbal response items there is the problem of memory. If the interval between measurement and remeasurement is short, the respondents may remember their earlier responses, making them appear more consistent with respect to the true content than they in fact are. To handle this problem some researchers employ a ***parallel forms*** test-retest procedure. A different, but presumably parallel form, of the measure is used at the second administration.

A second problem with the test-retest approach has to do with the assumption that the errors of measurement ϵ_1 and ϵ_2 are uncorrelated. If the errors of measurement are in some sense systematic and not random, one would expect that the same sources of bias might operate each time measurement occurs, thereby making the assumption of uncorrelatedness of errors in measurement highly suspect. But in the simple test-retest model, there is no way to estimate correlated measurement error. However, several writers (24–26) have shown how the use of multiple indicators to measure the underlying variable can yield enough information in some cases to allow for the estimation of correlated measurement error. We shall examine this approach in a later section.

A third problem is that true change cannot be distinguished from unreliability in a simple test-retest reliability design, and obviously the longer the time interval between measurement and remeasurement, the greater the probability that respondents will have in fact changed on the underlying unobserved variable. It should be clear that if individuals have in fact changed, a low test-retest correlation does not necessarily mean that the reliability of one's measure is poor. Several attempts to deal with the true change versus unreliability problem by using multiple remeasurements have appeared in the last 10 years. Building on Heise's (27) pioneering work several writers have extended this work (28–31). All of them assume that (a) errors of measurement are uncorrelated across time, (b) one has measures on an item at three points in time, and (c) the unobserved latent variables are uncorrelated with the errors in measurement. For a summary of this work see Bohrnstedt (32).

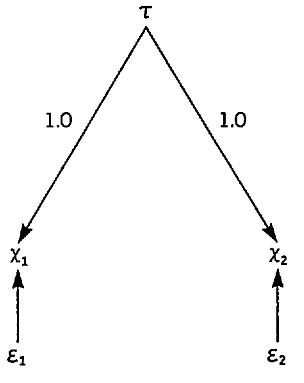


Figure 1. Schematic representation of the test-retest correlation as a reliability estimate.

Measures of equivalence

Whereas historically reliability has more often been assessed using measures of stability, in recent decades an alternative approach has gained in popularity—the use of measures of equivalence. Parallel tests administered at the same point in time and correlated to estimate reliability is an example of a measure of equivalence, although an outdated one. The assumption is made that the two tests have the same underlying true score and as such are equivalent and hence equally good measures of the true score. In a sense the two tests can be thought of as an instant test-retest. One need not be limited to a single pair of parallel items. Indeed, because reliability is partially a function of the number of items used, it makes good sense to build composites or scores that are sums of items so long as one can assume the items in a given score have the same underlying true score.

One of the earliest varieties of equivalence measures to appear were the ***split-half methods***. In the split-half approach to reliability the total number of items in a composite is divided into two halves and the half-scores are then correlated. Since the actual measure is twice as long as the half-score being correlated, the correlation is usually inserted in the Spearman-Brown prophecy formula with $n = 2$ (see equation 27) to get an estimate of the reliability of the total composite.

It should be mentioned in passing that split-half methods have also often been used in a test-retest format where one version is used at Time 1 and a second version is used at Time 2. In this case the splits are assumed to be parallel composites. This procedure is followed as a possible way to deal with memory effects that might confound test-retest reliability estimates using the same composite across time. In this case the simple zero-order correlation between the two halves is taken as the reliability estimate. That is, it is not inserted into the Spearman-Brown formula because the total composite is never employed, only the half-score.

The split-half techniques have gradually been replaced by internal consistency methods for the estimate of reliability with cross-sectional data. Internal consistency reliability estimates utilize the covariances among all the tests simultaneously rather than concentrating on a single correlation between two arbitrary splits.

Kuder and Richardson (33) were the first to devise a measure of equivalence that utilized all the covariances among the tests. However, their formulas KR20 and KR21 could only be used with dichotomous items. Hoyt (34) and Guttman (35) presented generalizations of the KR formulas for polychotomous items. But by far the most popular generalization has been coefficient α developed by Cronbach (36), where

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\Sigma_{i=1}^{n} V(y_i)}{\Sigma_{i=1}^{n} V(y_i) + 2\,\Sigma_{i<j}^{n}\Sigma^{n} C(y_i, y_j)}\right] \quad (29)$$

$$= \frac{n}{n-1}\left[1 - \frac{\Sigma_{i=1}^{n} V(y_i)}{\sigma_x^2}\right]$$

where $x = \Sigma_{i=1}^{n} y_i$. Importantly, α in general is a lower bound to the reliability of an unweighted composite of n items. That is, $\rho_x \geq \alpha$. This important relationship was shown by Novick and Lewis (37). However, as Lord and Novick (18) have pointed out, α *equals* the reliability of the composite if the n measures are either parallel or tau-equivalent. (See equation 11 for a definition of tau-equivalent measures.) Indeed, if one's n measures are tau-equivalent or parallel, α is exactly equal to the Spearman-Brown prophecy formula, that is, to equation 27. In practice it is rare for one's items to be tau-equivalent and therefore α will give a lower bound estimate to reliability. However, unless one's items deviate substantially from being at least tau-equivalent (something one can get a quick-and-dirty estimate of by examining the interitem covariances for rough equality) α will give quite good estimates of reliability.

Two points deserve brief mention. First, if one's items are dichotomous it is easy to prove that α is exactly equal to the formula for KR20. Second, Cronbach (36) proved that α is equal to the average of all possible split-half correlations among the n tests. Although any particular split-half might be closer than α is to the true reliability, there obviously is no a priori way to know this. This demonstrates why it is better to estimate the reliability of an n-item composite using α instead of an arbitrary split-half estimate.

In the National Survey of Personal Health Practices and Consequences (38), persons 20–64 years of age were asked a set of seven questions dealing with their frequency of participation in selected leisure-time activities. We can use the responses to these seven items to determine what the internal consistency reliabilities are, using α, for both the males and the females. The question asked was, Please tell me how often do you . . . ? The seven areas listed then were:

1. Go swimming in the summer (SWIM)
2. Take long walks (WALK)
3. Work on a physically active hobby such as dancing or gardening? (HOBBY)
4. Go jogging or running? (JOG)
5. Ride a bicycle? (BIKE)
6. Do calisthenics or physical exercise? (EXERCISE)
7. Participate in other active sports not mentioned above? (OTHSPORT)

The responses for each item were coded into one of four ordered categories: never, rarely, sometimes, or often.

Assuming that the responses to these items are coded 1, 2, 3, and 4, respectively, and then are added into a single score for each person, we can compute the sample covariance matrix (table 1) and compute the α estimates for the males and the females separately. Using the data in table 1, we estimated α = .65 for the males and α = .68 for the females. No firm rule exists on how low the reliability of a composite can be before it is considered to be of use. However, some measurement specialists suggest that .70 is about the minimum value. In fact, by the standards required in the physical sciences, a reliability of .70 is very low indeed. In other words, these items, which all putatively measure physical fitness, have low internal consistency reliabilities for both the males and the females. Recall from equation 26 that reliability is a function of both the average correlation among the items and the number of items. An

Table 1. Covariance and correlation matrices for physical activity items

Item	1	2	3	4	5	6	7
			Males (*N* = 988)				
1. SWIM	1.133	.122	.162	.344	.257	.313	.453
2. WALK	.115	1.000	.203	.202	.177	.123	.139
3. HOBBY	.141	.188	1.161	.122	.113	.178	.268
4. JOG	.292	.182	.102	1.227	.192	.562	.415
5. BIKE	.236	.173	.102	.169	1.047	.210	.219
6. EXERCISE	.248	.104	.140	.429	.173	1.402	.435
7. OTHSPORT	.355	.116	.207	.312	.179	.306	1.440
			Females (*N* = 1,490)				
1. SWIM	1.356	.169	.216	.265	.355	.349	.407
2. WALK	.149	.954	.239	.235	.208	.240	.246
3. HOBBY	.174	.229	1.138	.191	.171	.243	.273
4. JOG	.230	.244	.182	.975	.207	.390	.361
5. BIKE	.280	.196	.147	.193	1.184	.220	.373
6. EXERCISE	.256	.209	.194	.337	.173	1.378	.395
7. OTHSPORT	.301	.217	.221	.315	.296	.290	1.343

Note: Covariances are above the main diagonal; correlations below. The main diagonal contains item variances.

inspection of table 1 reveals that a number of the correlations between the items are not particularly high, which accounts for the somewhat low internal consistency reliability estimates.

We will return to the analysis of these data several times to determine which of the items are the best indicators of activity and which are worst. But as we shall see, dropping items that correlate less well will not improve the reliability of a construct based upon them because, again, the reliability of a composite is a function of *both* the number of items in it as well as the average correlations among the items.

If one can assume that all the item variances are equal, the formula for α simplifies to

$$\alpha = \frac{n\bar{\rho}_{ij}}{1 + (n - 1)\bar{\rho}_{ij}} \tag{30}$$

where $\bar{\rho}_{ij}$ is the average correlation among the n items. Even when the item variances are only roughly equal, equation 30 usually provides a quick-but-not-too-dirty estimate of α. For the data in table 1, for example, the α estimates for the males and the females, using equation 30, are .64 and .68, respectively. The estimate for the males is .01 less than the estimate for α using equation 29 and exactly the same for the females. These results are consistent with our previous experience using both equations 29 and 30. Because most researchers routinely compute item intercorrelations, equation 30 can be used very quickly to compute a quick-but-not-too-dirty estimate of α. Of course, when one has decided upon which items to include in the composite index, equation 29 should be used to provide the best possible estimate of α. Although, as we shall see in the next section, α is not necessarily the best estimate of the internal consistency reliability of a composite of n items.

Exploratory factor analysis and reliability

Because one of the explicit purposes of factor analysis (39) is to reduce a set of measures to some smaller number of latent unobserved variables, it is not surprising that several writers have been interested in the relationship between factor analysis and the reliability of measurement (40–46). There are various models for factor analysis (39), all of which can be broadly classified as either exploratory or confirmatory.

In *exploratory factor analysis* one is primarily interested in data reduction, that is, the representation of n manifest variables by m unobserved latent variables ($m<n$). Unfortunately it turns out that for any set of manifest indicators there is an infinite number of solutions that can account for the observed covariances among the observed variables equally well. To be able to obtain a unique solution, some constraints need to be placed on the solutions. In exploratory factor analysis the usual one is to "rotate" the initial solution to the most "simple structure." Although there are various analytic techniques for achieving simple structure (e.g., varimax and promax), they all have roughly the same goal in mind. In the ideal solution each item would be strongly related to one and only one latent factor. In this ideal situation one would have located the most "pure" indicators of the underlying factors. Because this ideal is rarely (if ever) achieved with real data, one instead settles for solutions that minimize the *complexity* of the measures, where complexity refers to the number of significant nonzero coefficients linking an observed measure to the factors. The general model is given by:

$$z_i = \lambda_{i1}F_1 + \lambda_{i2}F_2 + \cdots + \lambda_{im}F_m + b_iS_i + c_iE_i \tag{31}$$

where z_i is an observed item i (in standardized form, i.e., transformed to have a mean of zero and standard deviation of unity); F_1, F_2,..., F_m are the latent "common" factors; S_i is a "specific" factor; E_i is a random measurement component; and λ_{ij}, b_i, and c_i are coefficients associated with the parameters. A common factor is one with at least two items with significant coefficients linked to it. A specific factor is a source of variation from a latent variable associated specifically with test or item i. It is assumed that the specific factor and measurement error are mutually uncorrelated. Furthermore, it is assumed that the common factors are uncorrelated with the specific factors and measurement errors. To summarize, each item is seen to be a function of three independent sources of variance—m common factors, one specific factor, and measurement error. In fact, however, in cross-sectional studies, there is no way to estimate uniquely the variance associated with the specific factor and measurement error. Hence, the factor models, which are in fact estimated, are of the form:

$$z_i = \sum_{j=1}^{m} \lambda_{ij}F_j + d_iU_i \tag{32}$$

where $U_i = S_i + E_i$ and is called the "uniqueness" of an item.

The details of how the λ_{ij} are estimated are far too complex and detailed to present here. The interested reader can pursue the topic in Harman (39).

Consider a model with a single common factor:

$$z_i = \lambda_iF + d_iU_i \tag{33}$$

This model is quite similar to the usual formula for composing an item into "true" and "error" components (see equation 1). It is this similarity that undoubtedly has led to the attempts to link factor analysis to reliability theory. There are two important differences between equation 33 and equation 1. First, the factor model in equation 33 assumes that each item may be differentially related to the underlying unobserved variable, whereas the true score model assumes that each item is linked to the underlying true score with a coefficient of unity. Second, although not a necessity, most of the programmed factor analysis methods standardize the variables, whereas the observed variables in the true score model normally are assumed to be in their natural metric.

The fact that the factor analysis model allows for differentially "good" items can be seen as an advantage over the classical true score approach where all items must be equally good to estimate the reliability of a composite. (Recall that Cronbach's α is a lower bound to reliability unless one has items which are tau-equivalent or parallel.) Applying Heise and Bohrnstedt's (42) model, in which they adopt Lord and Novick's (18) definition of reliability as the ratio of true score to observed variance to equation 33 yields their estimate of reliability, called Ω. Specifically,

$$\Omega = 1 - \frac{\Sigma_{i=1}^{n} V(y_i) - \Sigma_{i=1}^{n} V(y_i) h_i^2}{\Sigma_{i=1}^{n} \Sigma_{j=1}^{n} C(y_i, y_j)} \qquad (34)$$

where h_i^2 is the communality of item i. (The communality of an item is the amount of variance in the item that can be explained by the common factors (39). In the case of a single factor the communality of item i is simply λ_{ij}^2.) Although it is easy to demonstrate that Ω is a better estimate of reliability than Cronbach's α, there are several problems with the Heise-Bohrnstedt approach. First, there are several exploratory factor-analytic models that can be used to estimate equation 32, all of which give slightly different estimates of the λ_{ij} and hence the h_i^2. Second, one may find that a single-factor model does not fit the observed covariances very well; hence additional factors may be needed to account for the factors. Third, when more than a single factor is needed to account for the data, one is faced with the problem that an infinite number of rotations will fit the data equally well. To deal with the case of multiple factors, Heise and Bohrnstedt introduce the concepts of validity and invalidity. In their model, validity refers to what was earlier called *theoretical* validity—the correlation between one's measure and the construct it is presumed to measure. *Invalidity* refers to variance in a measure that is reliable because of factors other than those one intends to measure with a set of items. Heise and Bohrnstedt (42) demonstrate that the reliable variance can be decomposed into the sum of valid and invalid variances when the factors are constrained to be uncorrelated.

Bentler (41) independently developed a measure similar to that of Heise and Bohrnstedt, which he called α_0. Armor (48), Allen (44), and Smith (46) have all developed measures similar to Ω or have refined the original Heise-Bohrnstedt measure. Most of these measures are compared and contrasted in an excellent review article by Greene and Carmines (47).

Despite the limitations just outlined, the estimation of reliability using exploratory factor-analytic methods continues to be useful and popular. A computer program is now available (49) that computes six different reliability coefficients, item-to-total correlations, and the results of a canonical factor analysis.

For the data on males in table 1, $\Omega = .68$ and $\alpha_0 = .72$, suggesting that the internal consistency reliability of .65 estimated by α is indeed too low. By contrast, the Ω values for the data on females, $\Omega = .68$ and $\alpha_0 = .69$, are not significantly different from the estimates given by α. In our experience, the values of all the coefficients are relatively close, but there are some cases in which the estimates of Ω and α_0 are somewhat higher than those of α and hence are to be preferred as estimates of the internal consistency of a composite.

A latent variable approach to the measurement of physical fitness

After our excursions through measurement and reliability theories, we now turn our attention more directly to the possible utility of a latent variable approach for the assessment of physical fitness. As we pointed out in the introductory section of this chapter, it seems unlikely that any single measure or test can be thought of as a gold standard for assessing physical fitness. Rather, we argue that the outcomes we observe, whether from physical instruments or verbal responses of behavior, are best thought of as error-laden measures of a *latent variable*, or more likely a set of interrelated unobservable latent variables, that account for the variation and covariation we observe in our measures.

For the sake of illustration only, assume a single construct called physical fitness exists and it can account for $\dot{V}O_2$ max uptake (x_1), the number of minutes it takes to run a mile (x_2), the number of pushups that one can do in a minute (x_3), and body fat as accounted for by the hydrostatic procedure (x_4). If we call the latent variable τ, then the set of equations that would link these four measures to the latent variable are given by:

$$\begin{aligned} x_1 &= \beta_1 \tau + \epsilon_1 \\ x_2 &= \beta_2 \tau + \epsilon_2 \\ x_3 &= \beta_3 \tau + \epsilon_3 \\ \text{and } x_4 &= \beta_4 \tau + \epsilon_4 \end{aligned}$$

where the β_i are weights that take into account that the observations may be relatively better or worse indicators of the latent variable and the ϵ_i are random measurement errors associated with the measures. Figure 2 shows how the four equations would look graphically if we assume that the arrows represent causal effects. That is, we make the assumption that physical fitness is an unobserved variable that causally determines the observations we make on these four variables, and that the four observed variables are imperfect fallible measures. We will not get into a discussion of the determinants of the latent variables although there are almost certainly genetic and behavioral ones; instead, we leave that issue to those who do research in the area.

In the example shown in figure 2, we have assumed that four rather different observed measures are due to a single latent variable. Though not expert in the field of physical fitness, our guess is that physical fitness is a

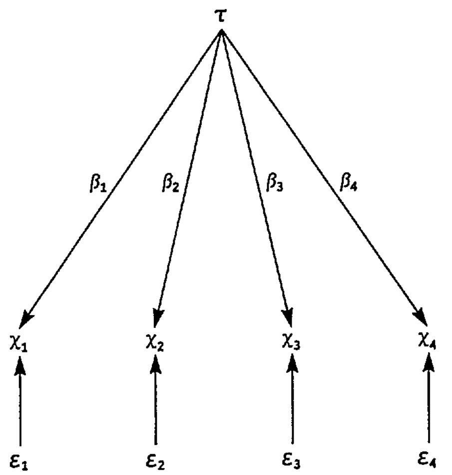


Figure 2. Hypothetical latent variable conceptualization of physical fitness with four fallible observed measures.

multivariate domain. Three subdomains that might make sense are *aerobic fitness, muscular fitness,* and *activity fitness,* for example. We do not offer this list as either theoretically correct, nor do we believe it to be exhaustive. Instead, our purpose is to get those who work in this area to think of the domain as more complex than the unidimensional term "physical fitness" implies. The researchers in the area must determine what the relevant dimensions are. For the sake of illustration, assume that the three areas identified above make sense. Then one can specify a richer model, such as that shown in figure 3. Notice that we have (a) assumed that the observed measures are congeneric (see equation 12) and (b) that the various types of fitness (shown by τ_1, τ_2 and τ_3 in figure 3) are correlated with one another, and (c) we have assumed that the measurement errors are uncorrelated with each other.

Before discussing how models such as those shown in figures 2 and 3 can be estimated, the notion of *domain* needs to be discussed. One can imagine a *domain of meaning* that a particular construct such as physical fitness is intended to measure. One's measures have *content validity* (50) to the degree that one has representatively sampled from the domain of meaning in choosing and designing those measures. One can think of a domain as having various *facets* (51) and just as one can use stratification to obtain a sample of persons, one can use stratification principles to improve the content validity of a construct that one wants to measure. Facets of the construct physical fitness might include aerobic fitness, muscular strength, muscular endurance, and so on.

Whereas content validity has received close attention in the construction of achievement and proficiency measures within the fields of psychology and educational psychology, it has generally been ignored in other disciplines. Although it may sound like a good idea to sample the facets of a domain, the fact is that a domain such as physical fitness cannot be enumerated in the same way that a population of persons or objects can be.

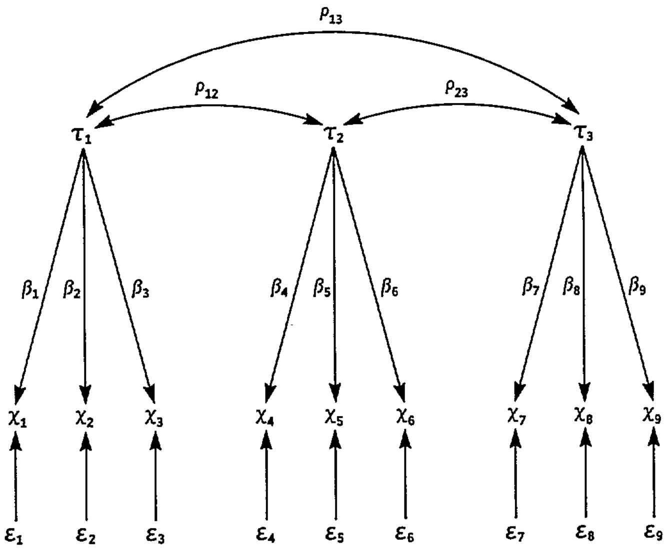


Figure 3. Three latent variable conceptualization of physical fitness with nine congeneric observed measures.

But two broad guidelines can be provided that may be of help to the researcher regardless of his or her discipline. First, one should search the literature to see how various others in the discipline have used or defined the construct and search for facets of meaning while doing so. Second, one should rely on his or her own observations and insights and ask whether they yield additional facets not considered by other researchers in the area. Using these two guidelines, the domain should now be stratified into its major facets making certain that the stratification is exhaustive, that is, all major facets are represented. If one or more strata appear by themselves to have a complex of meanings, they should be divided further into substrata. Then one should inventory extant measures by provisionally assigning them to one of the strata or substrata. Ideally when this task is completed one will have a minimum of three or four measures of each of the strata or substrata.

The approach we have outlined above makes the tacit assumption that the stratifications can be thought of as reflecting latent variables. What one wants to do is make certain that one has included all the latent variables needed to exhaust the domain and that one has multiple measures of each of the posited latent variables. Ideally all of our measures would be congeneric (see equation 12). The next step would be to administer the entire set of measures to the population under study. We now turn our discussion to the analysis of a set of items constructed in this way.

Confirmatory factor analysis

Confirmatory factor analysis (52), unlike exploratory factor analysis, allows one to test whether the data observed fit a *hypothesized* factor structure such as the one shown in figure 3. In particular, one specifies that the parameters in the model are one of three types. They can be either freely estimated, fixed to some value a priori (usually zero or one), or constrained to equal another parameter. Recall from equation 12 that congeneric measures are related to a single underlying latent variable (or factor) τ. This implies that the coefficients for each item are fixed to zero for all but one of the factors. Recall also that parallel and tau-equivalent measures (equations 7–10 and 11) assume that all of the measures bear the same relationship to the underlying latent variable of interest. By contrast, the definition of congeneric items allows items to be differentially weighted as a function of their relationship to the latent variable. The weight applied to the item is the β_i in equation 12. (Also see figure 3.) If one did a confirmatory factor analysis in which one constrained all of the β_i to be equal and the errors of measurement were constrained to be equal as well, one would be testing whether the items are parallel. If one constrained only the β_i to be equal, but freely estimated the error variances, one would be testing whether the items were tau-equivalent. But there is no reason to expect one's measures to meet such restrictive assumptions, and therefore, we ordinarily assume that the measures are congeneric and hence treat both the β_i and the variances of the measurement errors as parameters to be freely estimated.

Once we have estimated the weights with confirmatory factor analysis, given that the variance of an item x_i is given by equation 13, i.e., the sum of the weight squared plus the item's random error variance, it follows from the definition of reliability given in equation 15 that

$$\rho_i = \beta_i^2/(\beta_i^2 + \theta_i^2) \tag{35}$$

Now let x, μ, β, and e be column vectors of order n with elements x_i, μ_i, β_i, and e_i, respectively. Then a set of congeneric measures can be rewritten in vector form as

$$\mathbf{x} = \mu + \beta\tau + \mathbf{e} \tag{36}$$

If Σ is the covariance matrix of x and θ^2 is a diagonal matrix of error variances, then $E(\mathrm{xx}') = \Sigma$ and

$$\Sigma = \beta\beta' + \theta^2 \tag{37}$$

Equation 37 is the basic equation of the factor model with one common factor. If one can assume that the x_i follow a multivariate normal distribution, the parameters of the model can be efficiently estimated by Jöreskog's maximum likelihood method (49). The method yields large sample standard errors for all of the parameter estimates as well as the overall chi-square goodness-of-fit statistic referred to previously, which allows one to *test* the assumption that one's measures are congeneric.

While the definition of congeneric measures comes much closer to the assumption most survey researchers make, compared to those of parallel measurement and tau-equivalence, it is still unlikely that the model adequately describes the measurement characteristics of most of the items used. As Alwin and Jackson (52) show, it is difficult to imagine measures so "pure" that their variation is due to a single construct. As they point out, a set of items that can be fit by a single factor and hence may appear to be congeneric will often prove not to be when factored with items from other content domains.

One of the reasons that it is difficult to find measures that are congeneric is that the goodness-of-fit test, which is used to evaluate the fit of the data to the model, is sensitive (a) to sample size (53) and (b) to violations of the assumptions of multivariate normality (54). Large sample sizes and violations of distributional assumptions both lead to large χ^2 values relative to their degrees of freedom, and hence it is difficult to fit *any* data set to a congeneric model using traditional α levels of, say, .05 or larger. For this reason, any factor analysis that uses a maximum likelihood estimation method, whether exploratory or confirmatory, often will yield unacceptably high χ^2 values using traditional statistical criteria. To deal with this problem in confirmatory analyses, Bentler and Bonnet (53) have proposed two fit indices which are less sensitive to these problems. Sobel and Bohrnstedt (55) have suggested the limitations of their approach except for those cases where one has no prior knowledge of the factor structure. Since little systematic work has been done on the structure of measures of physical fitness, the models proposed by Bentler and Bonett would seem appropriate.

We will consider only one of the two methods of fit suggested by Bentler and Bonett, their normed fit index Δ:

$$\Delta_{0i} = 1 - \frac{\chi_i^2}{\chi_0^2}$$

where χ_i^2 is the chi-square observed for the model fit and χ_0^2 is the chi-square under the model of independence of one's measures, that is under the assumption of no factors. Such a model implies that all the observed covariances are zero. The normed fit index Δ_{0i} has the virtue of being bounded by 0 and 1.00, where 1.00 implies a perfect fit of the data to the hypothesized model and zero implies that the model of independence is the correct one. Since the standard error of Δ_{0i} is not known, only rules of thumb can be applied to determine the goodness of fit for a given hypothesized model. The "norm" that seems to have emerged states that values of .90 or higher imply a good fit. We will use the Bentler-Bonett coefficient in evaluating the models below.

Despite the fact that it is rare to find items that are congeneric, the congeneric model provides an ideal to be sought in choosing items. The best items are those that are "pure" with respect to a single underlying construct. If one attempts to choose items with the congeneric model in mind, then the theoretical validity of a composite score based on them should be high.

If one's items are congeneric, they can be combined into a linear composite:

$$y = \mathbf{w}'\mathbf{x} \tag{38}$$

where $\mathbf{w}'$ is a 1 × n vector of weights applied to the n ×

I-item vector **x**. Now substitution of equation 36 into 38 yields

$$y = \mathbf{w}'\mu + (\mathbf{w}'\beta)\tau + \mathbf{w}'\epsilon \qquad (39)$$

Taking the variance of equation 39 and applying the definition of reliability given in equation 15 Jörkeskog (43) shows that the reliability of y is given by

$$\rho = \mathbf{w}'\beta\beta\mathbf{w}/\mathbf{w}'\mathbf{s}\mathbf{w} \qquad (40)$$

He further shows that ρ is a maximum when the w_i are chosen to be proportional to β_i/θ_i^2. In the examples below, we have fixed all of the $w_i = 1.00$, because most composites are based on items that are unit weighted.

Let's now look at how the seven physical activity items fit a congeneric model for the males and females separately, using the data from table 1. The data were analyzed using LISREL VI (56). The results are shown in table 2.

For the males, first note that $\chi^2_{14} = 82.92$ (n.s.) indicating that the seven items do not fit the congeneric model well, at least using traditional statistical criteria to evaluate fit. Nevertheless, for pedagogical reasons we have estimated the reliability of a composite based on the seven items, under the assumption that they are congeneric, and have computed the item reliabilities as well. Notice that the reliability estimated is .65, which is virtually identical to that estimated with α. Note further, that three of the items have quite low β_i as well—WALK, HOBBY, and BIKE. And since the β_i are used in computing item reliabilities, the reliabilities of these three items are low as well. It should be pointed out that if the items were accounted for by a congeneric model, the estimate of the reliability of the composite is an estimate of the true reliability. However, when the data do not conform to a congeneric model, as is true in this case, then the reliability estimate is a lower bound of the composite's reliability.

Table 2. Fitting the 7 physical activity items to a congeneric model

Item	β_i	Θ_i^2	ρ_i
	Males (N = 988)		
SWIM	.554*	.825*	.271
WALK	.266*	.929*	.051
HOBBY	.297*	1.073*	.076
JOG	.671*	.777*	.367
BIKE	.350*	.925*	.117
EXERCISE	.678*	.942*	.328
OTHSPORT	.678*	.980*	.319
	Females (N = 1,490)		
SWIM	.565*	1.036*	.235
WALK	.402*	.793*	.169
HOBBY	.403*	.976*	.143
JOG	.529*	.695*	.287
BIKE	.480*	.953*	.195
EXERCISE	.607*	1.009*	.267
OTHSPORT	.687*	.871*	.351

*Significant at the .05 level.

Note: Male: $\chi^2_{14} = 82.92$, $p < .000$, $\rho = .65$, $\alpha = .64$. Female: $\chi^2_{14} = 63.09$, $p < .000$, $\rho = .68$, $\alpha = .68$

Since a single factor model does not fit the data well, we can ask how many factors are needed to account for the covariances for the males. The most expeditious way to answer this question is to factor the items using a maximum likelihood *exploratory* factor analysis extracting one, two, and three factors, respectively. This strategy was followed in analyzing the data for the males in table 1 using EFAP II (57).

Since maximum likelihood is almost always used (although other estimation options exist within the program) to estimate the parameters when using EFAP, the use of the χ^2 value as a fit statistic has the same problems associated with it as it does in LISREL, i.e., it is sensitive to both sample size and to violations of multivariate normality. To deal with this problem, Tucker and Lewis (58) propose a "reliability" coefficient that is designed to measure the variance in the observed covariances that is due to the common factors. The coefficient is given by

$$\rho = (M_k - M_{(k+1)})/(M_k - 1) \qquad (41)$$

where $M_k = \chi^2_k/df_k$ and $M_{(k+1)} = \chi^2_{(k+1)}/(df_{(k+1)}$. The χ^2 and degrees of freedom are those associated with maximum likelihood factor analytic solutions where k and $k + 1$ common factors have been extracted. Equation 41 makes it clear that if the ratio of the chi-square to the degrees of freedom approaches unity for $k + 1$ factors, the Tucker-Lewis coefficient also approaches unity, although it is theoretically possible for the coefficient to in fact exceed unity. The results of the EFAP analyses are shown in table 3. As indicated earlier, no hard-and-fast rules exist for determining when factorization is complete, although it has been suggested that it should continue until the Tucker-Lewis coefficient is .90 or greater. Using this criterion, three factors are needed to fit the data. If three factors are extracted, the χ^2 is 9.09 with 3 degrees of freedom, which approaches but still does not meet the traditional $\alpha = .05$ level of significance.

Table 3. Maximum likelihood factoring of 7 physical activity items

Number of factors	χ^2	df	p	Tucker-Lewis ρ
		Males (N = 988)		
0	743.40	21	***	—
1	82.92	14	***	.86
2	37.93	8	***	.89
3	9.09	3	.028	.94
		Females (N = 1,490)		
0	1157.54	21	***	—
1	63.09	14	***	.94
2	25.99	8	.001	.96
3	4.07	3	.254	.99

As indicated earlier, the Heise-Bohrnstedt (42) procedure allows one to estimate the theoretical validity and invalidity of a composite. In this case, the validity of the composite is .825 and the invalidity is .0006, which suggests that it is the first factor that is contributing most of the reliable variance to the composite. Although the contributions of the second and third factor are statistically significant, their contribution to the reliable variance is small.

Taken as a whole, what conclusions can we draw about these items? If we assume that a general latent variable called "frequency of participation in leisure-time activity" underlies these items, then it is clear that three of the seven items are not good measures for the males. Indeed, it can be shown that the reliability for a composite composed of the four best items—SWIM, JOG, EXERCISE, and OTHSPORT—is slightly higher than for the seven item score, .66 compared to .65. Therefore, if one were to compute a composite measure of the frequency of participation in leisure-time activity from these items, it should probably include only those four items.

How do the results of the analyses for the women compare to those for the men? As the results in table 2 indicate, the coefficients for all of the items are more similar in size than was true for the data on males. While the best items are the same four items that were best in the data for males, SWIM, JOG, EXERCISE, and OTHSPORT, the coefficients for the other three items were not nearly as small as the coefficients for these items in the data on the males. Furthermore, note from table 1 that while $\chi^2 = 63.09$, which though not statistically significant, is smaller suggesting that the items conform better to a congeneric model than did the data for the males. The reliability of a composite built on the assumption that the data are congeneric is .68, which is identical with the estimate provided by α but is slightly larger than that for the males. An examination of table 3 reveals that, although three factors are required to generate an acceptable statistical fit, in fact, the single factor model provides a good fit using the Tucker-Lewis coefficient, which equals .94. Furthermore, the Heise-Bohrnstedt estimated theoretical validity for the composite is .824 and the estimated invalidity is .0006 further suggesting that a single factor can account for most of the reliable variance.

What general conclusions can be drawn from the congeneric model from these analyses? First, whereas the assumptions of the congeneric model are certainly more compatible with the kinds of items and tests used by researchers doing research on physical activity than are those associated with parallel or τ-equivalent measures, it becomes clearer and clearer as we analyze more and more data sets that if one has more than four or five items, it is difficult to fit them to a congeneric model. In spite of the lack of statistical fit and even though the data may not fit the model at conventional levels of statistical significance, if one chooses one's measures carefully, they often will fit the congeneric model using fit measures less sensitive to sample size and distributional assumptions than the chi-square statistic, measures such as the Tucker-Lewis (58) and the Bentler and Bonett (53) measures of goodness of fit. Second, the moral seems clear. While the tools for assessing internal consistency reliability have become more refined and useful, what counts most in the final analysis is care and thoughtfulness in the design of the measures in the first place. As the analysis of the physical activity items for the males makes clear, a set of measures may be assumed to be measuring a single underlying latent variable, but only through careful analyses can one determine whether they are in fact measuring that construct.

Multiple group comparisons

One of the features available in LISREL, and in some of the other computer programs that allow one to analyze structural equation models with unobservable latent variables, is the ability to test the hypothesis that the factor structure, or subparts of the structure, are invariant across multiple populations.

We will not dwell on the technical aspects of the analysis here. The interested reader is encouraged to read the LISREL User's Manual (56) for the technical details. However, the general factor analysis model is as follows:

$$\mathbf{x} = \Lambda f + \epsilon \tag{42}$$

where x is a $n \times 1$ column vector of response variables, Λ is a $n \times m$ matrix of factor coefficients (or loadings as they are often called) and ϵ is a $n \times 1$ column vector of errors in measurement. If Σ is a $n \times n$ covariance matrix, then it can be shown that

$$\Sigma = \Lambda \, \Phi \, \Lambda' + \Theta_\epsilon \tag{43}$$

where Φ is a $m \times m$ matrix of covariances among the latent variables and Θ_ϵ is a $n \times n$ matrix of measurement variances. Ordinarily Θ_ϵ is taken to be a diagonal matrix, but LISREL allows one to relax this assumption and in fact to allow for correlated measurement errors. What equation 43 states is that the covariances can be seen as a function of three sets of parameters—the factor coefficients Λ, the interrelationships among the latent variables Φ, and the covariances, if any, of the measurement errors Θ_ϵ.

If the factor structure were the same across q populations, then the elements of Λ, Φ, and Θ_ϵ would all be invariant across the populations. In fact, it would be rare to find such a good fit of the data to a hypothesized model. To be able to make cross population comparisons, at a minimum, one must be able to demonstrate that the factor coefficients, the Λ, are invariant across groups. If these cannot be shown to be invariant, the implication is that the items are not equally good indicators of the latent variables across groups. This has the important implication that one cannot make meaningful comparisons across groups using the means and standard deviations of a composite built from the items. If the factor coefficients are not invariant, it implies that the "meaning" of the items vis-a-vis the latent variables differs. Importantly, some or all of the items may still be measuring the same construct across groups, but if the analyses suggest that some of the items should be in the construct in population 1, a different subset of them in population 2, and so on, computing means on the composites and comparing them across groups makes no sense.

In carrying out multiple group analyses we do a series of statistical tests. First we do a chi-square test to determine whether the covariance matrices differ across groups. If we reject the test that they are the same across groups, then we undertake a set of hierarchical tests. First we ask whether a model where all of the free parameters are unconstrained fits the data any better than a model where the factor coefficients Λ are constrained to be equal across groups. If one accepts the invariance hypothesis for the factor coefficients, then it is possible to build meaningful composites for use in across-group comparisons. If we must reject the invariance hypothesis for the Λ, as pointed out above, the items appear not to be measuring the same constructs across groups. Assuming, however, that the invariance hypothesis for the elements of Λ hold, next we test to see if the Φ matrices can be considered invariant, i.e., whether we can accept the hypothesis that the correlations among the latent variables are the same across populations. Finally, if the hypothesis of invariant matrices holds across populations, we move on to the final test, to see if the error variances can be considered to be equal, by constraining the elements of the Θ_ϵ matrices to be equal.

We did a multiple group analysis of the seven physical activity items in table 1, comparing the factor structures of the males and the females. The overall test for equality of the two covariance matrices yielded a $\chi^2 = 77.0$ (n.s.), which means that we must reject the hypothesis that the two covariance matrices are equal. Next we tested a model where only the elements of the factor coefficient matrices, the two Λ matrices, were constrained to be equal with a model where they were free. The χ^2 for the model where the Λ were free across the two groups was 146.50 with 28 degrees of freedom. When the two Λ matrices were constrained to be equal, $\chi^2_{28} = 172.24$. To determine whether these two models differ statistically, we must do a difference in χ^2 test. Recall that the difference between two χ^2 is itself a χ^2 with degrees of freedom equal to the difference in the degrees of freedom associated with the two χ^2 that make up the χ^2 difference test. In this case the difference is $\chi^2_6 = 25.74$. This difference is statistically significant far beyond even the $p = .001$ level, which means that we cannot accept the hypothesis that the Λ matrices are the same for the males and the females. This clearly supports the conclusion reached in the previous section that the seven physical activity items are not measuring exactly the same construct for the males as for the females. As a result, comparing males and females on a composite built from all seven items makes little sense. Because the items, SWIM, JOG, EXERCISE, and OTHSPORT both have relatively high loadings for the males and the females, they could be combined into a composite that would allow for cross-sex comparisons. However, as pointed out in the section immediately above, the reliability of the construct as estimated by α is only .66 for the males. For the females, the reliability of the four-item construct is only .62 (compared to $\alpha = .68$ for a composite built on all seven items). Sadly, the only conclusion one can draw is that these items cannot be considered as good indicators of the frequency of participation in leisure-time physical activity given their poor fit to a congeneric model and the low reliability estimates of composites built from them.

Ideally we would have had a set of data for this paper that would have shown more clearly how physical fitness or physical activity, or one of the subdomains of these concepts, fit a congeneric model perfectly or near perfectly. But such was not the case. What the analyses did indicate quite clearly, however, is the need for researchers in the area of physical fitness and physical activity to think more clearly about what it is they mean when they discuss physical fitness and activity and their measurement. Furthermore, we were unable to find a set of data to analyze where sets of tests had been constructed that would test whether it makes sense to think of physical fitness as a set of interrelated latent variables in the spirit of the model shown in figure 3. Ideally, someone will seriously explore whether it makes theoretical sense to think of fitness as a multidimensional construct with a latent structure. If so, we believe that the techniques described in this paper, as well as the other available latent variable models could be invaluable in helping to better measure physical fitness.

Some concluding comments

Our purpose has been to provide the reader with an overview of measurement theory that stresses the potential importance of conceiving of physical fitness and physical activity as latent variables. The knowledgeable reader will have discovered that our treatment is incomplete in at least two ways. First, whereas we have concentrated on how structural equation models with latent variables can be used to evaluate sets of tests and measurements chosen that putatively measure one or more subdomains related to a central construct, the techniques associated with structural equation modelling, such as LISREL VI (56), allow one not only to evaluate measurement models but to imbed the measurement models in a set of structural equations that link the latent variables to one another. We strongly recommend that the reader interested in latent variable models read further in the area. There are books and articles available at both the beginner's level (59, 60) and at the more advanced level as well (61–64). Second, we have only considered latent variable models where the latent variables and their observed measures are both assumed to be continuous variables. Increasingly, applied statisticians working in the area of latent variable modelling have been concerned about the limitations of such models. One concern has been that in LISREL the observed measures are assumed to follow a multivariate normal distribution. Two recently released computer programs allow one to relax this assumption, using arbitrary distribution theory methods (65, 66) if one so chooses.

Another limitation has been the assumption that one's observations are at the interval or ratio level of measurement. A program called LISCOMP (67) has recently been released, which estimates the measurement part of a structural equation model assuming no more than ordered categorical variables. Muthen and Short (chapter 28) use LISCOMP to analyze the same set of seven physical activity items we analyzed above using LISREL. The results are very similar in this case, but if variables measured at the ordinal level are badly skewed, the results can be quite different when using LISREL instead of LISCOMP. Although a program has not yet been released, the estimation technique has been developed (68) that will allow one to use *any* combination of interval, ordinal, or nominal level indicators, again under the assumption that the latent variable is continuous.

Another group of models allows one to assume that the latent variable is categorical rather than continuous (69–71) and is called *latent class analysis.* A very readable text at the introductory level has recently been released (72). In addition, McCutcheon presents a short overview and an example using latent class analysis in chapter 29. Although not available commercially, a program to do latent class analysis, called MLLSA, is available from its author, Clifford Clogg (73).

As the reader can see, a whole armament of techniques is now available for the analysis of models that assume one or more latent variables depending upon what assumptions one needs to make about (a) the measurement level of the latent variables themselves and (b) the measurement level of the observed variables. Our suspicion is that thinking of physical fitness and physical activity as latent constructs will have a payoff for researchers in this area.

References

1. Coombs, C. H., R. M. Dawes and A. Tversky (1970). *Mathematical Psychology: An Elementary Introduction.* Englewood Cliffs, NJ: Prentice-Hall.
2. Krantz, D. H., R. Luce, P. Suppes and A. Tversky (1971). *Foundations of Measurement, Vol. I.* New York: Academic Press.
3. Pfanzagl, J. (1971). *Theory of Measurement* (2nd ed.). New York: Wiley.
4. Roberts, F. S. (1979). *Measurement Theory.* Reading, MA: Addison-Wesley.
5. Stevens, S. S. (1959). Measurement, psychophysics, and utility. In C. W. Churchman and P. Ratoosh (Eds.), *Measurement: Definitions and Theories.* New York: Wiley.
6. Suppes, P. and J. L. Zinnes (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 1, pp. 1–76). New York: Wiley.
7. Coombs, C. (1964). *Theory of Data.* New York: Wiley.
8. Torgerson, W. S. (1958). *Theory and Methods of Scaling.* New York: Wiley.
9. Luce, R. D. and L. Narens (1987). Measurement scales on the continuum. *Science,* 236, pp. 1527–1532.
10. Borgatta, E. F. and G. W. Bohrnstedt (1981). Level of measurement: Once over again. In G. W. Bohrnstedt and E. F. Borgatta (Eds.), *Social Measurement: Current Issues.* Beverly Hills: Sage.
11. Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin,* 87, pp. 564–567.
12. Stevens, S. S. (1975). *Psychophysics: Introduction to its Perceptual, Neural and Social Prospects.* New York: Wiley.
13. Thomas, H. (1982). IQ, interval scales, and normal distributions. *Psychological Bulletin,* 91, pp. 198–202.
14. Townsend, J. T. and G. A. Ashby (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin,* 96, pp. 394–401.
15. Cliff, N. (1982). What is and isn't measurement. In G. Keren (Ed.), *Statistical and Methodological Issues in Psychological and Social Science Research.* Hillsdale, NJ: Erlbaum.
16. Duncan, O. D. (1985). *Notes on Social Measurement.* Beverly Hills: Sage.
17. Zimmerman, D. W. (1975). Probability spaces, Hilbert spaces, and the axioms of test theory. *Psychometrika,* 40, pp. 395–412.
18. Lord, F. M. and M. Novick (1968). *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley.
19. Bohrnstedt, G. W. (1969). Observations in the measurement of change. Pp. 113–136 in E. F. Borgatta (Ed.), *Sociological Methodology: 1969.* San Francisco: Jossey-Bass.
20. Warren, R. D., J. K. White and W. A. Fuller (1974). An errors in variables analysis of managerial role performance. *Journal of the American Statistical Association,* 69, pp. 886–893.
21. Fuller, W. A. and M. A. Hidiroglou (1978). Regression estimates after correction for attenuation. *Journal of the American Statistical Association,* 73, pp. 99–104.
22. Bohrnstedt, G. W. (1969). A quick method for determining the reliability and validity of multiple item scales. *American Sociological Review,* 34, pp. 542–548.
23. Curtis, R. F. and E. F. Jackson (1962). Multiple indicators in survey research. *American Journal of Sociology,* 68, pp. 195–204.
24. Blalock, H. M. (1970). Estimating measurement error using multiple indicators and several points in time. *American Sociological Review,* 35, pp. 101–111.
25. Hannan, M. T., R. Rubinson and J. T. Warren (1974). The causal approach to the measurement error in panel analysis: Some further contingencies. Pp. 293–323 in H. M. Blalock, Jr., *Measurement in the Social Sciences.* Chicago: Aldine.
26. Jöreskog, K. G. and D. Sörbom (1979). *Advances in Factor Analysis and Structural Equation Models.* Cambridge, Massachusetts: Abt Books.
27. Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review,* 34, pp. 93–101.
28. Wiley, D. E. and J. A. Wiley (1970). The estimation of measurement error in panel data. *American Sociological Review,* 35, pp. 112–117.
29. Werts, C. E., K. G. Jöreskog and R. L. Linn (1971). Comment on 'the estimation of measurement error in panel data.' *American Sociological Review,* 36, pp. 110–112.
30. Wiley, J. A. and M. G. Wiley (1974). A note on correlated errors in repeated measurements. *Sociological Methods and Research,* 3, pp. 172–188.
31. Hargens, L. L., B. F. Reskin and P. D. Allison (1976). Problems in estimating error from panel data: An example involving the measurement of scientific productivity. *Sociological Methods and Research,* 4, pp. 439–458.
32. Bohrnstedt, G. W. (1983). Measurement. Pp. 69–121 in P. Rossi, J. Wright and A. Anderson (Eds.), *Handbook of Survey Research.* New York: Academic Press.

33. Kuder, G. F. and M. W. Richardson (1973). The theory of the estimation of test reliability. *Psychometrika,* 2, pp. 135–138.
34. Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika,* 6, pp. 153–160.
35. Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika,* 10, pp. 255–282.
36. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika,* 16, pp. 297–334.
37. Novick, M. R. and C. Lewis (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika,* 32, pp. 1–13.
38. Danchik, K. M. (1981). Highlights from Wave I of the National Survey of Personal Health Practices and Consequences: United States, 1979. *Vital and Health Statistics,* Series 15, No. 1. Washington: Government Printing Office.
39. Harmon, H. H. (1967). *Modern Factor Analysis* (2nd edition). Chicago: University of Chicago Press.
40. Cattell, R. B. and J. Radcliffe (1962). Reliability and validity of simple and extended weighted and buffered unifactor scales. *British Journal of Statistical Psychology,* 15, pp. 113–128.
41. Bentler, P. M. (1968). Alpha-maximized factor analysis (Alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika,* 33, pp. 335–346.
42. Heise, D. R. and G. W. Bohrnstedt (1971). Validity, invalidity and reliability. Pp. 104–129 in E. F. Borgatta and G. W. Bohrnstedt (Eds.), *Sociological Methodology: 1971.* San Francisco: Jossey-Bass.
43. Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika,* 36, pp. 109–134.
44. Allen, M. P. (1974). Construction of composite measures by the canonical factor regression method. Pp. 51–78 in H. L. Costner (Ed.), *Sociological Methodology: 1973–74.* San Francisco: Jossey-Bass.
45. Smith, K. W. (1974). On estimating the reliability of composite indexes through factor analysis. *Sociological Methods and Research,* 2, pp. 485–510.
46. Smith, K. W. (1974). Forming composite scales and estimating their validity through factor analysis. *Social Forces,* 53, pp. 168–180.
47. Greene, V. L. and E. G. Carmines (1980). Assessing the reliability of linear composites. Pp. 160–175 in K. Schuessler (Ed.), *Sociological Methodology: 1980.* San Francisco: Jossey-Bass.
48. Armor, D. J. (1974). Theta reliability and factor scaling. Pp. 17–50 in H. L. Costner (Ed.), *Sociological Methodology: 1973–74.* San Francisco: Jossey-Bass.
49. Levine, L. M. and G. W. Bohrnstedt (1982). New developments in the assessment of reliability: The LINREL programs. *Educational and Psychological Measurement,* 42, pp. 195–199.
50. American Psychological Association (1974). *Standards for Educational and Psychological Tests.* Washington: American Psychological Association.
51. Guttman, L. (1959). A structural theory for intergroup beliefs and action. *American Sociological Review,* 24, pp. 318–328.
52. Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika,* 34, pp. 183–202.
53. Bentler, P. M. and D. G. Bonett (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin,* 88, pp. 588–606.
54. Boomsma, A. (1983). *On the Robustness of LISREL (Maximum Likelihood Estimation Against Small Sample Size and Nonnormality.* Ph.D. Dissertation, University of Groningen.
55. Sobel, M. E. and G. W. Bohrnstedt (1985). Pp. 152–178 in N. Tuma (Ed.), *Sociological Methodology: 1985.* San Francisco: Jossey-Bass.
56. Jöreskog, K. G. and D. Sörbom (1984). *LISREL VI. Analysis of Linear Structural Relationships by the Method of Maximum Likelihood.* Mooresville: Scientific Software.
57. Jöreskog, K. G. and D. Sörbom (1978). *EFAP (Version II). Exploratory Factor Analysis Program.* Mooresville: Scientific Software.
58. Tucker, L. R. and C. Lewis (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika,* 38, pp. 1–10.
59. Long, J. S. (1983). *Confirmatory Factor Analysis.* Beverly Hills: Sage.
60. Long, J. S. (1983). *Covariance Structure Models: An Introduction to LISREL.* Beverly Hills: Sage.
61. Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal Modelling. *Annual Review of Psychology,* 31, pp. 419–456.
62. Bentler, P. M. (1980). Linear structural equations with latent variables. *Psychometrika,* 45, pp. 289–308.
63. Everitt, B. S. (1984). *An Introduction of Latent Variable Models.* London: Chapman and Hall.
64. Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation, and testing. Pp. 265–287 in P. R. Krishnaiah (Ed.), *Applications of Statistics.* Amsterdam: North-Holland.
65. Bentler, P. M. (1985). *Theory and Implementation of EQS: A Structural Equations Program (Version 2.0).* Los Angeles: BMDP Statistical Software.
66. Schoenberg, R. (1987). *LINCS. Linear Covariance Structure Analysis.* Kensington: RJS Software, Inc.
67. Muthen, B. (1987). *LISCOMP. Analysis of Linear Structural Equations with a Comprehensive Measurement Model.* Mooresville: Scientific Software.
68. Arminger, G. (1986). Latente Variablen Modelle auf der Basis von Mischverteilungen. Pp. 483–506 in M. Beckmann, K. Gaede, K. Ritter and H. Schneeweiss (Eds.), *Methods of Operations Research,* Part I.
69. Lazarsfeld, P. F. and N. W. Henry (1968). *Latent Structure Analysis.* Boston: Houghton Mifflin.
70. Goodman, L. A. (1979). On the estimation of parameters in latent structure analysis. *Psychometrika,* 44, pp. 123–128.
71. Clogg, C. C. and L. A. Goodman (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association,* 79, pp. 762–771.
72. McCutcheon, A. L. (1987). *Latent Class Analysis.* Newbury Park: Sage.
73. Clogg, C. C. (1977). Unrestricted and restricted maximum likelihood latent structure analysis: A manual for users. Working paper 1977–09. Population Issues Office, Pennsylvania State University, University Park, PA.

Chapter 28

Applying Regression and Factor Analysis of Categorical Variables to Fitness and Exercise Data

Bengt Muthén, Ph.D.
Lynn Short, M.A.
University of California, Los Angeles

In this chapter we will describe new statistical methodology for categorical data and illustrate how this methodology can be applied to health data, including measurements of healthy lifestyles, fitness and exercise. Because the methodology for categorical variables is not a familiar area to many researchers, we will start with relatively simple models that will be gradually made more complex.

Regression analysis with categorical response

In order to fix ideas, let us consider some data from the National Survey of Personal Health Practices and Consequences (NSPHPC) regarding alcohol consumption in the United States (1). This data was collected in two waves; the first was in the spring of 1979, with a reinterview of the same respondents 1 year later. Consider the response to the question, "Do you drink alcoholic beverages?", recording the respondent as nondrinker ($y = 0$) or drinker ($y = 1$). As an explanatory (independent, exogenous) variable we may for illustrative purposes consider age recorded as 20–34 years ($x = 1$), 35–44 years ($x = 2$), 45–54 years ($x = 3$), 55–64 years ($x = 4$), 65 years and over ($x = 5$). Table 1 gives some descriptive statistics for these variables. We note that the proportion of drinkers tends to decrease with age, although for males there is a certain increase at 55–64 years.

Dichotomous response

Although we will treat the x variable as an "ordinary" interval scaled variable, we will consider the 0/1 y variable values merely as labels for the response categories. We will then consider a model that concerns itself with the probability of observing a certain y category given a certain x value, rather than the standard regression approach that concerns the prediction of y values. This leads to a standard so-called probit or logit regression model (see, e.g., (2–4)). Here we will consider two alternative formulations of the probit model—the conditional probability curve formulation and the latent response variable formulation. The latter formulation will be the one that we use when generalizing the analysis to multivariate response and factor analysis modeling.

Conditional probability formulation

In the standard formulation of the probit model we assume,

$$\Pr(y = 1|x) = \Phi(\alpha + \beta x) \tag{1}$$

Table 1. Descriptive statistics for drinking related to age

	Age group					
	20–34 years	35–44 years	45–54 years	55–64 years	Total	Percent
Female						
Total	251	150	121	134	656	
Percent drinker	91.2	78.7	62.8	57.5		
Nondrinker	22	32	45	57	156	23.8
Drinker	229	118	76	77	500	76.2
Male						
Total	245	163	120	64	592	
Percent drinker	91.8	88.3	62.5	87.5		
Nondrinker	20	19	45	8	92	15.5
Drinker	225	144	75	56	500	84.5

where Φ stands for the standard normal distribution function and α and β are the parameters to be estimated. The function is shown in figure 1. Hence, the model explicitly recognizes that as opposed to ordinary linear regression the conditional expectation function (equation 1) is nonlinear in order to accommodate the requirement of predicted ys falling between 0 and 1.

In terms of our example, we would assume a negative value of β, such that the probability of the registered response "drinker" decreases with increasing age category. The nonlinear conditional probability function makes it possible to describe the plausible situation that an age category increase from $x = 1$ to $x = 2$ perhaps does not correspond to as large a decrease in probability of "drinker" as the increase from $x = 3$ to $x = 4$.

Latent response variable formulation

In this formulation we consider a latent response variable for each individual,

$$y^* = \gamma x + \zeta \tag{2}$$

where γ is a parameter and ζ is a random residual variable. For a given x value, the latent response variable values for different individuals may not be the same because of other factors uncorrelated with x. There is also a threshold parameter on the latent response variable, such that

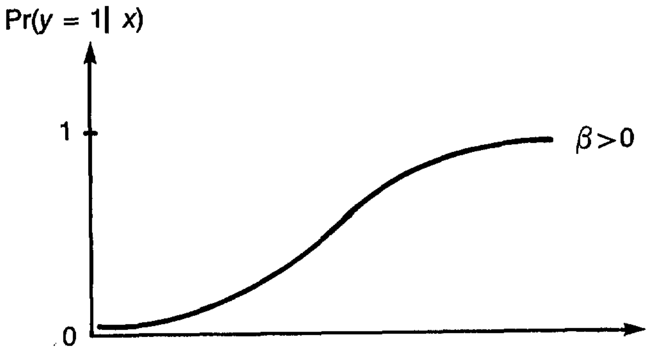


Figure 1. Conditional probability curve formulation.

$$y = \begin{cases} 1 & \text{if } y^* > \tau \\ 0 & \text{if } y^* \leq \tau \end{cases} \tag{3}$$

We also assume that ζ has mean zero, and variance one is independent of x and is normally distributed. This situation is shown in figure 2. It then follows that

$$\begin{aligned} \Pr(y=1|x) &= \Pr(y^* > \tau|x) \\ &= \Phi\,(-\tau+\gamma x) \end{aligned} \tag{4}$$

so that the conditional probability formulation and the latent response variable formulation have identical observed variable implications. The only difference between the models is a trivial reparametrization,

$$\begin{aligned} \alpha &= -\tau \\ \beta &= \gamma \end{aligned} \tag{5}$$

Returning to the alcohol example, we note that, for each age category, the model predicts a certain proportion of drinkers. If the model is correct in the population, transforming each of the corresponding population probabilities to standard normal values would yield "probits" falling along a straight line. In figure 3, this is carried out for the sample of females and males. It is seen that in both cases the sample probits reflect a

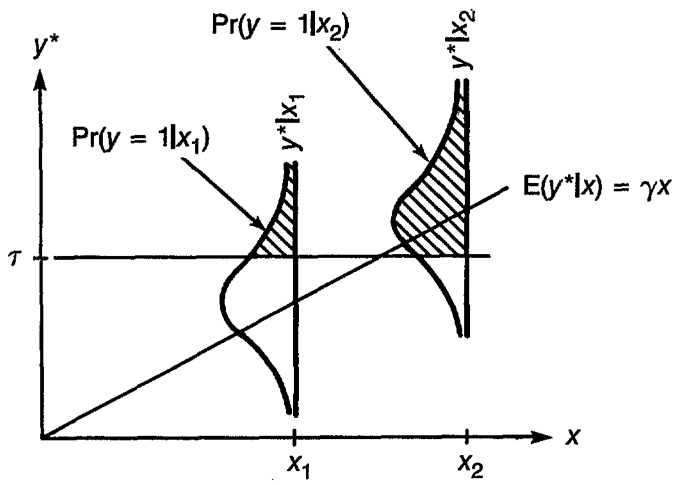


Figure 2. Latent-response variable formulation.

negative relationship. For females, linearity seems to hold very well; for males, there seems to be a nonlinearity because of the dip in drinking propensity for the category 45–54 years. This is a sex difference which, if significant, would be of interest to investigate further from a substantive point of view. The probit regression slope estimates are −0.395 for females, with a standard error of 0.050, and −0.272 for males, with a standard error of 0.064. We note that the probit slopes obtain significant negative values; the ratio of estimated slope to standard error is approximately normally distributed in large samples.

Ordered polytomous response

In addition to drinking-nondrinking information, the number of drinks consumed at one sitting was recorded. The category labels are 1 for 1–2 drinks per sitting, 2 for 3–4 drinks per sitting, and 3 for 5 or more drinks per sitting. Taken together with the 0 category of nondrinkers, we then have an ordered, four-category variable. This y variable will be called ALAMT.

The probit model is readily extended to describe this situation. We may consider the latent response variable formulation of (2), where in this case

$$y = \begin{cases} 3 & \text{if } y^* \geq \tau_3 \\ 2 & \text{if } \tau_2 \leq y^* < \tau_3 \\ 1 & \text{if } \tau_1 \leq y^* < \tau_2 \\ 0 & \text{if } y^* < \tau_1 \end{cases} \tag{6}$$

The latent response variable, which may be thought of as a "propensity to drink," hence has a set of thresholds, the values of which are increasing; figure 4 illustrates this situation. With polytomous response, it is of interest to consider the following conditional probabilities implied by the above model:

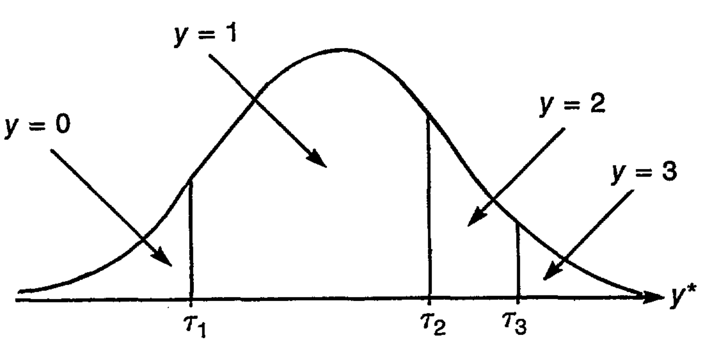


Figure 4. Latent-response variable and 4-category observation.

$$\Pr(y = 1 \text{ or } 2 \text{ or } 3 \mid x) = \Phi(-\tau_1 + \gamma x) \tag{7}$$

$$\Pr(y = 2 \text{ or } 3 \mid x) = \Phi(-\tau_2 + \gamma x) \tag{8}$$

$$\Pr(y = 3 \mid x) = \Phi(-\tau_3 + \gamma x) \tag{9}$$

Considering the arguments of the standard normal distribution function, it is once again seen that the population probits are linearly related to x. In this case, however, each x value gives rise to three probits so that there are three lines, the difference between which can be expressed in terms of the distances between the thresholds. For the distance between the 1 or 2 or 3 and the 2 or 3 probit lines, we have

$$\tau_2 - \tau_1 \tag{10}$$

and for the distance between the 2 or 3 and the 3 probit lines, we have

$$\tau_3 - \tau_2 \tag{11}$$

In words, this means that the three probit functions all have to be not only linear but also parallel. An important piece of information is contained in the distances between the lines. This gives us information about the

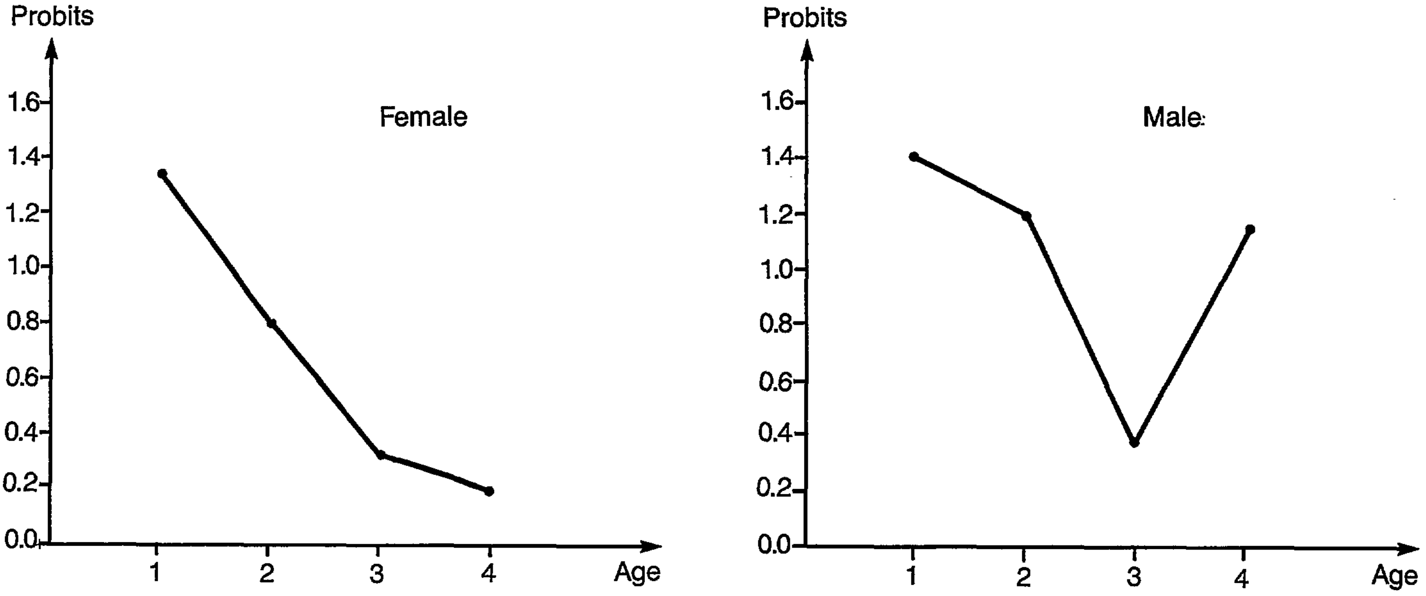


Figure 3. Probit for dichotomous dependent variable for drinking on age.

category widths on the y^* axis. Such information can be useful in inducing whether quantitative scoring of the categories is feasible.

ALAMT regressed on age

In figures 5 and 6 probit plots for ALAMT are given for the samples of females and males. It is seen in both figures that there is a larger difference between the "line" at the top for categories 1, 2, 3 and the other two lines at the bottom than between the two bottom lines. This means that on the latent-response variable (the "propensity to drink" variable) the $\tau_2 - \tau_1$ distance is larger than the $\tau_3 - \tau_2$ distance, so that we have the situation of figure 4. The substantive conclusion is that, when relating ALAMT to age, the step from not drinking to drinking 1–2 drinks per sitting is considerably larger than the step from drinking 1–2 drinks per sitting to drinking 3–4 drinks per sitting.

The probit slope estimates from the regression with polytomous response are -0.364 (s.e. = 0.040) for females and -0.227 (s.e. = 0.044) for males. The estimated differences for $\tau_2 - \tau_1$ and $\tau_3 - \tau_2$ are 1.778 and 0.863 for females and 1.595 and 0.546 for males. We conclude that, although the model seems to fit less well for males, there seems to be no large sex differences in the perception of the ALAMT response categories.

It should be noted from this example that using regression methodology explicitly developed for categorical response has a considerable advantage over traditional methods. If we instead assume that y (ALAMT) can be scored 0–3 and treated as an interval-scaled dependent variable in an ordinary regression, we would not have found the above-described difference in scale steps; the lack of equidistance would most likely have gone unnoticed.

ALAMT on income

Let us consider another example of the strength of categorical variable methodology over traditional methods not suited to these kinds of response variables. In this study, annual family income before taxes was recorded as <\$5,000 ($x = 1$), \$5,000–\$9,999 ($x = 2$), \$10,000–\$14,999 ($x = 3$), \$15,000–\$24,999 ($x = 4$), >\$24,999 ($x = 5$). The three sample probit plots for each sex is given in figure 7 for females and in figure 8 for males.

In each figure, the top line describes how the probability of drinking (at all) varies with income. At least for males there seems to be a clear positive relationship. The two bottom lines in each figure describe how drinking 2 or 3 and 3 drinks, respectively, varies with income. For these latter lines there seems to be no clear relationship

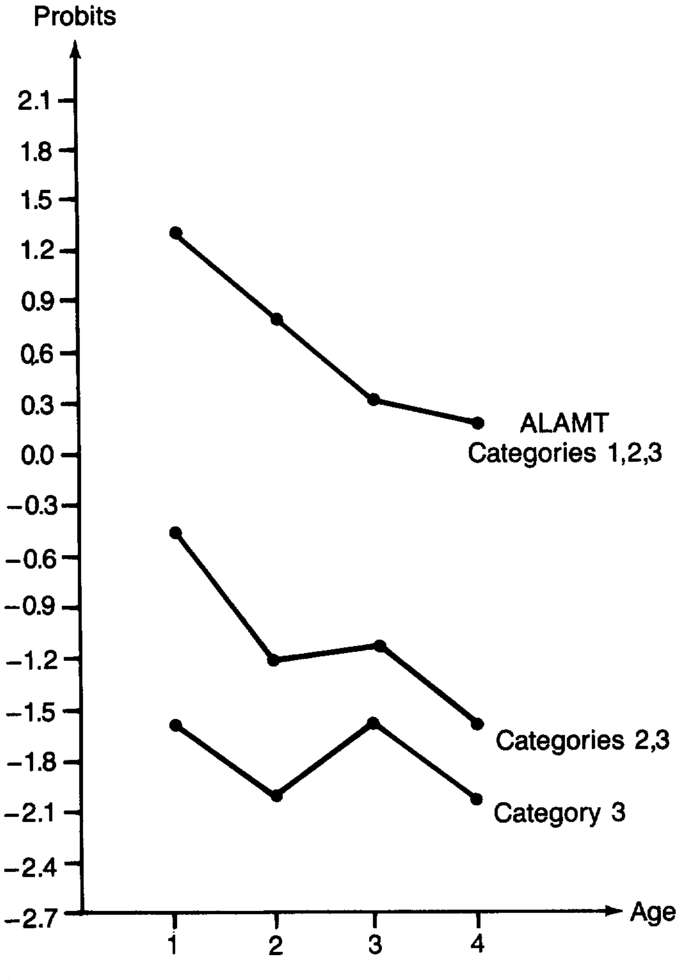


Figure 5. Probits for ALAMT on age for females.

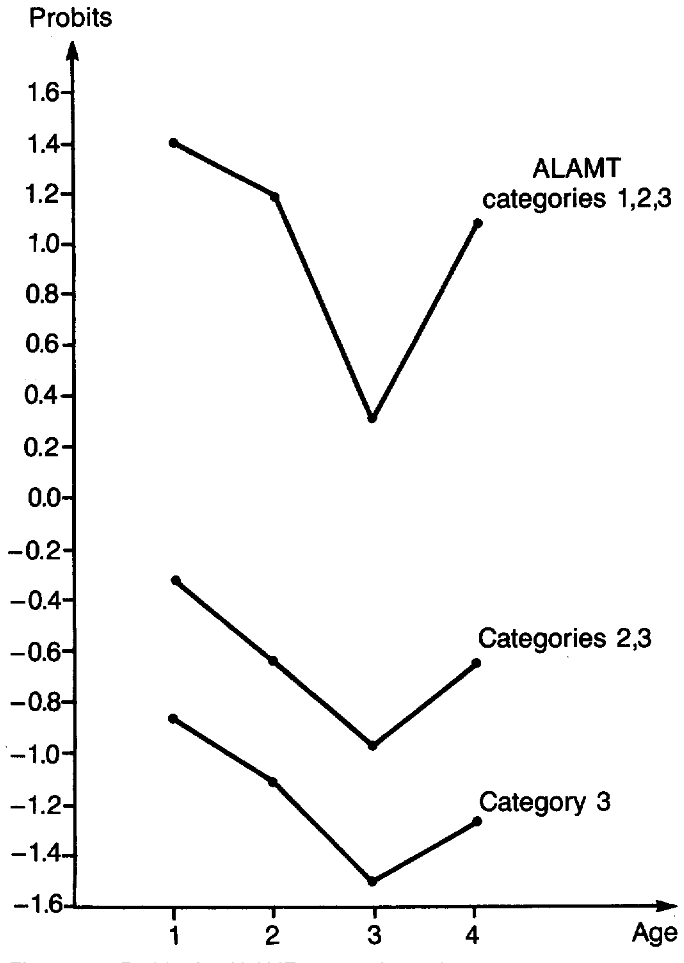


Figure 6. Probits for ALAMT on age for males.

with income at all. Hence, it may well be that the probability of drinking goes up with increasing income, but the amount consumed per sitting does not.

The ordered polytomous regression model presented above would not be directly suitable to this type of situation because the model requires parallel lines. Drinkers and nondrinkers would have to be described separately. There is, however, other categorical response variable methodology that could be used. There are so-called generalized Tobit models (5, 6) that describe the probability of drinking by one set of regression slopes and the amount of drinking by another set. Hence, one does not even have to assume that the same set of x variables are influential in the two relations. Such methods will not be described here, however.

At the same time, ordinary regression analysis treating ALAMT as an interval variable would most likely overlook the issue completely. In fact, such a regression analysis was carried out with an estimated slope of 0.174 (s.e. = 0.063) for females and 0.191 (s.e. = 0.039) for males. These estimates give the false impression of a very strong relationship between ALAMT and income, presumably mostly because of the influence income has on drinking at all.

Several good discussions of categorical variable methodology of the above type exist nowadays. In our presentation, we have solely worked with probit type models because they are easily generalizable to multivariate and latent variable extensions, such as factor analysis of dichotomous and ordered categorical variables. However, with univariate response, a more common model is perhaps the logit regression model, where the conditional probability curve of equation 1 is described by the logistic distribution function. The practical difference between these approaches is almost negligible for univariate response. A good introductory and application-oriented overview of various univariate response-modeling approaches is given in Aldrich and Nelson (7). They describe how to interpret estimates and make judgments about model fit. They also make clear what the dangers of using ordinary regression could be. An article by McKelvey and Zaviona (8) specifically deals with ordered polytomous response and includes a real data example showing the contrast between using correct methodology versus ignoring the measurement problem of the dependent variable. More technical treatments are given in the excellent overview article by Amemiya (9), in the comprehensive but still lucid econometric application-oriented book by Maddala (10), and in the research articles of Anderson (11) and McCullagh (12).

It should be noted that the bibliography contains a large number of citations in the field of categorical and latent variable modeling, extending beyond the refer-

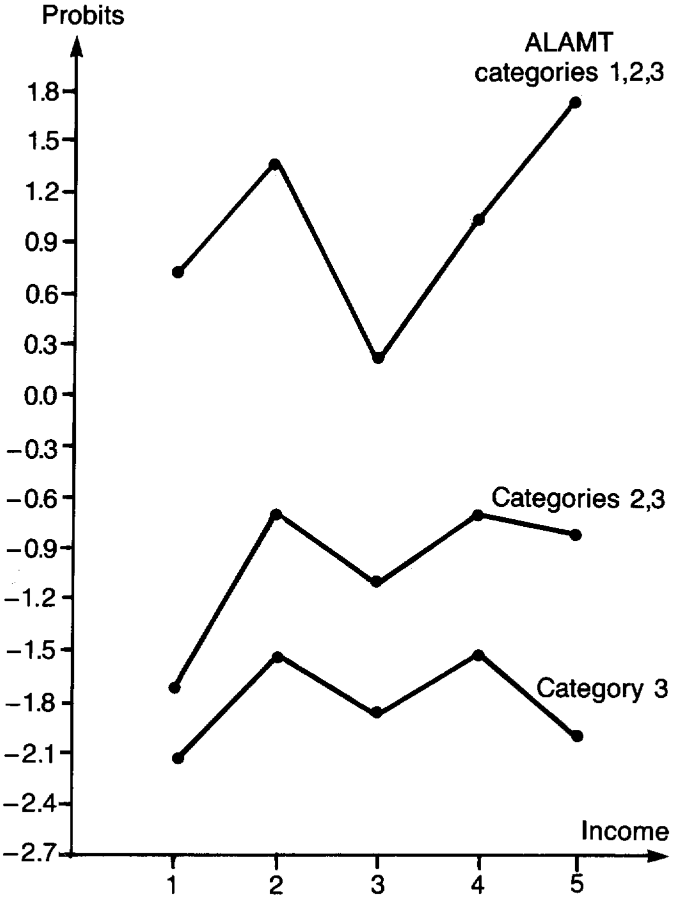


Figure 7. Probits for ALAMT on income for females.

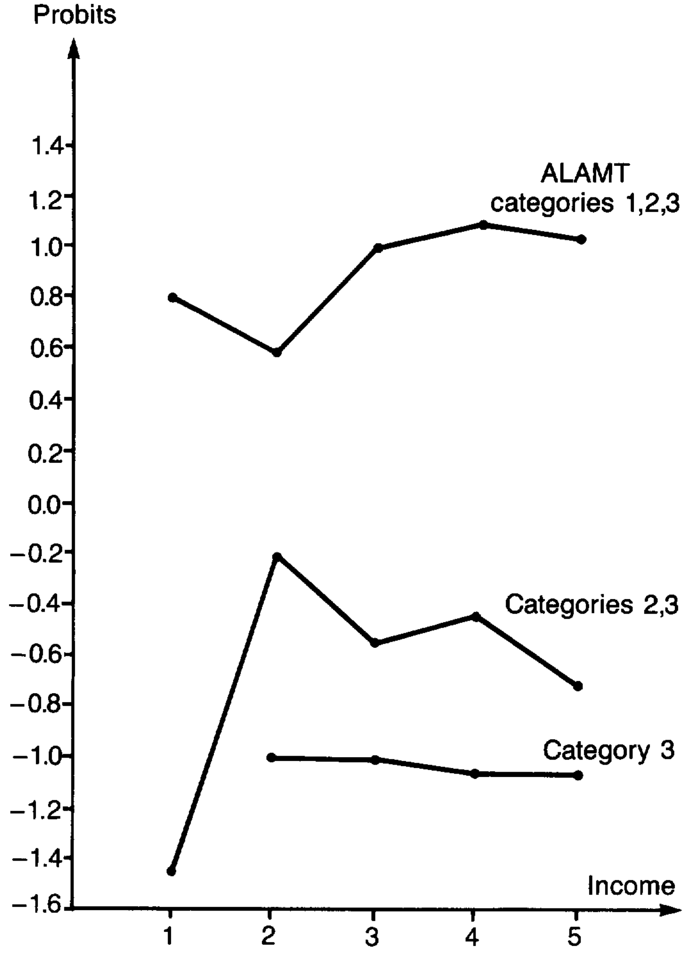


Figure 8. Probits for ALAMT on income for males.

ences discussed in the text. The readings should be of interest to a "health statistics" oriented readership, because most of the citations are from the social or behavioral sciences; borrowing ideas from neighboring sciences can often prove fruitful.

The above regression-modeling techniques are directly generalizable to multivariate response situations. In Muthén (13), a general methodology for the estimation of multivariate systems of equations was developed. Muthén (14) and Maddala (10) give overviews of various approaches in this area, largely drawing from econometric work. It may be noted that very powerful methodologies exist for the analysis of continuous variables in systems of causal equations, e.g., in so-called path analysis or simultaneous equation systems. In terms of applications with categorical variables, however, almost nothing has been done. This would seem to be an area of great future potential. In the section dealing with factor analysis, several references to multivariate modeling of this kind are however given for situations where latent variables are hypothesized.

Multivariate modeling may be of particular interest when longitudinal data are available. In figure 9, a so-called path diagram is shown as a hypothetical example for a possible longitudinal analysis. Dependent variables are alcohol amount (ALAMT) and the frequency of alcohol consumption (ALFRQ) measured at time 1 and time 2. Because of the possible differences in relationships between drinkers and nondrinkers, one may want to limit the analysis to drinkers. Both dependent variables then have three response categories. Possible independent variables include income, education, age, number of jobs where employed full time during past 5 years (JOBNO), and some physical activity measures (PHYSACT). Time 1 or time 2 measurements are considered. We make the assumption that time 2 dependent variables also depend on the corresponding time 1 measurements. Time 1 independent variables are assumed to remain relatively constant, assuming a short period between time 1 and time 2. Such a model can be estimated and tested in the framework of the Muthén (13) LISCOMP program. Specific hypotheses such as time-invariant independent variable slopes can also be tested. However, this will not be carried out here for lack of space.

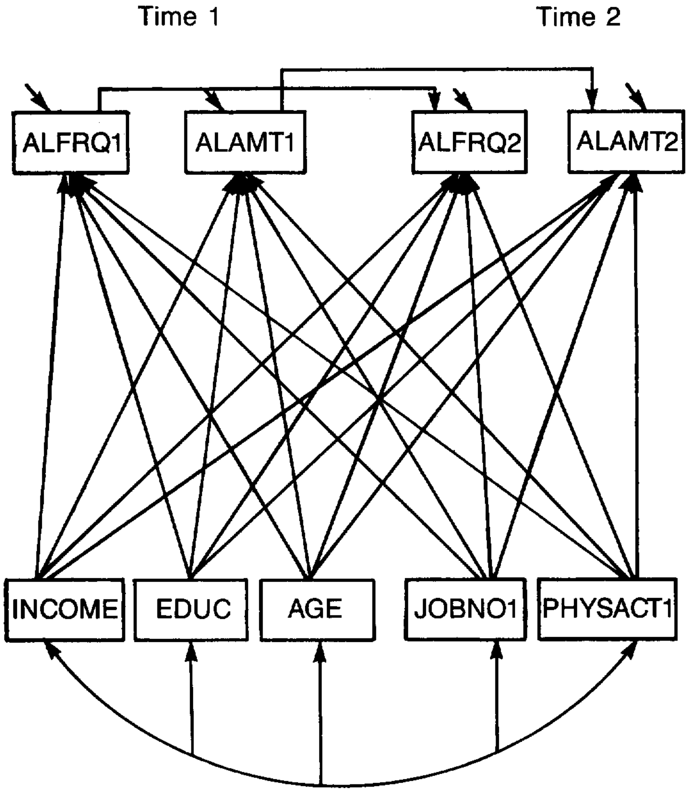


Figure 9. A hypothetical multivariate longitudinal model.

Factor analysis of categorical data

Using the same data set as previously, consider responses to the questions, "How often in the past month have you felt: Cheerful and light hearted? Loved and wanted? Downhearted and blue? Lonely?" The response categories for all four of these questions were: never, rarely, sometimes, and very often. Don't know and refused were additional categories. These responses were utilized in creating a single HAY (how are you?) score. In creating this score, the implicit assumption is presumably made that the four variables "measure the same thing"; although each variable by itself may only give a limited view of the respondents state of mind, the composite score may give a more valid and reliable picture. The investigation of whether these variables indeed measure the same thing and if so how well they are able to do that can be answered within the realm of factor analysis. It is not our intention to give a general description of factor analysis here; instead we will describe some specific issues related to the factor analysis of dichotomous and ordered categorical variables as are commonly found in rating scales of the above type.

Theory of factor analysis for categorical variables

Factor analysis of categorical variables may be conveniently linked to the modeling used above in the section entitled Regression analysis with categorical response. For the *j*th observed response variable we may assume an underlying latent response variable y^* as in equation 2

$$y^*_j = \lambda_j \eta + \epsilon_j \qquad (12)$$

Here, η is the vector of factor variables, and λ_j is the vector of regression slopes or factor loadings; and ϵ_j denotes the residual or measurement error in the *j*th variable's measurement of the factors. The relationship between the *y*s and the y^*s is still the threshold formulation of equation 3 or, with more than two response categories, of the type given in equation 6. The interesting feature here is that we are considering a regression on exogenous (independent) variables that are not observed but latent. The reason that such a regression analysis can be carried out without knowing the η values beforehand is that there are multiple response variables y; information on the regression slopes (the loadings) is obtained from the covariation of these y variables.

With a set of p y variables, there will be p equations of the type of equation 12. The factor analysis model specifies that a certain number of factors are sufficient to account for all covariation among the response variables. This means that the ϵ variables for different y variables are assumed to be uncorrelated. If this cannot describe the data well, a larger number of factors may be tried.

As with categorical regression, one may also formulate the model without latent response variables, considering a conditional probability curve for the regression of each y on the η. In this case, the assumption of uncorrelated residuals is replaced by the assumption of so-called conditional independence, i.e., given the η, the y variables are independent. The latter formulation is customary in so-called item response theory or latent trait theory used in the analysis of, e.g., achievement test items, scored right or wrong (see, e.g., (15)).

Factor analysis of categorical variables is a rather new field. Maximum likelihood estimation has been developed only in the last decade by Bock and Aitkin (16), building on earlier work by Bock and Lieberman (17). This is because of the rather complicated computations that arise from the categorical nature of the variables. Although Bock and Aitkin consider the conditional probability formulation of the model, we will again concentrate on the latent-response variable formulation as utilized in the works of Christoffersson (18), Muthén (19), Muthén and Christoffersson (20), and Muthén (21).

In the approach of Muthén (13), limited information from pairs of variables is used in order to first estimate the correlations among the latent-response variables. Empirical evidence has shown that the use of limited information relative to the full information approach of Bock and Aitkin (16) involves a very small loss of information; estimates and their standard errors are very close. The latent-response variables are assumed to be multivariate normal, in line with the assumption of normal y^* given x in the regression formulation of equation 2. With dichotomous response, this means that the estimated correlations are so-called tetrachoric correlations; and with ordered categorical response, polychoric correlations are obtained (for an overview of the treatment of such correlations, see, e.g., Muthén (14), and references therein).

One may note that the appropriate choice of correlation coefficients for categorical data is a much debated issue. In terms of dichotomous variables, there has been a long debate between advocates of ordinary Pearson correlations, so-called phi coefficients, and proponents of tetrachoric correlations. From a theoretical point of view, the phi coefficients can be said to be deficient because of the fact that their size is dependent upon the univariate distribution of the variables. The maximum obtainable correlation is not always 1.0. In factor analysis this can give rise to a distortion of the underlying correlational structure so that more than the correct number of factors are needed to explain the phi coefficients well (the so-called "difficulty factor" issue). Similar problems arise with more than two ordered categories, particularly when the univariate distributions exhibit skewness. Such problems have been studied by Olsson (21, 22) and Muthén and Kaplan (23).

The correlational distortions referred to above may be illustrated by considering bivariate normal latent-response variables with a correlation of 0.50 that are categorized into a new pair of variables. The correlation between the latter variables can differ quite considerably as illustrated in table 2. Here, D19 stands for dichotomization into a variable with univariate proportions 10%, 90%, and 3SY stands for a trichotomous variable symmetrically split, RE stands for a rectangular distribution, NS stands for a negatively skewed distribution, PS stands for a positively skewed distribution, and CON stands for a variable that has not been categorized. For instance, the Pearson correlation for a positively and a negatively skewed trichotomous variable is found to be 0.34 (decimals are omitted). Hence, the true correlation is substantially underestimated. More importantly, the degree of underestimation varies depending on univariate characteristics, so that if the underlying variables' correlations follow, for example, a single-factor model, the observed variables' Pearson correlations may not obey such a structure. If, however, the number of scale steps increase, or more importantly the distributions are relatively symmetric, the attenuation and therefore the correlational distortion may be negligible. For instance, five category rating scales ("Likert scales") may well be treated as "continuous," interval-scaled variables if not strongly skewed, so that the use of Pearson correlations would be acceptable.

The use of underlying, or "latent," correlations inferred from observed variables is of course not without its dangers. Underlying normality of the y^* is assumed in the approach of Muthén (13) and this may not always be suitable. In the Bock and Aitkin (16) conditional probability curve formulation, this corresponds to the assumption of normal distribution function curves for each y given the factors, as was also assumed in the regression modeling, but a further assumption of normal factors is also needed. Even though this would often seem to be a reasonable assumption, testing the appropriateness of this assumption before using the correlational estimates would seem to be valuable. Recently, a technique for carrying out such a test has been proposed by Muthén and Hofacker (24).

Another drawback of using latent correlations is the fact that their sampling variability is considerably higher than for ordinary Pearson correlations. Brown and Benedetti (25) studied the sampling behavior of the tetrachoric correlation and found that, in the two-by-two tables, expected frequencies of at least 5 were needed for good estimates and at least 10 for good estimates of the standard errors of the estimates. With very skewed univariate distributions, this means that large sample sizes are needed for good estimation. Usually, samples of say 500 are sufficient, but sometimes samples of at least double that are desirable.

Returning to the approach of Muthén (13), the analysis of the estimated "latent" correlations is carried out by generalized least squares. Estimates of model parameters, here factor loadings, factor variances and covariances, and residual variances are obtained by minimizing the weighted residuals between the model correlations and the sample correlations,

$$F = (\mathbf{s} - \sigma)'\mathbf{W}^{-1}(\mathbf{s} - \sigma) \tag{13}$$

Here, the vector s contains the sample correlations, and the vector σ contains the corresponding population quantities, which are functions of the parameters. The matrix **W** is a so-called weight matrix, chosen as an estimate of the large-sample covariance matrix of the estimated correlations. The function value F, calculated for the final parameter estimates, provides a large-sample chi-square test of model fit.

The chi-square test may be used as in modern factor analysis of ordinary, interval-scaled variables. In the usual, so-called exploratory factor analysis (see e.g. Lawley and Maxwell (26)), the null hypothesis is that a certain number of factors account for the interrelationships among the observed variables. The counterhypothesis is that of an unrestricted correlation matrix. Hence, we would usually not be satisfied with a rejection of the null hypothesis, i.e., we would like to observe small chi-square values relative to the number of degrees of freedom (the number of correlations minus the number of parameters estimated). In so-called confirmatory factor analysis (see, e.g., Jöreskog (27, 28)), one makes the further hypothesis of a certain structure on the loadings and/or the factor covariance matrix. For instance, a priori substantive knowledge might postulate that certain variables do not measure certain factors, so that the corresponding loadings would be zero.

It may be noted that the chi-square test provided by Muthén's approach builds on a weight matrix specifically developed for dichotomous and ordered categorical variables. For instance, although generalized least squares and maximum likelihood estimation using standard techniques for interval-scaled variables could be carried out on these sample latent correlations, this would not yield the correct result in terms of chi-square testing. Usually this leads to overestimated chi-squares, again due to the fact that the latent correlations are more variable than ordinary Pearson correlations assumed by these standard techniques. This should be kept in mind when using software such as LISREL (28), because latent correlations are provided there, but not the correct weight matrix of equation 13.

The methods developed in Muthén (13) and related articles are presently being programmed in a package similar to LISREL. The software is called LISCOMP (Analysis of Linear Structural Equations with a Comprehensive Measurement Model) and includes regression, factor analysis, and structural equation modeling with any mixture of dichotomous, ordered categorical, and continuous normal or nonnormal observed variables. All analyses of this chapter were carried out by a research version of LISCOMP. The program uses a model

Table 2. Pearson correlations for true correlations = 0.50

	D19	D28	D37	D46	D55	D64	D73	D82	D91	3SY	3RE	3NS	3PS	4SY	4RE	4NS	4PS	5SY	5RE	5NS	5PS	CON
D19	25																					
D28	26	30																				
D37	26	30	32																			
D46	24	30	32	33																		
D55	23	28	31	33	33																	
D64	20	26	30	32	33	33																
D73	18	23	27	30	31	32	32															
D82	15	20	23	26	28	30	30	30														
D91	10	15	18	20	22	24	26	26	25													
3SY	26	32	35	36	37	36	35	32	26	41												
3RE	25	31	35	36	37	36	35	31	25	41	41											
3NS	29	33	35	36	35	33	30	26	20	39	39	39										
3PS	20	26	30	33	35	36	35	33	29	39	39	34	39									
4SY	27	33	36	38	38	38	36	33	27	43	43	40	40	44								
4RE	26	33	36	38	38	38	36	33	26	42	42	40	40	44	44							
4NS	30	35	36	36	35	33	30	27	20	40	39	40	34	41	41	41						
4PS	20	27	31	34	35	36	36	35	30	40	39	34	40	41	41	35	41					
5SY	28	34	37	38	39	38	37	34	28	44	43	41	41	45	45	42	42	46				
5RE	27	33	37	38	39	38	37	33	27	43	43	41	41	45	45	41	41	46	45			
5NS	31	35	36	36	35	33	30	26	20	40	39	40	34	41	40	42	34	42	41	42		
5PS	20	26	30	33	35	36	36	35	31	40	39	34	40	41	41	34	42	42	41	34	42	
CON	29	35	38	39	40	39	38	35	29	45	45	42	42	47	46	43	43	48	47	44	44	50
	D19	D28	D37	D46	D55	D64	D73	D82	D91	3SY	3RE	3NS	3PS	4SY	4RE	4NS	4PS	5SY	5RE	5NS	5PS	CON

specification and keyword input system similar to LISREL's for ease of transition between the programs.

Factor analyses of psychological well-being and exercise data

Analyses of HAY variables

Returning to the four "how are you?" variables, we note that we are dealing with four-category ordered categorical variables. In an initial analysis, we may score these categories 0, 1, 2, 3 and treat them as continuous, interval scaled variables. A maximum likelihood factor analysis may be carried out in order to obtain a chi-square measure of fit to the assumed single-factor model. In the female and male samples used earlier, the chi-square values obtained were 87.17 and 66.64, with two degrees of freedom. Hence we must reject the notion of a single factor explaining the relationships among these variables. Although more than one factor may underlie the variables, this cannot be tested because with two factors the model is already "underidentified," i.e., not testable because of the lack of a positive number of degrees of freedom.

Scrutinizing the univariate distributions of these variables, one finds that three out of the four have rather skewed distributions. For "cheerful" a very large percentage of respondents in this sample give the answer "very often" (females 62%, males 53%), and this is also the case for "loved" (females 71%, males 64%), whereas for "lonely" the category "never" obtained a large number of responses (females 38%, males 47%). In all these three cases, the ensuing distribution appears "censored," i.e., there is an unduly high "piling up" of cases at these endpoints. This is a situation where categorical variable methodology is potentially very useful.

As opposed to the ordinary variable methodology just applied, with the categorical variable methodology, the respondents in the end categories need not be assumed to have the same response values. The methodology explicitly recognizes that the people in the end category have a latent-response variable value that exceeds (or falls below) a certain threshold, but the latent-response variable values may differ. Only the probability of exceeding (or falling below) this threshold is modeled, and the censoring problem vanishes. Alternative methodology that is more directly developed for censored, continuous variables has recently been proposed in a factor analysis context by Muthén (29).

We may also remind ourselves of the cases studied in table 2, where it was recognized that skewed four-category variables exhibited an attenuation of underlying correlations. Using polychoric correlations instead of ordinary Pearson correlations would seem to be more appropriate.

A categorical variable factor analysis was nevertheless deemed unnecessary in this case due to the very high chi-square values observed with continuous variable methodology. From practical experience in Muthén and Kaplan (23), it was judged that, although chi-square is expected to be brought down considerably in situations like this one when switching to categorical variable methodology, the necessary drop from 87 and 66 to less than 6 (i.e., falling below the 5% critical value with two degrees of freedom) would not occur. The one-factor model would most likely still be rejected.

A judgment of this kind may be of good practical use in situations where a large number of variables is to be analyzed. Presently, the new categorical variable methodology is considerably more demanding from a computational point of view than ordinary methods, particulary when exceeding, say, 20 response variables. As is the case in univariate regression with dichotomous response, where ordinary regression methods are frequently used for exploration purposes instead of the more expensive logit regression, it is useful to have a feeling for the situations where a large difference is not to be expected.

Further scrutiny of the interrelations among the variables show for both sexes that the lack of fit may be the result of a direct relationship between cheerful and loved, such that the large group of respondents answering very often to cheerful also tend to answer very often to loved. This strong association can not be explained by the two variables being related to the same single factor but goes beyond that and is indeed because of a kind of response consistency pertaining to the category very often. This issue will not be further analyzed for this set of variables, but similar problems will be dealt with below in the context of some other variable sets.

Analysis of physical activity variables

In the Belloc and Breslow (30) analysis of the Alameda health data, physical activities were related to health status. Similar types of analyses are possible with data obtained through the NSPHPC (1). Responses to the questions "Here is a list of active things that people do in their free time. How often do you do any of these things? Active sports? Swimming or taking long walks? Working in the garden? Doing physical exercise?" were used to form a variable indicating level of physical activity. Together with certain other good health practices, this score is often used to form a composite score.

In the data we are considering, similar variables exist. We will consider responses to the questions "Please tell me how often you participate in these activities. First, how often do you: Go swimming in the summer? Take long walks? Work on a physically active hobby such as dancing or gardening? Go jogging or running?" In addition, we will use responses to the related questions "How often do you: Ride a bicycle? Do calisthenics or physical exercise? Participate in any other active sport I haven't already mentioned?" The response categories for these seven variables are in all cases: never, rarely, sometimes, often, i.e., four ordered categories.

If these variables are to be used to form a physical activity index, a factor analysis may be very useful. We would like to know if the variables can be said to measure a single dimension and how well each of the variables measure this dimension. If certain variables do not measure only one dimension, these variables should be singled out. The effect of using a score based on variables that measure different factors is that, when the different factors relate differently to other variables such as health status, the relationship between the score and the health status variable will be confounded.

We will apply categorical variable methodology to the analysis of these seven four-category variables, using the LISCOMP approach of Muthén (13). A sample of 1,490 females with complete data on the seven variables and a health status variable was used and also a sample of 988 males. As a first step, we compute the polychoric correlations among the variables for both females and males. For comparison, we also give the regular Pearson correlations (table 3). We note the attenuation effect on the Pearson correlations described earlier in connection with table 2. However, the overall structure of the correlations does not differ that much across the two methods of computation. This may be attributable to the fact that most of the seven variables are rather symmetrically distributed.

An important piece of information is obtained as a side product from the computation of the polychoric correlations. As described previously, the estimation "latent correlations" such as polychorics builds on the assumption of bivariate normality for underlying latent-response variables for each pair of variables. As we have mentioned, this is a model assumption that can and should be tested before using the correlations in a further analysis step, such as factor analysis.

The underlying bivariate normality assumption involves a model with a set of thresholds parameters for each of the two variables involved and a correlation parameter. With four categories, there are three thresholds for each variable, such that there are seven parameters in the bivariate normal model for each pair of variables. These are the parameters of the null hypothesis model. In each four-by-four table there are 16 cells, which implies that the alternative, unrestricted model has 15 parameters. We may then estimate the frequencies under the model and compare them with the observed frequencies and carry out a Pearson chi-square test of fit to the model with 15 − 7 = 8 degrees of freedom.

This procedure was utilized in the analysis of the 21 pairs of variables for females and for males. The overall picture of these tests is that the normality hypothesis is strongly rejected. This is an interesting conclusion given the earlier observation that the correlational structure does not change much when moving from Pearson correlations to polychorics. However, further scrutiny of the reason for rejection gives some useful insights into the response behavior for these variables.

Table 3. Pearson and polychoric correlations for 7 physical activity variables

Activity variable	SWIM	WALK	HOBBY	JOG	BIKE	EXERCISE
			Pearson correlation			
Females						
Walk	0.149					
Hobby	0.174	0.229				
Jog	0.230	0.244	0.182			
Bike	0.280	0.196	0.147	0.193		
Exercise	0.256	0.209	0.194	0.337	0.173	
Sports	0.301	0.217	0.221	0.315	0.296	0.290
			Polychoric correlation			
Walk	0.174					
Hobby	0.200	0.278				
Jog	0.310	0.325	0.253			
Bike	0.354	0.249	0.193	0.289		
Exercise	0.302	0.255	0.237	0.461	0.220	
Sports	0.370	0.265	0.272	0.424	0.378	0.359
Males						
			Pearson correlation			
Walk	0.115					
Hobby	0.141	0.188				
Jog	0.292	0.182	0.102			
Bike	0.236	0.173	0.102	0.169		
Exercise	0.248	0.104	0.140	0.429	0.173	
Sports	0.355	0.116	0.207	0.312	0.179	0.306
			Polychoric correlation			
Walk	0.140					
Hobby	0.174	0.233				
Jog	0.374	0.233	0.145			
Bike	0.295	0.210	0.129	0.252		
Exercise	0.291	0.121	0.174	0.539	0.224	
Sports	0.423	0.134	0.255	0.409	0.240	0.376

In tables 4 and 5, the polychoric analyses are given for the pairs of variables JOG, BIKE and JOG, EXERCISE for females and males. In each table, there are three sets of cell entries—observed frequencies, expected frequencies (i.e. frequencies predicted by the bivariate normal polychoric model), and chi-square elements. When summed over the cells, the latter give the overall Pearson chi-square; hence, these elements give the contribution from each cell to misfit.

Consider table 4 describing JOG, BIKE for females. Chi-square testing clearly rejects the polychoric model, but the major contributions, as reflected in the chi-square elements, seem to come from the four "northwest" cells involving the response categories "never" and "rarely." There is a particularly large contribution from the 2–2 bivariate cell rarely-rarely, 33.2. We note that for this cell the expected frequency under the polychoric model is 44.5, and the observed frequency is almost double that, 83. The corresponding JOG, BIKE frequencies for males, show a very similar picture. The expected frequency for rarely-rarely is 54.7, and the observed frequency is 86.

The same phenomenon is observed also in table 5, describing JOG, EXERCISE. In fact, this observation holds true for a majority of the 21 pairs, although for some pairs the misfitting bivariate cell was sometimes-sometimes. At the same time, the magnitude of these cell frequencies are not that much different from neighboring cells.

One possible interpretation of this phenomenon is as follows. A certain proportion of the respondents shows relatively little commitment to answering in a precise way, instead responding in a relatively neutral way. Although underlying normality of the latent-response propensities does hold approximately, the tendency to give a consistent, nonprecise answer makes for

Table 4. Frequencies for males and females for JOG, BIKE

	Female			
Observed table				
	1	2	3	4
1	577.00	111.00	168.00	90.00
2	68.00	83.00	50.00	38.00
3	57.00	35.00	54.00	33.00
4	45.00	31.00	32.00	18.00
Expected table				
	1	2	3	4
1	539.66	159.85	167.04	79.46
2	103.42	44.54	56.09	34.96
3	67.04	33.31	45.79	32.86
4	36.88	22.31	35.08	31.73
Chi-square element (approximate chi-square with 8 df = 79.629)				
	1	2	3	4
1	2.58	14.93	0.01	1.40
2	12.13	33.22	0.66	0.26
3	1.50	0.09	1.47	0.00
4	1.79	3.39	0.27	5.94
	Male			
Observed table				
	1	2	3	4
1	301.00	95.00	74.00	41.00
2	47.00	86.00	39.00	26.00
3	54.00	41.00	27.00	14.00
4	51.00	39.00	30.00	23.00
Expected table				
	1	2	3	4
1	274.04	128.81	72.43	35.72
2	84.38	54.70	36.68	22.24
3	51.13	38.09	27.87	18.91
4	43.45	39.40	33.02	27.13
Chi-square element (approximate chi-square with 8 df = 51.500)				
	1	2	3	4
1	2.65	8.88	0.03	0.78
2	16.56	17.91	0.15	0.64
3	0.16	0.22	0.03	1.28
4	1.31	0.00	0.28	0.63

a distortion of this picture. In this case, the misfit of the polychoric model should be ignored and the estimated frequencies and ensuing correlation taken as a "smoothing" or correction/improvement of distorted data—i.e., the lower predicted frequency for rarely-rarely is the truer picture of reality. One could even consider calculating statistics, e.g., Pearson correlations, from these estimated frequencies.

An alternative interpretation is that the underlying latent-response propensities are not normal, but that there is a particular peak in the bivariate distribution in the region which gives rise to the rarely-rarely response. In this case, the Pearson correlation should be used as the more appropriate measure of correlation, although some small amount of attenuation is to be expected because of categorization.

The interesting practical implication is, however, not so much a choice of Pearson or polychorics but rather that the polychoric analyses point out further considerations that should be made. Given the high degree of consistent misfit for rarely-rarely, it would be of interest to scrutinize the response patterns over all seven variables for the samples at hand. Perhaps one can find certain "outlier" type response patterns, corresponding to individuals that give seemingly careless responses, such as rarely throughout. Such an analysis was not carried out here.

In the further analyses of these data, it was instead decided to dichotomize the four-category variables into never-rarely versus sometimes-often. To some extent,this may reduce the potential misclassification errors, at least if one is willing to assume that most of the overrepresentation of the rarely response comes from people who rightfully belong in the never category. If there is substantive reason for believing instead that rarely respondents are more likely to come from the sometimes

Table 5. Frequencies for males and females for JOG, EXERCISE

Female

Observed table

	1	2	3	4
1	388.00	128.00	245.00	185.00
2	22.00	63.00	82.00	72.00
3	18.00	25.00	71.00	65.00
4	9.00	13.00	23.00	81.00

Expected table

	1	2	3	4
1	363.79	163.15	255.06	163.99
2	43.49	34.80	77.49	83.22
3	21.83	21.24	55.82	80.11
4	7.88	9.80	32.64	75.68

Chi-square element
(approximate chi-square with 8 df = 60.255)

	1	2	3	4
1	1.61	7.57	0.40	2.69
2	10.62	22.85	0.26	1.51
3	0.67	0.66	4.13	2.85
4	0.16	1.04	2.85	0.37

Male

Observed table

	1	2	3	4
1	261.00	81.00	92.0	77.00
2	27.00	74.00	48.00	49.00
3	20.00	28.00	45.00	43.00
4	10.00	12.00	36.00	85.00

Expected table

	1	2	3	4
1	243.19	111.42	96.02	60.36
2	45.56	42.79	53.91	55.74
3	19.75	24.47	37.77	54.01
4	9.51	16.32	33.29	83.88

Chi-square element
(approximate chi-square with 8 df = 51.698)

	1	2	3	4
1	1.30	8.31	0.17	4.58
2	7.56	22.77	0.65	0.82
3	0.00	0.51	1.38	2.25
4	0.03	1.14	0.22	0.01

category, it would have been better to collapse never-rarely-sometimes into one category versus the often category.

At the same time, it was decided to try to relate the physical activity responses to health status. This latter variable is also categorical, measuring limitations of activity level with the categories: severely limited, moderately limited, symptomatic, healthy-low energy, and healthy-high energy. This variable was also dichotomized, with the first three categories constituting the low health status category. The samples were then split into low females ($N = 538$), high females ($N = 952$), low males ($N = 238$), and high males ($N = 750$).

The single factor model was first tested for each of the four groups of respondents. The degrees of freedom for testing this model is 14. For low females, the resulting chi-square value was 12.11, which means that the single-factor model cannot be rejected. The 5% critical value is 23.69. For the group high females, however, the chi-square of 56.63 results in a very strong indication of misfit for the single-factor model. For males, the chi-squares were 35.31 and 53.81, respectively, for the low and high groups, also indicating misfitting single-factor models. We note a considerable discrepancy in model fit across these subgroups.

Further analyses were carried out for the high groups in order to achieve better fitting models. An exploratory, two-factor analysis for females resulted in a "Heywood" case. This means that a residual variance tends to go toward a large negative estimate without reaching a proper solution. This situation is unacceptable in that the solution, if it could be found, falls outside the region of permissible parameter values; the negative variances make the two-factor model untenable. For high males, the exploratory two-factor analysis also resulted in a Heywood case. Such situations are common when there are not a sufficient number of variables to "define" a second or third factor; although a single-factor solution is not acceptable, the misfit is the result of additional correlations among the variables that cannot be accounted for in any factor analysis model.

As a final attempt at finding an appropriate model, the single-factor models were relaxed somewhat by allowing certain residuals to be correlated, i.e., deviating from the standard exploratory model. This may sometimes be appropriate when a second factor is not well defined by the variables. The LISCOMP program provides (mis-)fit indexes to aid in such a search for better fit. Scrutinizing the results of the single-factor models for high females and males revealed that one might try to include residual correlations between WALK and HOBBY and between JOG and BIKE. In both groups the single-factor model predicted too low a correlation between WALK and HOBBY and too high a correlation between JOG and BIKE. There seems to be no clear-cut interpretation for these correlations, and one may speculate that they are the result of the potential response consistency effects discussed above.

Including the residual correlations in the single-factor models did indeed bring down the chi-square values towards more acceptable regions. For females, the value was now 26.70; and for males, the value was 26.89. With 12 degrees of freedom, these values are, however, still unacceptable on the 5% level (critical value = 21.03).

We conclude that the single-factor model can only be said to fit well for the low health status female group. This has important practical implications when comparing some composite physical activity score across sex and health status groups, because we cannot be sure that the composite measures the same underlying phenomenon. The physical activity variables may indicate one thing in one group and another thing in another group. Indeed, only for the low female group is there a reason to consider a single composite physical activity score. This should be a serious concern in judging the validity of the physical activity measurements.

The factor loading estimates for the low female group are given in table 6. It is seen that all loadings are significant and have relatively small standard errors. The JOG variable seems to be the most reliable measure of the "physical activity factor," whereas EXERCISE is the least reliable measure.

If there had been a good fit to the single-factor model in all four groups considered, a further multiple-group analysis would be of interest. To firmly establish that the set of variables measure the same thing, we would not only want to obtain a good fit to the single-factor model in each group but also have equality of loadings for each variable across groups. Testing of such hypotheses can be accomplished in a simultaneous analysis of several groups, as described for the categorical case by Muthén and Christoffersson (20) and Muthén (13).

If the measurements operate in the same way in all subgroups, a chi-square test of group-invariant loadings should not lead to rejection. Certain variables may, however, contribute to high chi-square values, indicating that these variables are not suitable measures in group comparisons. If the set of variables is found to exhibit group-invariant measurement parameters, however, a powerful further analysis is possible. The groups may then be compared with respect to the estimated values of each group's factor mean and factor variance. This is a more appropriate measure of group differences than a comparison of the levels of an observed variable composite, because the

Table 6. Factor loading estimates for 7 physical activity variables low health status, female group

Activity variable	Loading	Standard error
Swim	0.569	0.056
Walk	0.642	0.053
Hobby	0.579	0.056
Jog	0.750	0.058
Bike	0.541	0.063
Exercise	0.485	0.058
Sports	0.657	0.053

$N = 538$

latter contains measurement errors. For applications of such methodology, we refer to Muthén (31).

The regression-type modeling discussed earlier may be combined with the factor analysis or measurement modeling just discussed. The factors may be related not only to categorical grouping variables such as sex but also to other continuous variables, be they other factors or other observed variables. Such a combination of regression modeling and factor analysis modeling is the topic of structural equation modeling. Categorical and continuous variables may be dealt with in the framework of the most general version of the model in Muthén (13). For applications of this type of methodology, see Muthén (14) and Muthén and Speckart (32).

Implications for measurement and analysis of fitness and exercise data

The above analyses should make clear that useful methodology does exist for the proper analysis of categorical or qualitative data. Although the categorical nature of the variables can be ignored, this usually tends to provide a less correct and, more importantly, a less informative analysis. The degree of detail that is desired will only be provided when the response process is realistically modeled by methods specifically designed for the type of variables at hand.

This was exemplified in the regression analyses with ordered categorical alcohol consumption responses as dependent variables. Even though response categories are ordered, there was a strong indication that equidistance of the categories did not hold. Furthermore, because the probability of consuming alcohol may be positively related to income, there was no indication that the amount consumed was related at all to income. Ordinary regression analysis would most likely overlook both of these properties of the data.

Other examples were found in the factor analyses with categorical data. In the analyses of physical activity data, detailed analysis of the responses showed interesting response consistency effects. With appropriate modeling for categorical data, it was found that a simple measurement model was in several instances not a suitable description of the data at hand.

The availability of new categorical variable methodology has important implications for design of measurement instruments. Clearly, there is now a greater freedom of choice when it comes to decisions regarding number and type of response alternatives. Two response categories can be as easily accommodated as more expansive five-category measurements. Equidistance of the scale steps need not be an a priori, untested assumption, but one which can be tested in the data. Even if there is distinct "piling up" of cases at endpoints of the scales, the new methodology avoids modeling errors.

The measurement design efforts can now be concentrated around what is most suitable for the substantive problem at hand and around judgments regarding the degree of detail in the information that can realistically be extracted from the respondents.

It is still true, of course, that the more fine-tuned, still reliable information that can be obtained from the respondents the better. In this connection, it should be kept in mind that, with categorical measurements, there is inevitably less information available, giving rise to a larger degree of imprecision in model parameter estimates. The point is, however, that there may be a definite advantage in avoiding unreliable responses by appropriate choice of measurement scales, despite such a loss in precision; the loss of precision may not be a real one.

It would be of interest in future methodological research to study the issue of such presumed tradeoffs between reliability and sampling variability. In large-scale health surveys, however, low precision of estimates would usually not be a serious concern. Even if subgroups of respondents are studied, it is usually not a probelm to obtain sufficiently large sample sizes. In such subgroupings, one may perhaps use the very rough rule of attempting sample sizes of at least 500 for most of the categorical variable analyses illustrated.

The factor analysis of categorical fitness and exercise data requires more than a good choice of measurement scale. This was clearly experienced in the limited set of analyses carried out in this chapter and in other health data factor analyses not reported. Effective use of factor analysis can only be made if the choice of variables to be analyzed has been made carefully.

In the physical activity variables that we studied (and these may indeed not be a set of variables for which a factor analysis was contemplated), there were potential modeling problems in that several variables would seem to have direct relationships to each other. For instance, there were questions both regarding how often you go jogging and how often you do physical exercise—a respondent may naturally include jogging in physical exercise and this fact alone will complicate the factor analysis.

If a factor analysis is deemed of importance, several variables hypothesized to measure different aspects of the factors should be measured. If multidimensionality is expected, each dimension should have a representative set of measurements. Pilot studies should be carried out using a large set of variables to aid in the choice of a suitable subset to be used in the large scale studies. In conclusion, nicely behaving factor solutions cannot be expected unless they were planned for by design of measurement.

References

1. Golden, P.M., 1982. *Public use data tape documentation: Wave I and wave II of the National Survey of Personal Health Practice and Consequences 1979–1980.* U.S. Department of Health and Human Services. Public Health Service. Office of Health Research, Statistics, and Technology. National Center for Health Statistics. Hyattsville, MD.

2. Bliss, C.I., 1935. The calculation of the dosage mortality curve (appendix by R.A. Fischer). *Annals of Applied Biology, 22,* 134–167.
3. Finney, D., 1971. *Probit Analysis.* Cambridge: Cambridge University Press.
4. Cox, D.R., 1970. *The analysis of binary data.* London: Methuen.
5. Tobin, J., 1958. Estimation of relationships for limited dependent variables. *Econometrica, 26,* 24–36.
6. Amemiya, T., 1984. Tobit models: A Survey. *Journal of Econometrics, 24,* 3–61.
7. Aldrich, J.H. and F.D. Nelson, 1984. ***Linear probability, logit, and probit models. Quantitative applications in the social sciences,*** No. 45. Sage Publications, Inc., Beverly Hills.
8. McKelvey, R.D. and W. Zavoina, 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology, 4,* 103–120.
9. Amemiya, T., 1981. Qualitative response models: A survey. *Journal of Economic Literature, 19,* 1483–1536.
10. Maddala, G.S., 1983. ***Limited-dependent and qualitative variables in econometrics.*** Cambridge: Cambridge University Press.
11. Anderson, J.A., 1984. Regression and ordered categorical variables. *Journal of the Royal Statistical Society, B,* 46, 1–30.
12. McCullagh, P., 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society, B42,* 109–142.
13. Muthen, B., 1984. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132.
14. Muthen, B., 1983. Latent variable structural equation modeling with categorical data. *Journal of Econometrics,* 1983, *22,* 43–65.
15. Lord, F.M., 1980. ***Applications of item response theory to practical testing problems.*** Hillsdale, NJ.: Lawrence Erlbaum Assoc.
16. Bock, R.D. and M. Aitkin, 1981. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443–459.
17. Bock, R.D. and M. Lieberman, 1970. Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35,* 179–197.
18. Christoffersson, A., 1975. Factor analysis of dichotomized variables. *Psychometrika, 40,* 5–32.
19. Muthen, B., 1978. Contributions to factor analysis of dichotomous variables. *Psychometrika, 43,* 551–560.
20. Muthen, B. and A. Christoffersson, 1981. Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46,* 485–500.
21. Olsson, U., 1979. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44,* 443–460.
22. Olsson, U., 1979. On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research, 14,* 485–500.
23. Muthen, B. and D. Kaplan, 1984. A comparison of some methodologies for the factor analysis of non-normal Likert variables. Accepted for publication in *British Journal of Mathematical and Statistical Psychology,* 1985.
24. Muthen B. and C. Hofacker, 1984. Testing the assumptions underlying tetrachoric correlations. Presented at the Psychometric Society meeting, Santa Barbara, CA. Submitted for publication.
25. Brown, M.B. and J.K. Benedetti, 1977. On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika, 42,* 347–355.
26. Lawley, D.N. and A.E. Maxwell, 1971. ***Factor analysis as a statistical method.*** London: Butterworth.
27. Joreskog, K.G., 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.
28. Joreskog, K.G., 1977. Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.), *Applications of statistics.* Amsterdam: North-Holland.
29. Muthen, B., 1985 (b). Tobit factor analysis. Presented at the Fourth European Meeting of the Psychometric Society, Cambridge, England, July 2–5.
30. Belloc, N.B. and L. Breslow, 1972. Relationship of physical health status and health practices. *Preventive Medicine, 1,* 409–421.
31. Muthen, B., 1981. Factor analysis of dichotomous variables: American attitudes toward abortion. In D.J. Jackson and E.F. Borgatta (Eds.), *Factor analysis and measurement in sociological research: A multidimentional perspective.* London: Sage Publications.
32. Muthen, B. and G. Speckart, 1985. Latent variable probit ANCOVA: Treatment effects in the California civil addict programme. University of California, Los Angeles. Accepted for publication in *British Journal of Mathematical and Statistical Psychology.*

Bibliography

Aigner, D.J. and A.S. Goldberger, 1977. ***Latent variables in socioeconomic models.*** Amsterdam: North-Holland.

Aitchison, J. and S.D. Silvey, 1957. The generalization of probit analysis to the case of multiple responses. *Biometrika, 44,* 131–140.

Amemiya, T., 1973. Regression analysis when the dependent variable is truncated normal. *Econometrica, 41,* 997–1016.

Amemiya, T., 1978. The estimation of a simultaneous equation generalized probit model. *Econometrica, 46,* 1193–1206.

Andersen, E.D., 1980. ***Discrete statistical models with social science applications.*** Amsterdam: North-Holland.

Ashford, J.R. and R.R. Sowden, 1970. Multivariate probit analysis. *Biometrics, 26,* 535–546.

Bartholomew, D.J., 1980. ***Factor analysis for categorical data.*** *Journal of the Royal Statistical Society, B42,* 293–321.

Bentler, P.M., 1980. Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31,* 419–456.

Bentler, P.M. and D.G. Weeks, 1980. Linear structural equations with latent variables. *Psychometrika, 45,* 289–308.

Bentler, P.M., 1983a. Simultaneous equations as moment structure models: With an introduction to latent variable models. *Journal of Econometrics, 22,* 13–42.

Bentler, P.M., 1983b. Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika,* Vol. 48, *4,* 493–517.

Bielby, W.T. and R.M. Hauser, 1977. Structural equation models. *Annual Review of Sociology,* 3, 137–161.

Bishop, Y.M.M., S.E. Fienberg and P.W. Holland, 1975. ***Discrete multivariate analysis: Theory and practice.*** Cambridge, MA.: MIT Press.

Bock, R.D., 1975. ***Multivariate statistical methods in behavioral research.*** New York: McGraw-Hill.

Bock, R.C. and L.V. Jones, 1968. ***The measurement and prediction of judgement and choices.*** San Francisco, CA.: Holden-Day.

Boomsma, A., 1983. ***On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality.*** Ph.D. dissertation, University of Groningen, Groningen, The Netherlands.

Browne, M.W., 1974. Generalized least squares estimates in the analysis of covariance structures. *South African Statistical Journal, 8,* 1–24. Reprinted in: D.J. Aigner and A.S. Goldgerger, (Eds.), 1977, ***Latent variables in socio-economic models.*** Amsterdam: North Holland.

Browne, M.W., 1982. Covariance structures. In D.M. Hawkins (Ed), ***Topics in applied multivariate analysis.*** Cambridge, MA.: Cambridge University Press.

Browne, M.W., 1984. Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Carroll, J.B., 1961. The nature of the data, or how to choose a correlation coefficient. *Psychometrika, 26,* 347–372.

Chen, C.-F., 1981. The EM approach to the multiple indicators and multiple causes model via the estimation of the latent variable. *Journal of the American Statistical Association, 76,* 704–708.

Clark, C., 1961. The greatest of a finite set of random variables. *Operations Research,* 145–162.

de Leeuw, J., 1983. Models and methods for the analysis of correlation coefficients. *Journal of Econometrics, 22,* 113–137.

Dempster, A.P., N.M. Laird and D.B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B 39,* 1–38.

Divgi, D.R., 1979. Calculation of the tetrachoric correlation coefficient. *Psychometrika, 44,* 169–172.

Edwards, A.L. and L.L. Thurstone, 1952. An internal consistency check for the methods of successive intervals and the method of graded dichotomies. *Psychometrika, 17,* 169–180.

Ferguson, G.A., 1941. The factorial interpretation of test difficulty. *Psychometrika, 6,* 323–329.

Ferguson, T.A., 1958. A method for generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics, 29,* 1046–1062.

Fienberg, S.E., 1977. *The analysis of cross-classified categorical data.* Cambridge: The MIT Press.

Fuller, E.L. and Hemmerle, 1966. Robustness of the maximum likelihood estimation procedure in factor analysis. *Psychometrika, 31,* 255–266.

Gibbons, R.D., 1981. Full information factor analysis of dichotomous variables. Presented at the Psychometric Society meeting, Chapel Hill, North Carolina.

Gibbons, R.D. and R.D. Bock, 1982. A probit model for trends in correlated proportions. Presented at the Psychometric Society meeting, Montreal.

Goodman, L.A., 1984. *The analysis of cross-classified data having ordered categories.* Cambridge: Harvard University Press.

Gumbel, E.J., 1961. Bivariate logistic distributions. *Journal of the American Statistical Association, 56,* 335–349.

Gurland, J., I. Lee and P.A. Dahm, 1960. Polychotomous quantal response in biological assay. *Biometrics, 16,* 382–398.

Heckman, J., 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica, 46,* 931–959.

Huba, G.J. and P.M. Bentler, 1983. Test of drug use causal model using asymptotically distribution free methods. *Journal of Drug Education, 13*(1), 3–17.

Huba, G.J. and L.L. Harlow, 1983. Comparison of maximum likelihood generalized least squares, ordinary least squares, and asymptotically distribution free parameter estimates in drug abuse latent variable causal models. *Journal of Drug Education, 13*(4), 387–404.

Huba, G.J. and L.L. Harlow, 1985. Robust estimation for causal models: A comparison of methods in some developmental data sets. In D.L. Featherman and R.M. Lerner (Eds.), *Life-span development and behavior, Vol. 6.* New York: Academic Press, forthcoming.

Huba, G.J. and J.S. Tanaka, 1983. Confirmatory evidence for three daydreaming factors in the short imaginal processes inventory. *Imagination, Cognition, and Personality, 3*(2), 139–147.

Huba, G.J., J.A. Wingard, and P.M. Bentler, 1981. A comparison of two latent variable causal models for adolescent drug use. *Journal of Personality and Social Psychology, 40,* 180–193.

Ito, K., 1969. On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures. *Proceedings of the International Symposium on Multivariate Analysis,* Vol. 2, P.R. Krishnaiah (Ed), 87–120, New York: Academic Press.

Jaspen, N., 1946. Serial correlation. *Psychometrika, 11,* 23–30.

Joreskog, K.G., 1971. Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426.

Joreskog, K.G., 1973. A general method for estimating a linear structural equation system. In A.S. Goldberger and O.D. Duncan (Eds.), *Structural equation model in the social sciences.* New York: Seminar Press, 85–112.

Joreskog, K.G., 1978a. Structural analysis of covariance and correlation matrices. *Psychometrika, 43,* 443–477.

Joreskog, K.G. and A.S. Goldberger, 1972. Factor analysis by generalized least squares. *Psychometrika, 37,* 243–260.

Joreskog, K.G. and D. Sorbom, 1984. LISREL VI; Analysis of linear structural relationships by maximum likelihood and least squares methods. Scientific Software, Inc.

Kendall, M. and A. Stuart, 1977. *The advanced theory of statistics.* Volume 2. New York: Macmillan Publishing Company.

Lawley, D.N., 1943. On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh, 61,* 273–287.

Lawley, D.N., 1944. The factorial analysis of multiple item tests. *Proceedings of the Royal Society of Edinburgh, 62-A,* 74–82.

LeCam, L., 1956. On the asymptotic theory of estimation and testing hypotheses. *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability, 5*(1), 129–156.

LeCam, L., 1980. Contributions to the discussion of Berkson (1980). *Annals of Statistics, 8,* 473–478.

Lee, L.F., 1979. Health and wage: A simultaneous equation model with multiple discrete indicators. Discussion paper no. 79–127. University of Minnesota, Minneapolis, MN.

Lee, S.-Y. and K.-L. Tsui, 1982. Covariance structure analysis in several populations. *Psychometrika,* forthcoming.

Lord, F.M. and H. Novick, 1968. *Statistical theories of mental test scores.* Reading, MA.: Addison-Wesley.

Maddala, G.S. and L.F. Lee, 1976. Recursive models with qualitative endogenous variables. *Annals of Economic and Social Measurement, 5,* 525–545.

Mardia, K.V., 1970. Measures of mutivariate skewness and kurtosis with applications. *Biometrika, 57,* 519–530.

Mardia, K.V., 1971. The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrika, 58,* 105–121.

Mardia, K.V., 1974. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankya, Series B, 36,* 115–128.

Mardia, K.V., 1975. Assessment of multinormality and the robustness of Hotelling's T2 test. *Applied Statistics, 24,* 163–171.

Mardia, K.V. and P.J. Zemroch, 1975. Measures of multivariate skewness and kurtosis. *Applied Statistics, 24,* 262–265.

Mardia, K.V., J.T. Kent and J.M. Bibby, 1979. *Multivariate analysis.* London: Academic Press.

Martinson, E.O. and M.A. Hamdan, 1971. Maximum likelihood and some other asymptotically efficient estimators of correlation in two way contingency tables. *Journal of Statistical Computation and Simulation, 1,* 45–54.

McDonald, R.P., 1974. Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27,* 82–99.

Mooijaart, A., 1983. Two kinds of factor analysis for ordered categorical variables. *Multivariate Behavioral Research, 18,* 423–441.

Morimune, K., 1979. Comparisons of normal and logistic models in the bivariate dichotomous analysis. *Econometrica, 47,* 957–976.

Muirhead, R.J., 1982. *Aspects of multivariate statistical theory.* New York: John Wiley and Sons, Inc.

Muraki, E., 1984. Implementing full information factor analysis: Testfact program. Presented at the annual meeting of Psychometric Society.

Muthen, B., 1976. Structural equation models with dichotomous dependent variables: A sociological analysis problem formulated by O.D. Duncan, Research report no. 76–19. Department of Statistics, University of Uppsala, Uppsala, Sweden.

Muthen, B., 1979. A structural probit model with latent variables. *Journal of the American Statistical Association, 74,* 807–811.

Muthen, B., 1982. Some categorical response models with continuous latent variables. In K.G. Joreskog and H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction.* Amsterdam: North-Holland.

Muthen, B., 1985 (a). A method for studying the homogeneity of test items with respect to other relevant variables. *Journal of Educational Statistics,* forthcoming.

Neyman, J., 1949. Contribution to the theory of the x2 test. *Proceedings of the First Berkeley Symposium in Mathematical Statistics and Probability,* 239–273.

Olsson, U., F. Drasgow and N.J. Dorans, 1982. The polyserial correlation coefficient. *Psychometrika,* Vol. 47, *3,* 337–347.

Pearson, K., 1900. Mathematical contributions to the theory of evolution, VII: On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, A-195,* 1-147.

Pearson, K., 1904. Mathematical contributions to the theory of evolution, XIII: On the theory of contingency and its relation to association and normal correlation. Drapers Company Research Memoirs, Biometric Series, No. 1.

Pearson, K., 1913. On the measurement of the influence of 'broad categories' on correlation. *Biometrika, 14,* 127-156.

Press, S.J. and J.M. Tanur, 1985. Multiple modeling of propensity to adopt residential energy conservation retrofits. Technical Report No. 208, Department of Statistics, Stanford.

Rubin, D.B. and D.T. Thayer, 1982. EM Algorithms for ML factor analysis. *Psychometrika, 47,* 69-76.

Samejima, F., 1969. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph* No. 17.

Samejima, F., 1972. A general model for free-response data. *Psychometrika Monograph* No. 18.

Sorbom, D., 1974. A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229-239.

Sorbom, D., 1978. An alternative to the methodology for analysis of covariance. *Psychometrika, 43,* 381-396.

Sorbom, D., 1982. Structural equation models with structured means. In K.G. Joreskog and H. Wold (Eds), *Systems under indirect observation: Causality, structure, prediction.* Amsterdam: North-Holland.

Steiger, J.H. and A.R. Hakstian, 1982. The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and Statistical Psychology, 35,* 208-215.

Tallis, G.M., 1972. The maximum likelihood estimation of correlation from contingency tables. *Biometrics, 18,* 342-353.

Tanaka, J.S., 1984. *Some results on the estimation of covariance structure models.* Ph.D. dissertation, University of California, Los Angeles.

Chapter **29**

Latent Class Analysis

Allan L. McCutcheon
University of Delaware

One of the central problems of survey analysis is that many variables of interest cannot be directly observed. Frequently, however, imperfect indicators of the variable of interest are directly observable; thus the challenge confronting the researcher is to relate the observed indicators to the latent (unobserved) variable of central interest. If the observed variables are scored at the interval level of measurement, if these measures have underlying multivariate normal distributions, and if the latent factor itself is a continuous construct then factor analysis may be used to relate the observed variables to the latent variables. Factor analysis is not satisfactory, however, when the observed variables are qualitative or categorical; that is, when the variables are scored as either nominal or ordinal. In epidemiologic research, in which the identification of "case" is a central feature, nominal and ordinal data are common.

In 1950 Lazarsfeld (1950a, 1950b) published two papers that set forth a new approach for examining latent variables with categorically scored observable indicators. In these papers, Lazarsfeld outlined a set of methods commonly called *latent-structure analysis.* Latent-*class* analysis—the primary focus of this paper—is a recently developed method for analyzing relationships among categorical variables that derives from Lazarsfeld's early work. In the mid-1970's, Goodman (1974a, 1974b) presented a method for solving the parameter estimation problems that had made latent class analysis inaccessible to earlier researchers.

Goodman (1974a, 1974b) developed several latent-class models for examining the relationships among several observed variables presumed to measure one or more categorical latent variables. These latent-class models, along with more recently developed models, offer analysts a wide range of techniques for examining relationships among categorical observed variables. The latent-class analysis model allows the researcher to assume that the relationships among the observed variables are noncausal in nature; in this model latent-class analysis may be viewed as a qualitative data analog to factor analysis. These analyses may be either *exploratory*, in which no a priori expectations are imposed on the relationship between the latent variable and the observed variables, or *confirmatory*, in which hypotheses relating the latent variable to the observed variables can be tested directly. Recent developments in latent-class analysis also enable simultaneous estimation of latent-class models in two or more groups (Clogg and Goodman, 1984, 1985). Simultaneous latent-class analysis allows across group comparisons so that researchers can 1) compare the relationship between the latent variable and the observed variables within each of the groups, and 2) compare the latent variable across the groups.

A quite different application of the latent-class model involves scalogram analysis of categorically scored items, the answers to which are hypothesized as forming an ordered scale. Clogg and Sawyer (1981) have demonstrated that latent-class models may be used to test a wide variety of hypotheses in scalogram analysis; and Clogg and Goodman (1986) have shown that latent-class models can be used to perform simultaneous scalogram analysis in two or more groups. A third application of the latent-class model is in examining hypothesized causal relationships among latent categorical variables in a manner analogous to structural equation models (Clogg, 1981; Madden and Dillon, 1982; Bergan, 1983).

Latent-class analysis, which has been developed largely within the context of survey research, has found many applications in the social sciences. Increasingly, however, applications of latent-class models are being found in different fields of research; Young (1983; Young, Tanner, and Meltzer 1982), for instance, has applied latent-class models to the evaluation of psychiatric diagnostic criteria, and Rindskopf and Rindskopf (1986) have proposed the use of latent class analysis for medical diagnosis. This paper discusses contemporary latent-class analysis methods and some of its potential uses, employing examples from research on physical fitness and activity.

Constructing and testing latent-class models

Latent-class analysis begins with a two- or higher-order dimensional contingency table of observations. A latent-class model is hypothesized regarding 1) the number of categories or classes (*T*) in the latent variable, 2) the proportion of the population associated with each of the *T* latent classes, and 3) the probabilities associating the observations in each of the classes with each of the levels of the observed variables. For example, Young (1983) examined three measures of psychiatric disorder—formal thought disorder (FTD), blunted affect (BA), and first-rank symptoms (FRS)—to test the hypothesis that these measures provided diagnostic criteria for evaluating a two-class model of the latent schizophrenia versus nonschizophrenia variable in his patient population. The probabilities of observations in the first class having any of the three observed measures present were very small (.000, .153, and .050, respectively), but those associated with observations in the second class were substantially higher (.563, .826, and .652, respectively). Moreover, from his sample of patients he estimated that .535 could be diagnosed as schizophrenic, and .465 could be diagnosed as nonschizophrenic.

As this example illustrates, latent-class models are mathematical representations of relationships between the hypothesized latent variable and the observed variables. Because Young's three observed indicators (FTD, BA, and FRS) were believed to be imperfect indicators of the latent variable, some patients in the nonschizophrenic class may exhibit symptoms of blunted affect (.153) or first-rank symptoms (.050), and some of the patients in the schizophrenic class need not exhibit all three symptoms.

The basic idea of latent-class analysis is that the latent variable "explains" the relationship between the observed variables. Using Young's example, the observed variables are interrelated—for instance, knowing that an individual has blunted affect increases the likelihood that he or she will also have first-rank symptoms present. Within each of the two classes, however, the observed variables are independent of one another—having blunted affect does not increase the likelihood that the individual will also have first-rank symptoms—a condition Lazarsfeld refers to as the axiom of local independence (Lazarsfeld and Henry, 1968). Thus, once the latent variable (schizophrenic/nonschizophrenic) is controlled, the association among the observed variables drops to zero, indicating that the latent variable "explains" the associations appearing when the latent variable is not controlled. Latent-class models are tested by assessing the extent to which the expected response patterns of the locally independent *T*-class model actually correspond to the observed response patterns.

Logic of latent-class analysis

The underlying logic of latent-class analysis can be illustrated with a relatively simple, hypothetical two-way contingency table such as that presented in figure 1.

These hypothetical data indicate that variables A and B are too interrelated for us to attribute the observed pattern to chance alone ($\chi^2 = 38.87$ with 1*df*, $p < .001$). Clearly, we must conclude that variables A and B are related in some manner. The (Pearson) chi-square test here is calculated in the usual manner, in which the expected proportion ($\hat{P}_{ij}$) in each cell is computed as the product of the marginal proportions of the appropriate row and column:

$$\hat{P}_{ij} = \hat{P}_i \times \hat{P}_j \tag{1}$$

and the expected cell frequency ($\hat{F}_{ij}$) is computed as

$$\hat{F}_{ij} = N \times \hat{P}_{ij} \tag{2}$$

where $i = 1,2$; $j = 1,2$; and N is the sample size.

Interpretations of relationships between two or more variables, such as that presented in figure 1, typically takes one of two forms. The first form of interpretation involves the idea of causality—that is, the

		Variable B		
		+	−	
Variable A	+	131	19	150
	−	82	68	150
		213	87	

Figure 1. Hypothetical two-item cross-tabulation.

analyst is inclined to suggest that either variable A or variable B is causally prior (i.e., the independent variable) to the other (i.e., the dependent variable). With causal interpretations, the latent-class model may be used as a categorical data analog to structure equation models. (These models will be explored in a later section as special cases of the *restricted* latent-class model.)

The second form of analysis, however, does not involve the issue of causal ordering; such noncausal relationships are characteristic of the most general form of the latent-class model. Rosenberg (1968) refers to noncausal relationships as symmetrical relationships, and he argues that they can be separated into five types (Rosenberg 1968: 3–7). *Alternative indicators of the same concept* is the type of symmetrical relationship that most concerns researchers in scale analysis—for example, several survey items may be highly interrelated because they are all indicators of the same underlying phenomenon (e.g., depression, stress, anxiety). *Parts of a common system or complex* is the type of symmetrical relationship in which items are interrelated due to common practice; thus, we are likely to believe that the association between nonsmoking and regular exercise is due to their being parts of a common "complex" we refer to as "fitness" or "lifestyle." The third type of symmetrical relationship, *functional interdependence of elements,* involves the interrelatedness of variables that is likely to be jointly present (or absent) due to their significance in the system under study; thus, in organizations with formal, abstract, impersonal rules, one also usually finds administrative rules, acts, and decisions recorded in writing; in organizations where one of these sets of features is absent, the others also tend to be absent. The fourth type of symmetrical relationship, *effects of a common cause,* involves the interrelatedness of variables due to some common causally prior variable; thus, among various nations we may observe an inverse correlation between the annual per capita kilowatt output and birth rates, though we are likely to view both as effects of the level of development or technology, and not electrical output as an effect of birth rate or birth rate as an effect of electrical output. The final type of symmetrical relationship Rosenberg discusses is the *fortuitous* relationship in which two or more variables represent coincidental correlations that have no logical relationship, such as the "rough chronological relationship between rock 'n' roll music and the onset of the space age" (1968: 6).

Latent-class analysis is a method for examining symmetrical relationships among discretely scored (categorical) variables; and Rosenberg's five types of interrelations represent several of the interpretations available in latent-class analysis. Scale analysts using latent-class methods interpret relationships between measures as the result of their being alternative indicators of the same concept. The interpretations of symmetrical relationships as parts of a common complex, as due to functional interdependence, or as effects of a common cause are used in the study of typologies. A researcher's choice of one interpretation over another depends on the dictates of theory and logic. As in the example discussed earlier, we may assume nonsmoking and regular exercise are noncausally related; we may believe that it is most reasonable to assume that the two are parts of a lifestyle or fitness complex. The analysis of typologies with latent-class analytic techniques represents the most general form of the latent-class model; this general model provides a focus for much of the explanation that follows. Recent developments in latent-class methods demonstrate that this analytic method provides insight into the comparative analysis of scales and typologies among different populations (Clogg and Goodman 1984, 1985, 1986). Consequently, when different populations represent different social groups (e.g., regional, ethnic, racial), latent-class analysis methods can be used to compare the scale characteristics and typologies based on identical measures used with each of the social groups. When we identify latent classes at two or more times in the same population of study (e.g., such as when survey researchers analyze several cross-sections in the study of trends), latent-class analysts can study historical changes in the scale characteristics and typologies identified using identical measures at different times in the same population.

Local independence

The hypothetical data reported in figure 1 suggest that variables A and B are not independent of one another, because the χ^2 of independence for these data (38.87) is significant well beyond the .001 alpha level. For expository purposes, we will assume that the relationship presented in figure 1 represents nonsmoking and dynamic exercise as parts of a common fitness complex. That is, each of these two activities are viewed as imperfect indicators of a latent fitness orientation variable. Thus, we hypothesize that persons who are fitness oriented are likely to report participation in a regular exercise regimen and are unlikely to smoke, and those who are not fitness oriented are unlikely to report participating in a regular exercise regime and more likely to smoke. The fitness orientation variable is said to explain the relationship between smoking and regular exercise—variable C explains the relationship between variables A and B.

If we assume for the moment that we have such a measure of fitness orientation—a variable C—then we might conceivably find a distribution such as that presented in figure 2. Here we see that at each level of variable C, the variables A and B are independent of one another (χ^2 = 1.44 with 2*df*), where the expected proportions are calculated as:

$$\hat{P}_{ijk} = \hat{P}^{\bar{A}C}_{ik} \times \hat{P}^{\bar{B}C}_{jk} \times \hat{P}^{C}_{k} \qquad (3)$$

$\hat{P}^{\bar{A}C}_{ik}$ is the estimated conditional probability of a case being at level *i* of variable A given that it is at level *k* of variable C; $\hat{P}^{\bar{B}C}_{jk}$ is the estimated conditional probability of B given the level of C; and $\hat{P}^{C}_{k}$ is the estimated probability of a case being at level *k* of variable C. For example, from figure 2, we estimate that the probability that respondents with fitness orientations report participating in regular exercise $\hat{P}^{\bar{A}C}_{11}$ is .80 (120/150) and the probability that they smoke $\hat{P}^{\bar{B}C}_{21}$ is estimated as .05 (8/150). The estimated probability that respondents with not-fitness orientations report participating in regular exercise $\hat{P}^{\bar{A}C}_{12}$ is .20 (30/150) and the estimated probability that they smoke $\hat{P}^{\bar{B}C}_{22}$ is .55 (79/150). Thus, the estimated probability that fitness-oriented respondents report both participation in regular exercise program and no smoking is equal to the product of their (conditional) probability of participating in a regular exercise program $\hat{P}^{\bar{A}C}_{11}$ and their (conditional) probability of not smoking $\hat{P}^{\bar{B}C}_{11}$. The joint probability that any given respondent participates in regular exercise is a nonsmoker and has a fitness orientation is estimated as .379 ($\hat{P}^{\bar{A}C}_{11} \times \hat{P}^{\bar{B}C}_{11} \times \hat{P}^{C}_{1} = .947 \times .800 \times .500$). To calculate the expected frequencies, we multiply $\hat{P}_{ijk}$ by the total number of cases:

$$\hat{F}_{ijk} = N \times \hat{P}_{ijk} = N \times \hat{P}^{\bar{A}C}_{ik} \times \hat{P}^{\bar{B}C}_{jk} \times \hat{P}^{C}_{k} \qquad (4)$$

where $\hat{F}_{ijk}$ represents the expected number of observations in each of the *i, j, k* cells of the table, and *N* represents the total number of cases reported in the table.

For the example data presented in figure 2, each of the observed cell entries is nearly equal to the expected frequencies for the chi-square test that variables A and B are independent of one another, conditional on the level of variable C, as defined in equation 3 (i.e., $f_{ijk} \approx \hat{F}_{ijk}$). Lazarsfeld and Henry (1968) refer to this condition as "local independence"; that is, when the relationships observed among a set of variables are found to be zero within the categories of some other variable, this set of variables is said to be locally independent with respect to this explanatory variable. Thus, from the data reported in figure 2, we see that variables A and B are locally independent of one another with respect to variable C.

The criterion of local independence, then, provides a method by which we can determine whether relationships observed among a set of variables can be characterized as due to some explanatory variable—in terms of our example, the relationship between regular exercise and smoking is due to the fact that both are components of fitness orientation, as illustrated by the local independence between these variables when fitness orientation is considered. If we had considered additional variables (e.g., diet monitoring) that were all found to be interrelated and if these measures were found to be locally independent of one another when variable C was considered, we would be inclined to interpret these findings as additional evidence that these behaviors are parts of a common complex referred to as "fitness orientation." Thus, when a set of interrelated variables are found to be locally independent within categories of some additional variable, we say that the additional variable explains the observed relationships—that the additional variable represents the true variable of interest, and that once it is considered all of the other measures are unrelated.

Formal latent-class model

The data presented in figure 2 show that we have measured a variable (C) that explains the symmetrical relationship between variables A and B. Typically, however, we are not so fortunate as to have measured the variable within whose categories the symmetrically related variables are locally independent. Thus, the object of latent-class analysis is to define a latent variable—that is, a set of classes within which the manifest variables are locally independent. If such a variable can be defined, then its classes represent either the latent types or the categorical scale locations described by the measured variables within the sampled population.

Before proceeding to a discussion of the characteristics of latent classes and how they can be estimated from observed data, we first return to a discussion of the data presented in figure 2. Recalling the earlier discussion, it was noted that within categories of variable C, variables A and B are completely independent of one another. This relationship is expressed in equation 3 by stating that the proportion of each of the *i, j, k* combinations is the product of the estimated conditional probability of variable A ($\hat{P}^{\bar{A}C}_{ik}$), the estimated conditional probability of variable B ($\hat{P}^{\bar{B}C}_{jk}$), and the estimated prob-

	Variable C			
	+		−	
	Variable B		Variable B	
Variable A	+	−	+	−
+	114	6	17	13
−	28	2	54	66

Figure 2. Hypothetical three-item cross-tabulation.

ability of falling in the particular category of variable C ($\hat{P}_k^C$). When the variable that designates the explanatory categories is not measured (i.e., when variable C is latent), the proportions in equation 3 are symbolized as π, and the latent classes are designated as a latent variable X with *T* classes (levels). Thus this equation is expressed as:

$$\pi_{ijt}^{ABX} = \pi_{it}^{\bar{A}X} \times \pi_{jt}^{\bar{B}X} \times \pi_t^X \qquad (5)$$

where π_{ijt}^{ABX} is the probability that a randomly selected case will be located in the *i, j, t* cell, $\pi_{it}^{\bar{A}X}$ is the conditional probability that a case in class *t* of the latent variable (X) will be located at level *i* of variable A, $\pi_{jt}^{\bar{B}X}$ is interpreted similarly as the conditional probability of being at level *j* of variable B, and π_t^X is the probability of a randomly selected case being at level *t* of the latent variable X. In general, the relationship is given as:

$$\pi_{ij\cdots mt}^{AB\cdots EX} = \pi_{it}^{\bar{A}X} \times \pi_{jt}^{\bar{B}X} \times \cdots \times \pi_{mt}^{\bar{E}X} \times \pi_t^X \qquad (6)$$

which states that the symmetrical relationships among the observed measures are locally independent within the latent classes—that is, that the probability of occurrence in each of the cells of the *i, j, . . ., m, t* cross-tabulation is expected to equal the product of the conditional probabilities ($\pi_{it}^{\bar{A}X}$, $\pi_{jt}^{\bar{B}C}$, etc.) and the probability associated with being in any particular class of the latent variable X (π_t^X). These two types of probabilities—the latent-class probabilities and the conditional probabilities—are the two primary types of parameters of latent-class analysis. Throughout most of this discussion we will consider the instance in which three observed variables are used to estimate the latent variable. That is, we will consider the case where

$$\pi_{ijkt}^{ABCX} = \pi_{it}^{\bar{A}X} \times \pi_{jt}^{\bar{B}X} \times \pi_{kt}^{\bar{C}X} \times \pi_t^X \qquad (7)$$

All of the discussion can, of course, be expanded to cases in which more than three observed variables are used in the analysis.

Latent class probabilities

There are two important aspects to the latent-class probabilities (π_t^X): the *number* of classes and the *relative sizes* of these classes. The number of classes (*T*) in the latent variable (X) represents the number of latent types defined by the observed cross-tabulation; in scale analysis, each class represents a discrete location on the scale. Thus, for example, if the latent variable has three classes, the population can be described as being either three "types" or three levels of an underlying (latent) continuum. The minimum number of identifiable classes in a latent variable is two, because local independence within a single class is equivalent to the observed indicators being independent of one another. As we shall see later, the maximum number of classes in the latent variable is determined by the number of observed indicators, the number of levels measured for each of the observed indicators, and the number of model restrictions that are imposed.

The relative size of each of the classes also provides significant information for the interpretation of the latent-class probabilities. The relative sizes of the latent-class probabilities indicate whether the population is relatively evenly distributed among the identified classes, or whether some of the latent classes represent relatively large segments of the population while other classes represent relatively small segments. The sum of the latent-class probabilities (π_t^X) over all *T* latent classes of the latent variable (X) must equal one:

$$\sum_t \pi_t^X = 1.00 \qquad (8)$$

Later we shall see that the relative sizes of the latent-class probabilities are also very useful for comparing latent structures in two or more populations. When the relative sizes of the latent classes differ significantly between two (or more) populations having similar latent structures, the differences represent distributional heterogeneity of the set of types (or levels of the scale) between the populations. When the different populations represent observations on the same population at two or more times, these differences represent changes in the distribution of the population across the classes over time (i.e., historical trends).

Each of the latent-class probabilities can be calculated by summing the products of the joint probabilities associated with each of the *i, j, k* levels of the observed variables and the conditional probabilities of an observation being in class *t* given that it is located at the *i, j, k* level with respect to the observed variables. That is,

$$\pi_t^X = \sum_{ijk} (\pi_{ijk} \times \pi_{ijkt}^{ABC\bar{X}}) \qquad (9)$$

where the joint probabilities associated with each of the *i, j, k* levels of the observed variables can be obtained by summing equation 7 over all *T* latent classes

$$\pi_{ijk} = \sum_t \pi_{ijkt}^{ABCX} \qquad (10)$$

The conditional probability of an observation being in class *t* given that it is located at level *i, j, k* with respect to the observed variables is obtained by dividing equation 7 by the joint probability associated with the *i, j, k* level of the observed variables

$$\pi_{ijkt}^{ABC\bar{X}} = \pi_{ijkt}^{ABCX} / \pi_{ijk} \qquad (11)$$

As Goodman (1979) has shown, equations 9–11 can be used to obtain maximum likelihood estimates of the latent class probabilities.

Conditional probabilities

The second type of latent-class parameters—the conditional probabilities—are comparable to the factor scores of factor analysis. The conditional probabilities are the probability that an individual in class *t* of the latent variable (X) is at a particular level of the observed variable. Using our earlier example from figure 2, the estimated conditional probability $\hat{P}^{\bar{B}C}_{11}$ represents the probability of the fitness oriented being nonsmokers, and $\hat{P}^{\bar{B}C}_{12}$ represents the estimated probability of the not-fitness oriented being nonsmokers. Similarly, the estimated conditional probability $\hat{P}^{\bar{A}C}_{11}$ is the probability of the fitness oriented reporting that they participate in a regular exercise regime, and $\hat{P}^{\bar{A}C}_{12}$ is the estimated probability of the not-fitness oriented reporting that they participate in a regular exercise regime. Thus, when the type variable is unobserved (e.g., when the fitness orientation variable C is a latent variable X), latent class analysis enables the researcher to identify the conditional probabilities ($\pi^{\bar{A}X}_{it}, \pi^{\bar{B}X}_{jt}, \ldots, \pi^{\bar{E}X}_{mt}$).

For each of the *T* classes of the latent variable, there is a set of conditional probabilities for each of the observed variables. Thus, if three observed variables have been used to define the latent classes, each of the classes will have three sets of conditional probabilities ($\pi^{\bar{A}X}_{it}, \pi^{\bar{B}X}_{jt}, \pi^{\bar{C}X}_{kt}$). Because each of the observed variables can be either dichotomous or polytomous, the number of distinct conditional probabilities for each of the observed variables is equal to the number of levels measured for that variable. That is, if an observed variable has two levels (e.g., exercise versus no exercise), there will be two associated probabilities: $\pi^{\bar{A}X}_{1t}$ the conditional probability that persons of class *t* exercise, and $\pi^{\bar{A}X}_{2t}$ the conditional probability that persons of class *t* do not exercise. Furthermore, if there are three observed variables, there are $I + J + K$ separate conditional probabilities for each of the *T* classes of the latent variable (X). We should note, however, that within each of the *T* latent classes the conditional probabilities for each of the observed variables sum to 1.00:

$$\sum_i \pi^{\bar{A}X}_{it} = \sum_j \pi^{\bar{B}X}_{jt} = \sum_k \pi^{\bar{C}X}_{kt} = 1.00 \qquad (12)$$

Equation 12 states that within each of the latent classes, members have a specific probability of being at each of the particular levels on each of the observed variables. Because the within class conditional probabilities must sum to 1.0, one of the conditional probabilities will be fully determined for each of the observed variables. Thus, in an analysis with three observed variables, there will be $(I - 1) + (J - 1) + (K - 1)$ conditional probabilities that need to be identified for each of the latent classes (i.e., $T[(I-1) + (J - 1) + (K - 1)]$).

To calculate the conditional probability that a member of class *t* will be at a given level of an observed variable, we first sum over the levels of the remaining observed variables the products of the joint probabilities associated with each of the *i, j, k* levels of the observed variables (π_{ijk}) and the conditional probability of an observation being in class *t* given that it is at level *i, j, k* of the observed variables ($\pi^{ABC\bar{X}}_{ijkt}$), and then divide that sum by the appropriate latent class probability (π^X_t).

$$\pi^{\bar{A}X}_{it} = \Big(\sum_{jk} \pi_{ijk}\pi^{ABC\bar{X}}_{ijkt}\Big)/\pi^X_t \qquad (13)$$

$$\pi^{\bar{B}X}_{jt} = \Big(\sum_{ik} \pi_{ijk}\pi^{ABC\bar{X}}_{ijkt}\Big)/\pi^X_t \qquad (14)$$

$$\pi^{\bar{C}X}_{kt} = \Big(\sum_{ij} \pi_{ijk}\pi^{ABC\bar{X}}_{ijkt}\Big)/\pi^X_t \qquad (15)$$

The conditional probabilities enable us to characterize the nature of the types defined by each of the latent classes. Within each of the latent classes, the conditional probabilities indicate whether individuals in class *t* are likely or unlikely to have characteristics of each of the observed variables. Thus, if we were investigating fitness orientation and found that one class was .05 likely to smoke and .80 likely to participate in regular exercise and the other class was .55 likely to smoke and .20 likely to participate in regular exercise, we might be inclined to typify the first latent class as fitness oriented and the second latent class as not-fitness oriented. It is from this usage that conditional probabilities have been described as analogous to factor loadings in factor analysis.

Model estimation and identification

A variety of methods have been proposed for estimating the model parameters. Anderson (1954) and Lazarsfeld and Henry (1968) provide determinantal methods for estimating the model parameters. These methods, however, can yield parameter estimates that are impermissible; that is, estimates that lie outside the interval (0,1) may occur by these methods. Several alternative estimation procedures that avoid this problem have also been suggested (McHugh 1956; Dayton and Macready 1976; Formann 1978; Haberman 1979). The method outlined in this paper was first suggested by Goodman (1974a, 1974b), and it can be used to find maximum likelihood estimates of the model parameters using an iterative proportional fitting algorithm. Clogg (1977) has implemented Goodman's iterative procedure in an extremely flexible and widely used program called MLLSA (Maximum Likelihood Latent Structure Analysis), which will be used to estimate the parameters presented in the examples below.

It is important to note that the estimates of the conditional and latent-class probabilities for latent-class models may not be unique. That is, there may exist more than one set of conditional and latent-class probabilities for any specified number of *T* latent classes—the maximum likelihood estimates may represent local, rather than global, maxima. As in most latent-structure analysis programs, analysts using MLLSA must provide initial, or start, values for the probabilities included on the right-hand side of equation 7; consequently, the analyst needs to try more than a single set of initial values. Also,

$$\begin{array}{ccccccc}
 & \pi_1^X & \cdots & \pi_{T-1}^X, & \pi_{11}^{\bar{A}X} & \cdots & \pi_{(K-1),T}^{\bar{C}X} \\
\pi_{111} & \dfrac{\partial\pi_{111}}{\partial\pi_1^X} & \cdots & & & \cdots & \dfrac{\partial\pi_{111}}{\partial\pi_{(K-1),\,T}^{\bar{C}X}} \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
\pi_{IJK-1} & \dfrac{\partial\pi_{IJK-1}}{\partial\pi_1^x} & \cdots & & & \cdots & \dfrac{\partial\pi_{IJK-1}}{\partial\pi_{(K-1),\,T}^{\bar{C}X}}
\end{array}$$

Figure 3. Partial derivatives matrix for assessing model identifiability.

because the program does not reset a parameter estimate once it has been estimated as either 1.0 or 0.0, alternative initial values may be desired to check the estimates.

The latent-class model is said to be identifiable when there is a unique set of parameters that can generate the observed probabilities. Because there are $IJK - 1$ nonredundant observed probabilities (π_{ijk}), Goodman notes that a necessary condition of identifiability is that the number of estimated model parameters not exceed $IJK - 1$. In the model specified in equation 15—the *unrestricted* model with *all parameters estimated* which Goodman (1974a) refers to as the 'basic set'—the number of estimated parameters is

$$(T - 1) + T[(I - 1) + (J - 1) + (K - 1)] = (I + J + K - 2)T - 1 \quad (16)$$

That is, we are estimating $T - 1$ latent-class probabilities and $(I - 1) + (J - 1) + (K - 1)$ conditional probabilities for each of the T latent classes (recalling the requirement of equations 8 and 12 that the latent-class and conditional probabilities sum to 1.00). A necessary condition of model identifiability, then, is

$$(IJK) - 1 > [(I + J + K - 2)T - 1] \quad (17)$$

Unfortunately, there are models for which equation 17 may be satisfied, but for which no set of unique model parameters exist. A simple example is in the instance of four dichotomous observed variables, and a three-class latent variable ($T = 3$), where equation 17 is satisfied ($15 > 14$), but the model is not identified (Goodman 1974a).

A necessary and sufficient condition for determining the *local identifiability* of a latent-class model is also provided by Goodman (1974a). This involves determining the rank order of a $(IJK-1)$ by $([I+J+K-2]T-1)$ matrix of partial derivatives of the nonredundant observed probabilities with respect to the nonredundant model parameters (figure 3).

Each cell element is a partial derivative of each row element with respect to the corresponding column element:

$$\frac{\partial\pi_{ijk}}{\partial\pi_t^X} = \pi_{it}^{\bar{A}X}\pi_{jt}^{\bar{B}X}\pi_{kt}^{\bar{C}X} - \pi_{iT}^{\bar{A}X}\pi_{jT}^{\bar{B}X}\pi_{kT}^{\bar{C}X}. \quad (18)$$

To be locally identified, the matrix in figure 3 must be of full column rank, or the rank must be equal to $(I + J + K - 2)T - 1$. Calculation of the rank of this matrix is performed by the MLLSA program and is reported along with the parameter estimates.

Unidentified models can be made identifiable by imposing restrictions on one or more of the model parameters. When restrictions are imposed, the right-hand side of equation 17 no longer reflects the number of nonredundant model parameters to be estimated. Consequently, restricting the model means that we must decrement the right-hand side of equation 17 by the number of nonredundant restrictions imposed. Although we will explore models with restrictions more thoroughly below, we first will examine latent-class models without restrictions.

Applications

Latent-class models may be used to address many questions in fitness survey research; the examples presented below represent examples of the flexibility of the latent-class model. An exhaustive inventory of the applications of the model and the MLLSA program are beyond the scope of this paper. Additional applications are documented in the literature (e.g., Clogg and Sawyer 1981; Clogg and Goodman 1984, 1985, 1986; McCutcheon, 1987).

The data used in the examples below come from the 1981 Canada Fitness Survey (CFS), a multistage household probability sample of Canadians between the ages of 7 and 69 years. Data on persons age 30–69 years are examined in the examples below. The cases are separated into males and females and cross-tabulated by three manifest variables: hypertension status (treated, undetected, normal), body mass index (normal, overweight), and leisure exercise (active, sedentary). The first two categories of the hypertension status variable represent the group with high blood pressure, and they have diastolic blood pressure readings of 90 mm Hg or greater. The body mass index is calculated as the ratio of weight (kg) to the square of height (meters) (W/H^2); those persons whose body mass index exceeds 25 are classified as overweight. The cell entries reported in table 1 are the unweighted cases from the CFS.

Unrestricted latent-class analysis

Perhaps the most general application of the latent-structure approach to data analysis is the unrestricted latent-class model. The unrestricted latent-class model can be viewed as a *data unmixing* procedure (Lazarsfeld and Henry 1968) for categorically scored data analogous to traditional exploratory factor analysis for continuous data; with unrestricted latent-class analysis, the analyst attempts to find a latent variable that explains the relationship between the observed categorical variables much as the traditional exploratory factor analysis model seeks to identify the latent factors that explain the correlations between the observed continuous variables. Unlike traditional exploratory factor analysis, however, latent-class analysis does not require distributional assumptions for the manifest variables such as the assumption of multivariate normal distributions required for factor analysis.

Suppose, for the moment, that we suspect that the three observed variables (hypertension status, body mass index, and leisure exercise) characterize different classes, or types, of individuals. We must first determine the number of possible classes of persons that we are able to identify from the three-way contingency table of the observed variables. Recalling equation 17, we see that when $(3 \times 2 \times 2) - 1 > (3 + 2 + 2 - 2)T - 1$, the largest number of unrestricted classes that can be identified is 2 (i.e., $T = 2$). The hypothesis that only one class exists (i.e., $T = 1$) means that we expect the observed variables to be independent of one another within one class—this is equivalent to the independence model. If this is the case, no latent variable is required to explain the relationships between the observed variables, because no relationships are found to exist between these variables. Consequently, the two-class model provides the only meaningful unrestricted latent-class model that can be identified using these three variables.

Statistics reported in table 2 allow us to evaluate the results of the unrestricted latent-class analysis of three-way cross-tabulation of the observed variables for males (table 1). The likelihood ratio chi-square (L^2) is preferred to the usual Pearson chi-square (χ^2), because it is possible to partition the L^2 as restrictions are added to the latent-class model (Goodman 1974a), and thus to allow efficient tests of model improvement as we will see later. In the three-variable case, this likelihood ratio chi-square is calculated as:

$$L^2 = 2\sum_{ijk} f_{ijk} \ln(f_{ijk}/\hat{F}_{ijk}) \tag{19}$$

where f_{ijk} is the observed frequency at levels ijk of the observed variables, and $\hat{F}_{ijk}$ is the expected cell frequency estimated by the specified model. The likelihood ratio chi-square is asymptotically distributed as χ^2, with the degrees of freedom associated with the specified model.

Table 1. Observed cross-tabulation of hypertension status, by body mass index by leisure exercise by sex: CFS, ages 30–69 years

Body mass and leisure exercise	Males			Females		
	Treated	Undetected	Normal	Treated	Undetected	Normal
Normal body mass:						
Active	20	71	372	27	57	555
Sedentary	31	165	718	87	138	1416
Overweight body mass:						
Active	47	152	352	67	52	211
Sedentary	120	342	786	183	180	670

The two additional statistics reported in table 2 have been recommended by Clogg (1981; Clogg and Sawyer 1981) as alternative indicators of model fit. The first of these statistics, the index of dissimilarity (Δ), is a useful measure of association because—unlike the L^2—it is not influenced by the sample size used in the analysis. This index becomes especially useful when we compare the fit of models between groups of unequal size, as we will see later. The index of dissimilarity is given by:

$$\Delta = \sum_{ijk}\left(\frac{|f_{ijk} - \hat{F}_{ijk}|}{2N}\right) \tag{20}$$

where N is the sample size. Clogg (1981) suggests interpreting the index of dissimilarity as the percent of observed cases that need to be shifted to align with the $\hat{F}_{ijk}$ of the specified model. The second statistic, the lambda (λ), is a measure of the overall association between the latent variable (X) and the observed measures (ABC), and it may be used to supplement the L^2 statistic in determining the overall model improvement achieved by introducing the latent variable in accounting for the association among the observed variables. Calculation of the lambda requires the introduction of some of the quantities we have encountered in the general discussion of the previous section and of some we will encounter below in the discussion of the estimated parameters for this two-class model. To calculate lambda, we must first determine the error (E_1) incurred under the assumption that all observations can be assigned to a single class (i.e., the independence model). After fitting the specified model, we note the largest, or *modal*, latent-class probability ($\hat{\pi}^X_{t'}$). The number of cases that would be inaccurately assigned if a one-class model was assumed, then, is given by:

$$E_1 = N(1 - \hat{\pi}^X_{t'}) \tag{21}$$

Next, we must determine the number of errors (E_2) incurred under the specified model. This requires that we calculate the error associated with each of the IJK cells of the original observed cross-tabulation (E_{ijk}). To do so, we calculate:

$$E_{ijk} = (1 - \hat{\pi}^{ABCX}_{ijkt.})f_{ijk} \tag{22}$$

Table 2. L^2, degrees of freedom, Δ, and λ of unrestricted, latent-class models: 1981 CFS males, ages 30–60 years

Model	L^2	Degrees of freedom	Δ (%)	λ
Independence	108.94	11	—	—
Two-class model	2.55	2	0.65	0.52

where $\hat{\pi}_{ijkt}^{ABCX}$ denotes the modal conditional probability that an observation in cell *ijk* will be in class *t* ($t = 1, \ldots, T$). The total number of errors incurred by predicting classes of *X* from the observed response is:

$$E_2 = \sum_{ijk} E_{ijk} \tag{23}$$

and the lambda measure of association is equal to:

$$\lambda = \frac{E_1 - E_2}{E_1} \tag{24}$$

The data reported in table 2 clearly indicate that we must reject the independence hypothesis that the three observed variables are unrelated ($L^2 = 108.94$, 11*df*). Because lambda is a measure of the improvement of fit over the independence model, it is not a meaningful statistic for the independence model itself. Also, we here follow the convention of not reporting an index of dissimilarity for the independence model (Clogg 1981).

Based on the data reported in table 2, however, we can accept the hypothesis that the three observed variables represent two latent classes for the males (age 30–69 years) in the Canada Fitness Survey. The likelihood ratio chi-square ($L^2 = 2.55$, 2*df*) indicates that the two-class model is quite acceptable at the .05 alpha level. The two additional indicators also indicate that the two-class model provides a good fit to the observed data. The index of dissimilarity (Δ) indicates that with the two-class model we would misclassify only 0.65 percent of the observed frequencies, and the lambda—which is bounded by 0.0 and 1.0—indicates that .52 of the errors made with the independence model are eliminated by the two-class model.

How can we characterize the two classes of the latent variable? To do this we must examine the conditional probabilities associated with each of the latent classes specified by the model. To characterize the distribution of the latent classes in the population, we must examine the latent-class probabilities. The conditional and latent-class probabilities identified by the two-class model are presented in table 3.

The conditional probabilities for each of the two latent classes are reported in the upper portion of table 3. We may use these conditional probabilities to characterize each of the two classes of the latent variable in a manner analogous to using factor scores to characterize the latent factors in factor analysis. Thus, we see that observations in the class I are characterized by a 35% probability of reporting participation in an active leisure exercise, a low probability of being overweight (.26), and a high probability of having normal (sub-90 mm Hg diastolic) blood pressure (.86). By contrast, respondents in class II are far more likely than class I respondents to be overweight (.81) and much less likely to have normal blood pressure (.57). Interestingly, participation in active leisure exercise appears to play relatively little role in discriminating between the low-risk (class I) and at-risk (class II) classes (.35 vs .29). In the section on model restrictions, below, we will see that model restrictions enable us to make efficient tests of whether leisure exercise plays a significant role in discriminating between these two classes. Recalling equation 12, within each class the conditional probabilities for each variable must sum to 1.0.

The latent-class probabilities are reported in the lower portion of table 3. The latent-class probabilities allow us to characterize the relative distribution of the two types of respondents in the population. As the data in table 3 indicate, approximately 56% (.559) of the Canadian males, ages 30–69 years in 1981, are classified as at risk, and 44% (.441) are classified as low risk. Recalling equation 8, these latent-class probabilities must sum to 1.0 (.559 + .441).

Sometimes the latent variable examined in latent-class analysis is an intermediate step in the overall analysis; the researcher wishes to explore the use of the latent variable as either an independent or dependent variable in some larger analytic framework. When this is the case, we would like to assign each of the original observations to one of the latent classes of the model. The assignment of observations to classes means that the latent classes can be used as a new variable in further analysis, analogous to the use of factors computed from the factor scores of factor analysis. The assignment of respondents to latent classes is on a cell-by-cell basis—all of the observations in any given cell of the originally observed contingency table are assigned to the same class. The data reported in the fourth column of table 4, for example, indicate that the observations in the four-cells (1,1,1), (1,1,2), (1,2,1), and (1,2,2) are assigned to class I, and the observations in the remaining 8 cells are assigned to class II. The response pattern for the 12 cells of the originally observed contingency table is reported in the first column, and it is reported in the order of body mass index (normal = 1, overweight = 2), hypertension status (normal = 1, undetected = 2, and treated = 3), and leisure exercise (active = 1, sedentary = 2).

Table 3. Estimated parameters for the two-class latent structure model: 1981 CFS males, ages 30–69

Manifest variable	Class (I)	(II)
Leisure activity:		
Active	.35	.30
Sedentary	.65	.70
Body mass index:		
Normal	.74	.19
Overweight	.26	.81
Hypertension status [1]		
Treated	.01	.11
Undetected	.12	.31
Normal	.86	.57
Latent class probabilities	.441	.559

[1] Summing to less than 1.0 due to rounding.

The quantities presented in table 4 allow us to rigorously assess the quality of predictions based on this unrestricted two-class model. The expected probabilities are calculated by multiplying the results of equation 10 by the number of observations in the analysis (N = 3,176). These probabilities are used to calculate the L^2 and the Δ and may be useful in analysis of residuals (Haberman 1979) for locating the source of poor model fits when the L^2 is unacceptably high. In the predicted-class column we report the class to which the modal probability ($\hat{\pi}^{\bar{A}BCX}_{ijkt}$) indicates the cell observations should be assigned. The probability-of-error column is calculated as the difference between 1.0 and the modal probability ($1 - \hat{\pi}^{\bar{A}BCX}_{ijkt}$). As noted by the information reported in the fifth column, the error associated with these class assignments differs for each of the 12 cells. These differential error rates indicate that we can have less confidence in assigning some cells to their modal class than we can in assigning other cells; for example, we are more confident of making few errors when assigning the normal hypertension status, normal body mass index respondents (1,1,1 and 1,1,2) to the low-risk class than we are of assigning the undetected hypertension status, normal body mass index respondents (1,2,1 and 1,2,2) to that class. Also, in cells that differ only on the leisure exercise variable, we see a rather negligible effect of leisure exercise on the probability of error, ranging from no difference for overweight males who are treated for hypertension (.03 vs .03) to a .06 difference for normal body mass men who had undetected (high) hypertension (.41 vs. .47).

This unrestricted model leaves ambiguous the question of whether leisure activity significantly discriminates between the two latent classes. More generally, although unrestricted latent-class models allow us to perform *exploratory* analyses of the associations among the observed and latent variables, they do not allow us to test specific hypotheses regarding the association of the observed variables to the latent classes. In the following section we take up the issue of model restrictions in *confirmatory* latent-class analysis.

Table 4. Assignment of respondents in latent classes for the two-class unrestricted model

Response pattern	Observed frequency	Expected frequency	Predicted class	Probability of error
(1,1,1)	372	369.19	1	0.16
(1,1,2)	718	720.88	1	0.19
(1,2,1)	71	76.47	1	0.41
(1,2,2)	165	159.00	1	0.47
(1,3,1)	20	15.94	2	0.29
(1,3,2)	31	35.52	2	0.24
(2,1,1)	352	353.68	2	0.31
(2,1,2)	786	784.24	2	0.26
(2,2,1)	152	149.06	2	0.11
(2,2,2)	342	345.48	2	0.09
(2,3,1)	47	49.65	2	0.03
(2,3,2)	120	116.89	2	0.03

Restricted latent-class analysis

Frequently, researchers have specific hypotheses they wish to test regarding either the association of the observed variables to the latent classes or the distribution of the latent classes (types) in the population. Hypotheses regarding the association between an observed variable and a latent class involve restrictions on the conditional probabilities, and hypotheses regarding the distribution of latent classes in the population involve restrictions on the latent-class probabilities. For example, below we will test the hypothesis that the relationship between the observed variable leisure activity and the low-risk class is identical to the relationship between the leisure activity and the at-risk class.

Model restrictions can be characterized as being of two general types: equality restrictions and exact indicator restrictions. Equality restrictions on conditional probabilities may be used to test a wide variety of hypotheses. Some, like those mentioned above for the observed leisure activity variable, test the hypothesisthat an observed variable is equally related to two (or more) latent classes. In the two-class model, this type of restriction for a dichotomous variable is formally stated as:

$$\pi^{\bar{A}X}_{11} = \pi^{\bar{A}X}_{12} \tag{25}$$

Because the conditional probabilities for the two levels of the dichotomous A variable must sum to 1.0 within each of the classes (equation 12), constraints on the conditional probabilities for the second level of dichotomous observed variables (i.e., $\pi^{\bar{A}X}_{21} = \pi^{\bar{A}X}_{22}$) are redundant. Other types of equality constraints can also be imposed on conditional probabilities. For example, in scalogram analysis, equality constraints may be placed on different observed variables within the same class to test the hypothesis that the error rates for the observed variables are equal (Clogg and Sawyer 1981; Clogg and Goodman 1986). Equality constraints on latent-class probabilities, on the other hand, tend to be limited to hypotheses regarding the equal distribution of the population into two or more classes. For example, we might test the hypothesis that the 1981 Canadian male population aged 30–69 years is equally distributed into the two classes. Formally, we state this hypothesis as:

$$\hat{\pi}^{X}_{1} = \hat{\pi}^{X}_{2} \tag{26}$$

When there are more than two latent classes ($T > 2$), equality restrictions may be placed on any or all of the latent-class probabilities to test equality hypotheses.

Exact indicator restrictions may be further subdivided into specific value and deterministic restrictions. Specific value restrictions on conditional probabilities test hypotheses regarding the a priori stated strength of an association between an observed variable and a latent class. If we knew that 7% of the population had been treated for hypertension, for example, we might wish to test the hypothesis that the treatment level for members of class I is the same as that for the entire population (i.e., $\pi_{11}^{\bar{B}X} = .07$). Similarly, we can impose specific value restrictions on latent-class probabilities to test hypotheses about a specific proportion of the population falling in a given class (e.g., $\pi_1^X = .33$). Because specific value restrictions require a priori reasons for restricting conditional or latent class probabilities to specific values, this type of restriction has not yet found widespread use in latent-class analysis.

Deterministic restrictions on conditional probabilities, on the other hand, have important applications in latent class analysis. Such restrictions state that observations can be assigned to a specific class on the basis of their value on one (or more) of the observed variables—that an assignment to a latent class is determined by the value on one (or more) of the observed values. An example of a deterministic restriction would be one in which all persons in class I are restricted to have normal hypertension status (i.e., $\pi_{11}^{\bar{B}X} = \pi_{12}^{\bar{B}X} = 0.0$, and $\pi_{13}^{\bar{B}X} = 1.0$). When imposing multiple deterministic restrictions on conditional probabilities, the analyst must take care to provide a potential class for all possible combinations of the observed variables. For example, in a two-class model with a deterministic restriction that all class I respondents have normal hypertension status and a deterministic restriction that all class II respondents have overweight body mass, the 287 persons with normal body mass and high (treated and undetected) hypertension status could not be assigned to either class; this may lead to extreme model identification problems. On the other hand, deterministic restrictions for latent-class probabilities are not meaningful because setting a latent class probability to 0.0 (or 1.0) would be restricting a class to contain none (or all) of the population.

In table 5 are reported the results necessary for evaluating the two-class model in which deterministic restrictions have been placed on the conditional probabilities of the hypertension status variable for class I; and an equality restriction has been placed on the conditional probabilities of the leisure activity variable for the two classes. The models in table 5 have been restricted hierarchically; that is, all of the parameters estimated for each model are also estimated for the previous model (except for the H_3 to H_4 step). Consequently, the unrestricted two-class model can serve as a baseline model with which the conditional probability restrictions can be evaluated. The improvement in fit L^2—obtained by partitioning (i.e., differencing) the model L^2—is asymptotically distributed as χ^2 with degrees of freedom equal to the difference between the models' degrees of freedom.

Table 5. L^2, degrees of freedom, Δ, and λ of the two-class models: 1981 CFS males, ages 30–60 years

Model	L^2	Degrees of freedom	Δ (%)	λ
Independence	108.94	11	—	—
H_1: Unrestricted	2.55	2	0.65	0.52
H_2	2.78	3	0.67	0.47
H_3	8.83	4	1.03	0.28
H_4	5.93	4	1.75	0.46

The improvement of fit L^2 ($[L^2_{H_2}|L^2_{H_1}] = 2.78 - 2.55 = 0.23$ with $3 - 2 = 1$*df*) for the deterministic restriction that class I respondents have not been treated for hypertension ($\pi_{11}^{\bar{C}X} = 0.0$) indicates that, at the .05 alpha level, we fail to reject the hypothesis that no respondents in the low-risk group have been treated for hypertension. In H_3 we add a second deterministic restriction to test the hypothesis that class I respondents do not have undetected hypertension ($\pi_{12}^{\bar{C}X}$), but we must reject this hypothesis given the unacceptably large improvement of fit L^2 ($[L^2_{H_3}|L^2_{H_2}] = 8.83 - 2.78 = 6.05$ with $4 - 3 = 1$*df*). Finally, in H_4 we impose an equality restriction on the conditional probabilities of the leisure activity variable ($\pi_{11}^{\bar{A}X} = \pi_{12}^{\bar{A}X}$). The small improvement of fit L^2 ($[L^2_{H_4}|L^2_{H_2}] = 5.93 - 2.78 = 3.15$ with $4 - 3 = 1$*df*) indicates that we can accept the hypothesis that leisure activity does not play a significant discriminating role between class I and class II type respondents. We should note that successive improvements of fit are calculated from models for which previously imposed restrictions have been accepted. Consequently, the improvement of fit for model H_4 is determined by comparison with the fit for model H_2.

With the exception of the rejected deterministic restriction H_3, the relatively small changes in the index of dissimilarity and the lambda provide further support for concluding that men in the low-risk class have not been treated for hypertension and that leisure activity has little discriminating effect for this population of men. The proximity of the leisure activity restriction's improvement of fit ($L^2_{H_4}|L^2_{H_2}$) to the .05 alpha level critical value (3.84 with 1*df*) and the sizable differences between the remaining probabilities of the two classes indicate that further restrictions on the latent-class model are unlikely to yield acceptable fits to the observed data.

The conditional and latent-class probabilities reported in table 6 represent the parameters for the final restricted, two-class model (H_4). As we can see, a few changes have occurred in the unrestricted conditional and latent-class probabilities. The conditional probabilities for the leisure activities variable have changed very little, converging on values intermediate to the unrestricted values for these parameters. Also, the latent-class probabilities have changed, indicating a higher proportion in the at-risk population and lower proportion in the low-risk population.

The use of equality restrictions is an extremely important aspect of simultaneous latent-class analysis, the topic of the section which follows. In simultaneous latent-class analysis the researcher examines the latent structure among manifest variables observed in two or more groups. An important issue in these analyses is where the latent structure in one group is similar to the latent structure in the other group(s). As we will see below, conditional probability equality restrictions across groups are essential for comparing the latent structures in two or more groups.

Simultaneous latent-class analysis

Often researchers are interested in knowing whether the latent structure among the observed variables in one group is similar to the latent structure among the same observed variables in other groups. In our example of the two latent classes among the three observed variables (hypertension status, body mass index, and leisure activities) for Canadian men, we may wish to know if two or more classes can be used to describe these same observed variables for Canadian women. Moreover, if two classes do appear to provide an adequate description for women, we may further ask whether the two classes for women are identical to two classes for men; that is, we would like to know if the class I women are equally likely to be overweight as are class I men, whether class II women are equally likely to participate in leisure activity as are class II men, and so forth. If the two classes for women differ from the two classes for men, we are likely to want to know the precise nature of the difference. Recent advances in latent-class methods (Clogg and Goodman 1984, 1985, 1986) allow the researcher to simultaneously specify latent-structure models for two or more groups and to impose across-group restrictions on the models to test whether the latent classes are identical in each of the groups.

When a set of observed variables (e.g., *A, B,* and *C*) are available for each of *S* groups, the four-way contingency table can be viewed as *S* three-way ($A \times B \times C$) tables. The conditional probability that an individual in the *s*th group is at the *ijk* level of the observed variables is

$$\pi_{ijks}^{\bar{A}\bar{B}\bar{C}G} = \pi_{ijks}^{ABCG}/\pi_{s}^{G} \qquad (27)$$

where *G* denotes the group variable, and $\pi_{ijks}^{ABCG\bar{X}}$ denotes the probability that an individual will be at level *ijks.*

If a *T*-class latent structure characterizes each of the groups, then the simultaneous latent-structure model may be expressed as

$$\pi_{ijks}^{\bar{A}\bar{B}\bar{C}G} = \sum_{t}\pi_{ijkst}^{\bar{A}\bar{B}\bar{C}G\bar{X}} \qquad (28)$$

where $$\pi_{ijkst}^{\bar{A}\bar{B}\bar{C}G\bar{X}} = \pi_{st}^{G\bar{X}}\pi_{ist}^{\bar{A}GX}\pi_{jst}^{\bar{B}GX}\pi_{kst}^{\bar{C}GX} \qquad (29)$$

Table 6. Estimated parameters for the restricted two-class latent structure model: 1981 CFS males, ages 30–69

Manifest variable	Class I	Class II
Leisure activity:		
Active	[1].32	[1].32
Sedentary	.68	.68
Body mass index:		
Normal	.75	.23
Overweight	.25	.77
Hypertension status:		
Treated	[2].00	.11
Undetected	.10	.31
Normal	.90	.58
Latent-class probabilities	.384	.616

[1] Equality restrictions imposed on conditional probabilities.

[2] Deterministic restriction imposed on conditional probability.

In equation 28 we see that $\pi_{ijkst}^{\bar{A}\bar{B}\bar{C}G\bar{X}}$ denotes the conditional probability that a person in group *s* will be at level *ijkt* of the observed and latent variables, closely parallel to the condition set forth in equation 10. In equation 29, $\pi_{st}^{G\bar{X}}$ is the conditional probability that a person in group *s* will be in latent class *t*, $\pi_{ist}^{\bar{A}GX}$ is the conditional probability that an individual will be at level *i* of *A*, given that he or she is in latent class *t* and group *s*; $\pi_{jst}^{\bar{B}GX}$ and $\pi_{kst}^{\bar{C}GX}$ are similar conditional probabilities for observed variables *B* and *C*.

Each of the parameters of equation 29 are probabilities and must sum to 1.0:

$$\sum_{t}\pi_{st}^{G\bar{X}} = 1.0 \qquad (30)$$

for $s = 1, \ldots, S$, and

$$\sum_{i}\pi_{ist}^{\bar{A}GX} = \sum_{j}\pi_{jst}^{\bar{B}GX} = \sum_{k}\pi_{kst}^{\bar{C}GX} = 1.0 \qquad (31)$$

for $s = 1, \ldots, S$ and $t = 1, \ldots, T$. Consequently, $S(T-1)$ latent-class probabilities and $TS[(I-1)+(J-1)+(K-1)]$ conditional probabilities must be estimated. Together, this basic set of parameters for the heterogeneous, unrestricted *T*-class model totals $S[(I+J+K-2)T-1]$ elements.

Clogg and Goodman (1984, 1985) show that the simultaneous *T*-class latent structure can be expressed as a deterministically restricted model for the $A \times B \times C \times G$ cross-tabulation. To do so, they define a variable *Y* from a two-way cross-tabulation of the latent variable and the group variable (i.e., $Y = G \times X$), and note that *Y* will have ($S \times T =$) *U* levels in which the correspondence between the *U* levels of *Y* and the *ST* levels of *GX* is:

$$u = [(1,1),(1,2), \ldots, (1,T),(2,1), \ldots, (S,T)] \qquad (32)$$

where the first set of *T* elements is for group 1, the next set of *T* elements is for group 2, and so forth to the *S*th set of *T* elements for group *S*. Because *Y* incorporates the latent variable *X*, *Y* is itself a latent variable.

Substituting the *Y* latent variable for the latent variable *X*, the (unconditional) probability that an individual will be at level *ijks* of the $A \times B \times C \times G$ cross-tabulation is:

$$\pi_{ijks}^{ABCG} = \sum_{u}\pi_{ijksu}^{ABCGY} \qquad (33)$$

where

$$\pi_{ijksu}^{ABCGY} = \pi_u^{Y} \pi_{iu}^{\bar{A}Y} \pi_{ju}^{\bar{B}Y} \pi_{ku}^{\bar{C}Y} \pi_{su}^{\bar{G}Y} \tag{34}$$

Thus, equation 34 is expressed as a special case of the basic latent-class model (equation 6) with U latent classes and four manifest variables A, B, C, and G. The correspondence between the U levels of Y and the group variable G (equation 32), Clogg and Goodman note, means that there is a redundancy in the $\pi_{su}^{\bar{G}Y}$ expressed in equation 34 that can be exploited in the model estimation. Because the first set of T elements in U are for the first group (i.e., $s = 1$), the second set of T elements in U are for the second group, and so forth, deterministic restrictions are imposed on the $\pi_{su}^{\bar{G}Y}$:

$$\begin{aligned} \pi_{su}^{\bar{G}Y} &= 1.0 \quad \text{for } u = (s - 1)T + t \quad (t = 1, \ldots, T) \\ \pi_{su}^{\bar{G}Y} &= 0.0 \quad \text{for } u = (s - 1)T + t \quad (t = 1, \ldots, T) \end{aligned} \tag{35}$$

That is, restrictions are imposed on the model conditional probabilities such that individuals in the first set of T latent classes ($u = 1, \ldots, T$) have a 1.0 probability of being in the first group ($s = 1$) and a 0 probability of being in any other group; individuals in the second set of T latent classes ($u = T + 1, \ldots, 2T$) have a 1.0 probability of being in the second group ($s = 2$) and a 0 probability of being in any other group; and so forth to the last set of T latent classes ($u = [S - 1]T + 1, \ldots, U$) which have a 1.0 probability of being in the Sth group and a 0 probability of being in any other group.

Differences and similarities between the latent classes of two (or more) groups can be characterized in several ways. The latent-class model which has no across-groups restrictions on the conditional probabilities (e.g., $\pi_{11}^{\bar{A}Y} = \pi_{1(T+1)}^{\bar{A}Y}$) or latent-class probabilities (e.g., $\pi_1^{Y} = \pi_{T+1}^{Y}$), is referred to as a *heterogeneous, unrestricted T-class* model. If the heteregeneous T-class model fits the observed data, we may then pursue the question of whether the T classes for the groups are similar or different. For example, although there are two classes for each gender, leisure exercise may play an important discriminating role between the two classes for women. To test whether the classes in one of the groups are like the classes in the other(s), the analyst may impose across-group restrictions on the conditional probabilities. If the simultaneous latent-class model continues to provide an adequate fit to the observed data when across-group equality restrictions have been imposed on all of the conditional probabilities, the latent structure is said to be *structurally* homogeneous. When one or more of the across-group equality restrictions on the conditional probabilities results in an unacceptably large improvement of fit L^2, the restriction must be rejected and the model is said to exhibit *partial* structural homogeneity.

Across-group equality restrictions on the latent-class probabilities may also be imposed on simultaneous latent-class models. These restrictions are likely to be most meaningful when across-group equality restrictions on the conditional probabilities indicate that the classes of the latent variable are similar in the different groups. When such (complete or partial) structural homogeneity is found between the groups, across-group equality restrictions on the latent-class probabilities allow efficient tests of hypotheses regarding the *distributional* homogeneity of the latent-class model. That is, when the researcher is satisfied that a class of the latent variable in one group is structurally similar to a class of the latent variable in another group, an acceptable improvement of fit L^2 for the across-group equality restriction on the respective latent-class probabilities of the groups indicates that the class represents an equally large (or small) proportion of the population in each of the groups. Thus, if we were to find an at-risk group for men and an at-risk group for women, we may wish to test the hypothesis that men and women are equally at risk. Because there is one redundant latent-class probability for each of the groups (equation 30), equality restrictions can be fitted to only $T - 1$ of these parameters for each group. When acceptable improvement of fit L^2 are found for across-group equality restrictions on all of the $T - 1$ conditional and latent-class probabilities, the simultaneous latent-class model is said to be *completely* homogeneous.

The cross-tabulation of the three fitness variables from the 1981 Canada Fitness Survey is used to illustrate simultaneous latent-class modelling. In addition to the three observed variables, we now include the sex of the respondent as the group variable (see table 1). On the basis of our earlier findings of the two-class model for males, the first simultaneous latent-class model we estimate is the heterogeneous two-class (per group) model. The L^2 and other fit statistics for this model are reported in table 7.

The likelihood ratio L^2 of independence clearly indicates that the three observed variables and the group variable are not mutually independent. The L^2 (4.66 with 4*df*), index of dissimilarity (0.63), and lambda (0.78) for the heterogeneous two-class latent structure model, on the other hand, indicate that this model fits the observed data quite well. These data, then, indicate that we can readily accept the hypothesis of two classes of respondent types for both men and women.

As we noted earlier, however, acceptance of an unrestricted, heterogeneous T-class model does not mean that the classes in one group necessarily correspond to the classes of the other group(s). In the current example, the acceptability of the heterogeneous two-class model does not allow us to conclude that the two classes for the men are the same as the two classes for the women. To assess the nature of the two classes for each gender, the conditional probabilities and condi-

Table 7. L^2, Degrees of freedom, Δ, and λ for the heterogeneous two-class model: 1981 CFS males and females ages 30–60 years

Model	L^2	Degrees of freedom	Δ (%)	λ
Independence	806.33	23	—	—
Heterogeneous model	4.66	4	0.63	0.78

tional latent-class probabilities (i.e., $\pi_{st}^{G\bar{X}} = \pi_u^Y/\pi_s^G$) are reported in table 8.

These data indicate a clear correspondence between the two classes for men and women. Relative to the class II respondents, the class I respondents in each group are more likely to be of normal body mass, more likely to have normal hypertension status, and more likely to participate in active leisure exercise. There also appear, however, to be differences between the conditional probabilities of the two-class model for men and those for women. For instance, in both of the classes men appear to be about 6% more likely than women to report participating in active leisure exercise (.35 vs .29 and .30 vs .24, respectively). Also, women appear less likely to be overweight than men when compared within their respective classes (.05 vs .24 and .76 vs .84, respectively), and more likely to have been treated for high blood pressure if their diastolic blood pressure is in excess of 90 mm Hg.

Although the data in table 8 indicate that the class I male and female respondents are similar, as are the class II male and female respondents, there may be significant differences between the two groups. The conditional latent-class probabilities also indicate an additional difference between the two groups—the proportion of males in each of the classes appears to differ from the proportion of females in each of the classes. Thus, although both groups appear to be distributed approximately 45% in one class and 55% in the other class, men are more likely to be in class II (.549) and women are more likely to be in class I (.457).

The data in table 9 report the summary statistics used to evaluate the simultaneous latent-class models as equality and deterministic restrictions are imposed on pairs of conditional probabilities. Each hypothesis represents the imposition of an additional restriction; and if a restriction provides an acceptable fit, it is retained in the later models. Thus, the hypotheses presented in table 9 are hierarchical.

The first hypothesis reported in table 9 is for the unrestricted, heterogeneous two-class model, which serves as the baseline model for testing the improvement in fit of the restricted models. The second hypothesis (H_2) imposes an across-group (sex) equality restriction on the treated level of hypertension status for class I respondents (i.e., $\pi_{111}^{\bar{B}GX} = \pi_{121}^{\bar{B}GX}$). This tests the hypothesis that men and women in class I are equally likely to report having been treated for hypertension. The third hypothesis (H_3) imposes a deterministic restriction that the likelihood of their having been treated for hypertension is 0.0 (i.e., $\pi_{111}^{\bar{B}GX} = \pi_{121}^{\bar{B}GX} = 0.0$), which tests the hypothesis that men and women in class I have zero probability of having been treated for hypertension. Each of these hypotheses indicate an acceptable improvement of fit ($L_{H_2}^2|L_{H_1}^2 = 0.0$, and $L_{H_3}^2|L_{H_2}^2 = 0.36$, each with 1*df*).

The fourth model (H_4) imposes an across-group equality restriction on the undetected level of hypertension status for class I respondents (i.e., $\pi_{211}^{\bar{B}GX} = \pi_{221}^{\bar{B}GX}$) to test the hypothesis that class I type men and women are equally likely to have undetected high blood pressure. The improvement of fit for this test is acceptable ($L_{H_4}^2|L_{H_3}^2 = 1.66$ with 1*df*), so we may also accept this model restriction. The next model (H_5) tests the hypothesis that class I men and women have a zero probability of having undetected high blood pressure (i.e., $\pi_{211}^{\bar{B}GX} = \pi_{221}^{\bar{B}GX} = 0.0$). When added to the restriction of H_3, this restriction translates into a test of the hypothesis that all class I men and women have normal (sub-90 mm Hg diastolic) blood pressure. We must reject this hypothesis ($L_{H_5}^2|L_{H_4}^2 = 23.13$ with 1*df*).

Briefly, we next test the hypothesis that class II men and women are equally likely to be overweight (H_6); that class II men and women are equally likely to have normal blood pressure (H_7); that class I men and women are equally likely to have normal body weight (H_8); and that class I men and women are all normal body weight (H_9). The first three of these hypotheses (H_6–H_8) involve across-group equality restrictions and the fourth involves (H_9) a deterministic restriction. All of these

Table 8. Estimated parameters for the unrestricted, heterogeneous two-class model: 1981 CFS males and females, ages 30–69

	Males		Females	
Manifest variable	Class I	Class II	Class I	Class II
Leisure activity:				
Active	.35	.30	.29	.24
Sedentary	.65	.70	.71	.76
Body mass index:				
Normal	.76	.16	.95	.24
Overweight	.24	.84	.05	.76
Hypertension status [1]				
Treated	.02	.11	.02	.20
Undetected	.14	.31	.07	.18
Normal	.85	.58	.92	.63
Conditional latent-class probabilities	.451	.549	.543	.457

[1] Summing to more or less than 1.0 is due to rounding.

Table 9. L^2, Degrees of freedom, Δ, and λ for several restricted two-class models: 1981 CFS males and females ages 30–60

Model	L^2	Degrees of freedom	Δ (%)	λ
Independence	806.33	23	—	—
H_1: Unrestricted	4.66	4	0.63	0.78
H_2	4.66	5	0.63	0.78
H_3	5.02	6	0.67	0.75
H_4	6.68	7	0.75	0.76
H_5	29.81	8	1.47	0.74
H_6	8.81	8	0.76	0.76
H_7	8.90	9	0.74	0.77
H_8	8.90	10	0.73	0.78
H_9	8.92	11	0.78	0.79
H_{10}	26.50	12	1.89	0.78
H_{11}	15.31	12	1.61	0.78
H_{12}	140.40	12	3.39	0.77
H_{13}	15.89	12	1.53	0.78
H_{14}	12.25	12	1.32	0.79

hypotheses provide an acceptable improvement in the model's fit and are thus accepted.

The next three hypotheses involve across-group equality restrictions on conditional probabilities, and they result in increases in model L^2 too large to be accepted. The first of these restrictions (H_{10}) tests the hypothesis that class II men and women are equally likely to report participating in active leisure exercise; the second (H_{11}) tests the hypothesis that class I men and women are equally likely to participate in active leisure exercise. The third restriction (H_{12}) tests whether class II men and women are equally likely to have been treated for hypertension.

The final two hypotheses involve within group, across-class restrictions to test whether activity level discriminates between the two classes for each of the groups. In H_{13} we test the hypothesis that women in class I are equally likely to engage in active leisure exercise as are women in class II. The improvement in fit ($L^2_{H_{13}}|L^2_{H_9} = 6.97$ with 1*df*) indicates that we must reject the hypothesis that class I and class II women do not differ in their level of leisure exercise. In hypothesis H_{14} we test the hypothesis that men in class I are equally likely to engage in active leisure exercise as are men in class II. Here we see the improvement in fit is acceptable ($L^2_{H_{14}}|L^2_{H_9} = 3.33$ with 1*df*), which indicates that we can accept the hypothesis that class I and class II men do not differ in their level of leisure exercise.

In table 10 are reported the conditional probabilities and the conditional latent-class probabilities for the final model (H_{14}) reported in table 9. Only the nonredundant conditional probabilities are reported for leisure activity and body mass index variables.

The conditional probabilities reported in table 10 appear to confirm our earlier characterization of class I respondents as low risk and class II respondents as high risk. Among both men and women, class I respondents are absolutely unlikely either to be overweight or to have been treated for hypertension. Moreover, persons in class I have only a 6% likelihood of having undetected high blood pressure and a 94% likelihood of having normal blood pressure. Somewhat less than one-third of the class I respondents of each gender participate in active leisure exercise (32% for men and 29% for women). The class II respondents, on the other hand, appear to be more at risk; 7 out of 10 class II respondents are overweight, and fewer than 2 out of 3 have normal blood pressure (64%). Interestingly, although class II type men and women are equally likely to have high blood pressure (36%), women are twice as likely as men to have been treated for the condition (19% vs 9%, respectively). Also, class II women are less likely than class II men to report participation in leisure activities (24% vs 32%, respectively), a factor which discriminates between class I and class II respondents for women only.

Table 10. Estimated parameters for the final, partially homogeneous two-class model: 1981 CFS males and females, ages 30–69

	Males		Females	
Manifest variable	Class I	Class II	Class I	Class II
Leisure activity:				
Active	[1].32	[1].32	.29	.24
Body mass index:				
Overweight	[2].00	[3].72	[2].00	[3].72
Hypertension status [1]				
Treated	[2].00	.09	[2].00	.19
Undetected	[3].06	.27	[3].06	.17
Normal [3]	.94	.64	.94	.64
Conditional latent class probabilities	.215	.785	.475	.525

[1] Within group, across class equality restrictions imposed.

[2] Across group deterministic restrictions imposed.

[3] Across group equality restrictions imposed.

Finally, the conditional latent-class probabilities reported in table 10 indicate that a substantial portion of the older (ages 30–69) Canadian male population may be categorized in class II. Recalling the unweighted, observed data from table 1, well over one-half (.57) of this group were classified as overweight, and one-fifth (.21) of those with normal body weight were characterized as diastolic hypertensive. In table 10, we see that one-fifth (.215) of the males are categorized as class I—our low-risk type—and nearly 4 out of 5 are categorized as class II—our at-risk type. Older Canadian women, on the other hand, are more than twice as likely to be categorized in class I (.475), even though one-half (.525) are categorized in class II.

Hopefully, the several models presented here help to illustrate some of the uses of latent-structure analysis and indicate the versatility of the modeling and hypothesis-testing capabilities of the latent-class approach. The reader is encouraged to consult Goodman (1974a), Clogg (1981), and Clogg and Goodman (1984, 1985, 1986) for descriptions of many other models in which model restrictions are used.

Conclusions

The purpose of this paper was to provide the researcher with an introduction to some of the many uses of latent-class modeling. The rapid advances in the latent-class approach have recently begun to bear fruit in the analysis of nominal and ordinal level survey data in many disciplines. The widespread availability of Clogg's MLLSA (1977) program and the flexibility of latent-class models for expressing and investigating hypotheses suggest an exciting potential for this approach in health and fitness research.

References

Anderson, T. W. (1954) On estimation of parameters in latent structure analysis. *Psychometrika* 19:1–10.

Bergan, J. R. (1983) Latent-class models in educational research. In W. E. Gordon (ed.) *Review of Research in Education.* Washington, D. C.: American Educational Research Association.

Clogg, C. C. (1977) Unrestricted and restricted maximum likelihood latent structure analysis: A manual for users. Working Paper 1977–09. University Park, PA: Population Issues Research Office.

Clogg, C. C. (1981) New developments in latent structure analysis. In D. M. Jackson and E. F. Borgatta (eds.) *Factor Analysis and Measurement* (pp. 215–246). Beverly Hills, CA: Sage.

Clogg, C. C. and Goodman, L. A. (1984) Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association* 79:762–771.

Clogg, C. C. and Goodman, L. A. (1985) Simultaneous latent structure analysis in several groups. In N. B. Tuma (ed.) *Sociological Methodology.* San Francisco: Josey-Bass.

Clogg, C. C. and Goodman, L. A. (1986) On scaling models applied to data from several groups. *Psychometrika* 51:123–135.

Clogg, C. C. and Sawyer, D. O. (1981) A comparison of alternative models for analyzing the scalability of response patterns. In S. Leinhardt (ed.) *Sociological Methodology*. San Francisco: Josey-Bass.

Dayton, C. M. and Macready, G. D. (1980) A scaling model with response errors and intrinsically unscalable individuals. *Psychometrika* 45:343–356.

Formann, A. K. (1978) A note on parameter estimates for Lazarsfeld's latent class analysis. *Psychometrika* 43:123–126.

Goodman, L. A. (1974a) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61:215–231.

Goodman, L. A. (1974b) The analysis of systems of qualitative variables when some of the variables are unobservable. Part I–A: Modified latent structure approach. *American Journal of Sociology* 79:1179–1259.

Goodman, L. A. (1979) On the estimation of parameters in latent structure analysis. *Psychometrika* 44:123–128.

Haberman, S. J. (1979) *Analysis of Qualitative Data.* Vol. 2: New Developments. New York: Academic Press.

Lazarsfeld, P. F. (1950a) The logical and mathematical foundations of latent structure analysis. In S. A. Stouffer et al. (eds.) *Measurement and Prediction.* Princeton, NJ: Princeton University Press.

Lazarsfeld, P. F. (1950b) The interpretation and computation of some latent structures. In S. A. Stouffer et al. (eds.) *Measurement and Prediction.* Princeton, NJ: Princeton University Press.

Lazarsfeld, P. F. and Henry, N. W. (1968) *Latent Structure Analysis.* Boston: Houghton Mifflin.

Madden, T. J. and Dillon, W. R. (1982) Causal analysis and latent class models: An application to a communication hierarchy of effects model. *Journal of Marketing Research* 19:472–490.

McCutcheon, A. L. (1987) *Latent Class Analysis.* Quantitative Analysis in the Social Sciences series. New York: Sage.

McHugh, R. B. (1956) Efficient estimation and local identification in latent class analysis. *Psychometrika* 21:331–347.

Rindskopf, D. and Rindskopf, W. (1986) The value of latent class analysis in medical diagnosis. *Statistics in Medicine* 5:21–27.

Rosenberg, M. (1968) *The Logic of Survey Analysis.* New York: Basic Books.

Young, M. A. (1983) Evaluating diagnostic criteria: A latent class paradigm. *Journal of Psychiatric Research* 17:285–296.

Young, M. A., Tanner, M. A., and Meltzer, H. Y. (1982) Operational definitions of schizophrenia: What do they identify? *Journal of Nervous and Mental Disease* 170:443–447.

Framingham Leisure Time Physical Activity Questionnaire

Andrew L. Dannenberg, M.D., M.P.H.
Peter W. F. Wilson, M.D.
National Heart, Lung, and Blood Institute

The Framingham leisure time physical activity questionnaire (figure 1) is adapted from the Minnesota Leisure Time Physical Activity questionnaire (1, 2) and includes some questions from the Paffenbarger questionnaire (3).

Kilocalorie assignments for the 82 activities are listed in table 1. For 63 activities, assignments were based on those published by Taylor (1) and Folsom (2). For six activities, other sources of kilocalorie assignments were identified (4, 5). For the remaining 13 activities, kilocalorie assignments were estimated by comparing the intensity of those activities to others in the list for which published assignments were available. Kilocalorie assignments are based on average intensity of involvement in an activity; the questionnaire does not measure actual intensity of involvement.

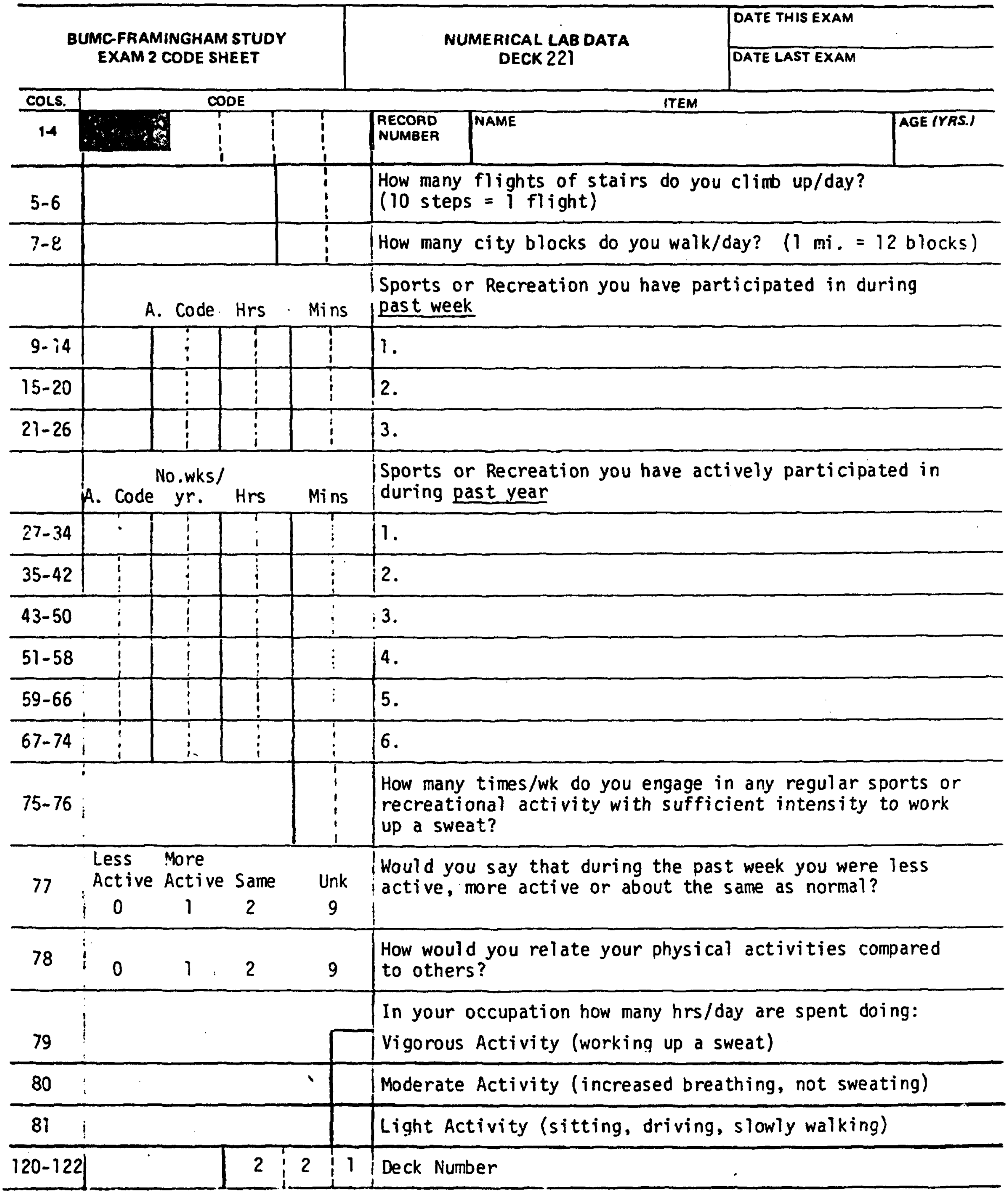
Participants in the Framingham Offspring Study (which include both offspring and spouses of offspring of the original Framingham cohort) (6) were questioned about physical activity during the Cycle 2 examination conducted between 1979 and 1983 at the Study Clinic in Framingham, Massachusetts. Approximately 90% of the subjects in the Offspring Study live in New England, primarily in the metropolitan Boston area.

Initial analysis of this data on activity levels in the Framingham Offspring was recently completed (7). The most common physical activities reported by this group are listed in table 2. Substantial seasonal variation in activities was observed; walking for pleasure was the only activity reported by a large proportion of subjects in every season. Activity levels correlated directly with HDL-cholesterol and inversely with heart rate, body mass index, and cigarettes smoked per day (7).

Other analyses of this data currently in progress or planned include comparison of physical activity by questionnaire to treadmill performance and to echocardiographic left ventricular mass and a comparison of activity levels in the Framingham Offspring to activity levels of their parents in the Framingham cohort.

References

1. Taylor HL, Jacobs DR, Schucker B, Knudsen J, Leon AS, Debacker G. A questionnaire for the assessment of leisure time physical activities. *J Chronic Dis* 1978; 31:741–755.
2. Folsom AR, Caspersen CJ, Taylor HL, et al. Leisure time physical activity and its relationship to coronary risk factors in a population-based sample: The Minnesota Heart Survey. *Am J Epidemiol* 1985; 121:570–579.
3. Paffenbarger RS Jr, Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol* 1978; 108:161–175.
4. Fox SM, Naughton JP, Gorman PA. Physical activity and cardiovascular health. *Mod Conc Cardiovasc Dis* 1972; 41:17–30.
5. Passmore R, Durnin JVGA. Human energy expenditure. *Physiologic Reviews* 1955; 35:801–840.
6. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: The Framingham Offspring Study. *Am J Epidemiol* 1979; 110:281–290.
7. Dannenberg AL, Keller JB, Wilson PWF, Castelli WP. Leisure time physical activity in the Framingham Offspring Study: description, seasonal variation and risk factor correlates. Submitted for publication, 1987.



BUMC-FRAMINGHAM STUDY EXAM 2 CODE SHEET	NUMERICAL LAB DATA DECK 221	DATE THIS EXAM DATE LAST EXAM

COLS.	CODE	ITEM
1-4	[redacted]	RECORD NUMBER NAME AGE *(YRS.)*
5-6		How many flights of stairs do you climb up/day? (10 steps = 1 flight)
7-8		How many city blocks do you walk/day? (1 mi. = 12 blocks)
	A. Code Hrs Mins	Sports or Recreation you have participated in during <u>past week</u>
9-14		1.
15-20		2.
21-26		3.
	A. Code No.wks/yr. Hrs Mins	Sports or Recreation you have actively participated in during <u>past year</u>
27-34		1.
35-42		2.
43-50		3.
51-58		4.
59-66		5.
67-74		6.
75-76		How many times/wk do you engage in any regular sports or recreational activity with sufficient intensity to work up a sweat?
77	Less Active 0 More Active 1 Same 2 Unk 9	Would you say that during the past week you were less active, more active or about the same as normal?
78	0 1 2 9	How would you relate your physical activities compared to others?
		In your occupation how many hrs/day are spent doing:
79		Vigorous Activity (working up a sweat)
80		Moderate Activity (increased breathing, not sweating)
81		Light Activity (sitting, driving, slowly walking)
120-122	2 2 1	Deck Number

Figure 1. Framingham Leisure Time Physical Activity Questionnaire.

Table 1. Kilocalorie assignments for Framingham Leisure Time Physical Activities in order of activity codes

Source	Kcal/ minute	Code	Activity	Source	Kcal/ minute	Code	Activity
M	3.5	1	Walking for pleasure	M	12	42	Soccer
M	4	2	Walking to work	M	3.5	43	Golf—power cart
M	8	3	Using stairs	M	5	44	Golf—pulling cart
M	6	4	Cross country hiking	M	5.5	45	Golf—carrying clubs
M	7	5	Backpacking	M	2.5	46	Mowing lawn—ride mower
M	8	6	Mountain climbing/hill climbing	M	4.5	47	Mowing lawn—power mower
M	4	7	Bicycling—work, pleasure	M	6	48	Mowing lawn—push mower
M	6	8	Dancing	M	4.5	49	Weeding, gardening
M	4.5	9	Home exercise	M	5	50	Spading, digging
M	6	10	Health club exercise	M	4	51	Raking lawn
M	6	11	Jog/walk	M	6	52	Snow shoveling
M	12	12	Running < 10 min/mile	F	7.5	53	Wood chopping
M	6	13	Weight lifting	M	3	54	Carpentry—inside
M	6	14	Water skiing	M	4.5	55	Painting—inside
M	3	15	Sailing—competition	M	6	56	Carpentry—outside
M	3.5	16	Canoeing—pleasure	M	5	57	Painting—outside
M	12	17	Canoeing—competition	M	3.5	58	Fishing—bank or boat
F	6	18	Kayaking, rafting	M	6	59	Fishing—wading
M	6	19	Swimming in a pool	M	6	60	Hunting—pheasants, grouse
M	6	20	Swimming at beach	M	5	61	Hunting—small game
M	7	21	Scuba diving	M	6	62	Hunting—large game
M	5	22	Snorkeling	F	9	64	Hockey
M	7	23	Skiing—downhill	M	3.5	66	Horseback riding
M	8	24	Skiing—cross country	B	5	67	Karate, martial arts
M	7	25	Skating	B	5	68	Car washing
M	7	26	Sledding	M	8	70	Snowshoeing
M	3	27	Bowling	B	4	71	Whiffle ball
M	4	28	Volleyball	B	5	73	Frisbee
M	4	29	Table tennis	F	2.5	74	Dirt bike riding
M	6	30	Tennis—singles	F	3.5	75	Car and boat repair
M	6	31	Tennis—doubles	B	8	77	Wood carrying
M	5	32	Softball, baseball	M	3	78	Sailing—pleasure
M	7	33	Badminton	B	5	80	Coaching sports
M	8	34	Paddleball	B	5	82	Playing with children
M	10	35	Racketball	B	8	83	Lacrosse
M	6	36	Basketball—nongame	P	5	85	Golf ball hitting
M	8	37	Basketball—game	B	5	90	Yard & barn chores
M	7	38	Basketball—officiating	B	5	93	Lobstering, clamming
M	8	39	Touch football	B	5	94	House building
M	10	40	Handball	B	5	95	Umpiring, referee
M	10	41	Squash	B	6	98	Wind surfing, board sailing

Sources:
M Minnesota: Taylor et al (1), Folsom et al (2).
F Fox et al (4).
P Passmore and Durnin (5).
B Best estimate.

Table 2. Physical activities with most total time spent in past year, number of mentions, median weeks of participation, and total caloric expenditure in Framingham Offspring, Cycle 2 examination

Rank	Activity	Total hours/year	Number mentions	Median weeks	Total Kcal/year × 1,000
Men (*N* = 1,598)					
1	Walking for pleasure	36,393	381	32	7,643
2	Weeding, gardening	25,560	327	20	6,901
3	Carpentry—outside	19,380	139	8	6,977
4	Mowing lawn—power mower	18,980	700	20	5,125
5	Golf—pulling cart	15,771	123	20	4,731
6	Running <10 min/mile	15,430	175	36	11,110
7	Swimming in pool	15,347	415	12	5,525
8	Wood chopping	13,334	287	6	6,000
9	Fishing—bank or boat	12,685	162	8	2,664
10	Bicycling—work, pleasure	11,917	213	16	2,860
Women (*N* = 1,762)					
1	Walking for pleasure	55,690	686	25	11,695
2	Weeding, gardening	39,567	531	20	10,683
3	Swimming in pool	21,209	576	12	7,635
4	Health club exercise	14,838	236	16	5,342
5	Dancing	11,539	316	7	4,154
6	Bicycling—work, pleasure	11,249	344	15	2,700
7	Bowling	9,472	193	24	1,705
8	Home exercise	9,355	183	32	2,526
9	Swimming at beach	7,614	438	8	2,741
10	Tennis—doubles	7,177	110	20	2,584

Index

☆ U.S. GOVERNMENT PRINTING OFFICE: 1989— 193 - 320/00695