



\*E001069\*

*Proceedings*

**REGIONAL  
MEDICAL  
PROGRAMS**

*National Conference and Workshop  
on EVALUATION, 1970*

U.S. DEPARTMENT of HEALTH, EDUCATION, AND WELFARE  
Public Health Service  
Health Services and Mental Health Administration  
Regional Medical Programs Service

**Regional Medical Programs**  
***National Conference and Workshop on Evaluation***

**Chicago, Illinois - September 28-30, 1970**

**Center for Continuing Education  
University of Chicago  
Chicago, Illinois**

## PREFACE

The Regional Medical Programs Conference and Workshop on Evaluation, held in September 1970, marks the first time that coordinators and staff members from all 55 Regional Medical Programs met to exchange views on evaluation and to assess their own activities and programs.

A number of factors and circumstances prompted the idea of such a conference-workshop. Most of the 55 Programs were at least three or even four years old. It was a natural time for stocktaking. Changes had been slowly taking place within the Programs and were subtly emerging; goals and objectives, and means and methods for achieving these ends, were being examined; and national priorities and budgetary restrictions were leading the Congress and the Administration to scrutinize federal programs more closely than ever. This current of events emphasized the need for greater self-assessment.

The impetus for the Conference lay largely in the Regions themselves, and most of the Conference planning and development was undertaken by the Regions. Moreover, the content of presentations and discussions were drawn directly from the evaluative work of the Regions. This fact illustrates more clearly than anything else the considerable strides that Regional Medical Programs have made in the past several years — not only in building up their evaluation capability, but also in putting it to good use.

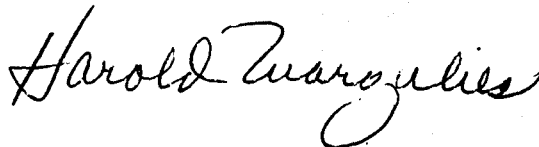
The Conference was significant in its purpose, development and content. Some of the issues posed were broad and generic to the program itself, such as is "change" really the mandate? Others were more specific to evaluation, e.g., how much should be spent on evaluation? Still others were directed to specific aspects of the Regional Medical Programs: What is the Regional Advisory Group's role in evaluation?

If there was a central issue posed by the Conference-Workshops, it must, I believe, have been capsulated by Dr. Donald Schon's presentation. If the whole Regional Medical Program is greater than the sum of its parts, those specific activities supported by it — as its proponents have long argued — then the total *program* must be a primary object of evaluation or assessment.

The Conference-Workshops provided few solutions to the great gamut of issues and problems that were raised. It did, however, make more explicit than ever before those questions that had to be answered. That in itself is a considerable accomplishment and an auspicious beginning.

Any measure of the relative success (or failure) of a conference such as this one must of course be deferred. Its major impact, its final contribution, will only emerge in the actions and changes which will follow.

I hope these "Proceedings" will be useful to those many persons who are concerned with, and who will carry out and evaluate, the Regional Medical Programs, their activities and their efforts. This volume itself provides a fair index of the range of both the interests and work of the Regional Medical Programs to date.



HAROLD MARGULIES, M. D.  
Acting Director  
Regional Medical Programs Service



# CONTENTS

	<i>Page</i>
An Approach to Evaluation for The Regional Medical Program .....	1
Donald A. Schon	
How Others See Regional Medical Programs and Evaluation	
Alexander M. Schmidt .....	21
Peter Fox .....	23
Richard S. Wilbur .....	25
John M. Blamphin .....	27
Luncheon Remarks .....	31
Harold Margulies	
HSMHA -- An Instrument for Improvement in Health Services .....	35
Vernon Wilson	

## Workshop Sessions

### WORKSHOP ON DATA

Participants .....	41
The Values and Limitations of National Data .....	41
Charles A. Metzner	
Data for Ambulatory Care Planning .....	44
J. William Gavett	
Information Systems to Meet Common Data Needs of Health Agencies ..	46
Katherine G. Bauer	
The Northern New England Regional Medical Program Health Planning Data Base .....	51
John E. Wennberg	

### WORKSHOP ON MEASURING CHANGES IN BEHAVIOR

Participants .....	55
Measuring Changes in Knowledge .....	55
William R. Crawford	
Measuring Changes in Clinical Performance .....	55
Barbara J. Andrew	

### WORKSHOP ON THE EVALUATION OF CHANGING HEALTH STATUS

Participants .....	61
Remarks by Moderator .....	61
Robert R. Carpenter	

	<i>Page</i>
The Value of Health Status Measures	
Sam Shapiro .....	62
Maureen M. Henderson .....	67
Discussion .....	72
How to Measure Health Status	
Howard R. Kelman .....	78
Discussion .....	83
The Relation of Process and End-Result Evaluation	
Charles E. Lewis .....	85
Discussion .....	90
 <b>WORKSHOP ON PROGRAM EVALUATION</b>	
Participants .....	99
Approaches to Program Evaluation .....	99
Harold W. Keairnes	
Program Evaluation Workshop — A Case Study .....	104
Harold W. Keairnes	
 <b>WORKSHOP ON RESOURCE ALLOCATION/ECONOMICS</b>	
Participants .....	109
Cost-Benefit and Cost-Effectiveness Analyses in the Health Field .....	109
John Glasgow	
Role of Social and Behavioral Scientists in RMP Evaluation .....	116
Michael Zubkoff	
The Application of Economic Analysis to Regional Medical Programs ...	117
James R. Jeffers	
Summary of Remarks .....	122
John E. Wennberg	
Social Scientists and the Process of Evaluation .....	122
Conrad Seipp	
Cost Benefit — Cost Effectiveness Studies, and Their Application to Allocation of Resources .....	126
Robert L. Berg	
Accountability and Decision-Making in the Iowa Regional Medical Pro- gram .....	130
Charles W. Caldwell	
Resource Allocation and the Evaluation Process .....	141
Charles L. Joiner	

## Special Interest Meetings

## STATISTICAL MODELS AND OPERATIONS RESEARCH

Participants .....	147
A 'Weighted Aggregate' Approach to R & D Selection .....	147
David H. Gustafson, Pai and Kramer	
Comments on an Evaluation Model for the Regional Medical Program ...	157
Vernon E. Weckwerth	
On Evaluation: A Tool or a Tyranny .....	159
Vernon E. Weckwerth	

## EVALUATION OF CORONARY CARE TRAINING

Participants .....	167
Evaluation of Coronary Care Training: Some Direct Observations of Per- formance in Hospital Practice .....	167
Rodger Shepherd	
Report on Xerox Study of Eleven National Coronary Care Training Centers .....	167
Daniele Deverin	
Evaluation of Coronary Care Unit Nurse Education in Washington and Alaska .....	170
Mariella Larter	

## A SYSTEMS APPROACH TO CORONARY CARE EVALUATION

Participants .....	175
A Study of Coronary Care Unit Effectiveness .....	175
M. A. Rockwell	

## EVALUATION OF INSTRUCTIONAL TECHNOLOGY PROJECTS

Participants .....	177
Summary of Session .....	177
Cecelia Conrath	

## EVALUATION OF PHYSICIAN EDUCATION

Participants .....	179
Summary of Session .....	179
William B. Munier	

## EVALUATION OF MULTIPHASIC SCREENING

Participants .....	181
--------------------	-----

	<i>Page</i>
Evaluation of Multiphasic Health Testing .....	181
Donald N. Logsdon	
 <b>EVALUATION OF STROKE – REHABILITATION PROJECTS</b>	
Participants .....	189
Evaluating Stroke and Rehabilitation Programs: An Overview .....	189
Charles M. Wylie	
An Evaluation of a Stroke Program in California .....	192
Bertram L. Tesman	
North Carolina Comprehensive Stroke Program .....	194
B. Lionel Truscott	
 <b>EVALUATION OF CANCER REGISTRIES</b>	
Participants .....	197
Use and Evaluation of Cancer Registries .....	197
Abraham Ringel	
Alternative Methodologies for Evaluation of Registries .....	198
George Linden	
Methodologies for Evaluating Effectiveness and Value of Registries .....	200
Charles R. Smart	
 <b>EVALUATION OF REGIONAL ADVISORY GROUPS</b>	
Participants .....	207
Regional Advisory Groups as a Factor in the Regionalization Process ...	207
David A. Pearson	
Bases for Regional Advisory Group Evaluation .....	211
Paul E. White and Van Hove	
 <b>MEDICAL CARE EVALUATION</b>	
Participants .....	219
ABCD Strategy of Patient Care Assessment .....	219
John W. Williamson	
 <b>EVALUATION OF NEW CATEGORIES OF MANPOWER</b>	
Participants .....	225
Evaluation of New Categories of Manpower .....	225
Harriet Kitzman	

**TRAINING FOR EVALUATORS**

Participants .....	227
Summary of Session .....	227
Marion Leach	

<b>ATTENDANCE BY ORGANIZATIONAL AFFILIATION .....</b>	<b>229</b>
---	------------

# AN APPROACH TO EVALUATION FOR THE REGIONAL MEDICAL PROGRAM

DONALD A. SCHON, President  
Organization for Social and Technical Innovation

## Introduction

The questions in which we are primarily interested are these:

- What are the criteria, methods, and measures pertinent to evaluation of the activities of the Regional Medical Program?
- How can evaluation be linked most effectively to the planning process?
- What are the appropriate roles for those engaged in evaluation at project, regional, and national levels?

These questions have a deceptively simple ring. They raise, in fact, not only the special problems stemming from the nature, context and history of RMP but several more fundamental questions of theory concerning the evaluation of *any* activity.

## Section 1 Toward a General Theory of Evaluation

Evaluation is an essential part of intelligent individual and organizational behavior.

It is the process through which individuals or organizations perceive the consequences of action, assess their meaning for future action, and reformulate plans and policies.

Within this framework evaluation serves three distinct purposes:

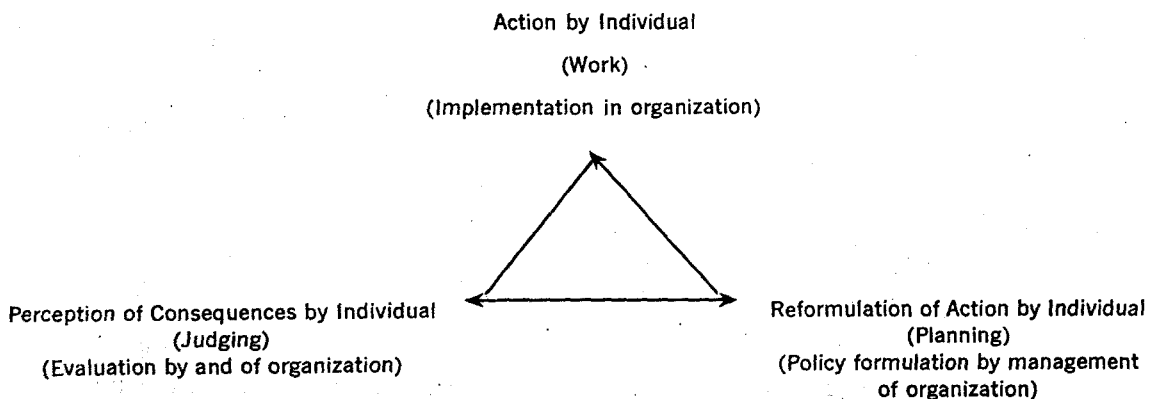
*Justification:* to defend what's planned or what has been done. We justify in order to assign reward or punishment (as in "grading"), to decide what resources to commit to an activity, or simply to place an activity on a scale of excellence. In any case, justification concerns itself with identifying what has been done, or what is proposed, and appraising it against some standard.

*Control:* to monitor an on-going activity in order to make it conform to standard.

*Learning:* to change activity, to do it better. Learning may be limited to the selection of means to achieve goals or to conform to standards, or it may encompass change in the goals and standards themselves.

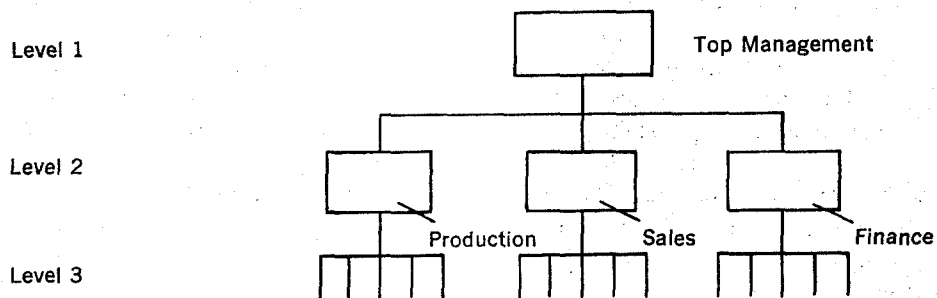
For any program such as RMP, there are always demands for justification, control and learning. But it is not always recognized that these several purposes have different implications for methods and systems of evaluation.

We are accustomed to think about evaluation from the point of view of a rational manager who supervises



the business of an organization or program. The rational manager takes as his reference point a *systems rationale* — that is, a set of formal objectives, operations for achieving them, and methods for appraising the effectiveness of operations in achieving objectives. In a business firm, the systems rationale makes reference to profits and return on investment; in the public housing system, to the provision of standard housing for persons cut off from access to the market; in the health care system, to improvement in people's health, in the quality of care, or in equitable access to care.

According to the rational manager's model of evaluation, the systems rationale is fixed and given. Justification



consists, then, in assessing the impact of past or proposed activity on established systems objectives. How effective are these activities in meeting objectives? How efficiently do they use resources? Control consists in monitoring ongoing activity to make it conform to established standards. Learning is limited to the selection of means for achieving objectives.

The evaluation process appropriate to the rational manager's model depends on the assumption that everybody in the system is to some extent a rational manager. People's accountabilities for activities within the system are supposed to mirror the systems rationale.

Within the organization or program, as within the systems rationale, activities are organized hierarchically. Each person is accountable for the activities of his component, whose goals are keyed, in turn, to the objectives of the system. The job of evaluation is to compare accountabilities with the actual behavior of individual components within the system. Evaluation tends, then, to become an auditing process in which a third party assesses behavior in terms of the systems rationale, and sends information toward the top of the system. On the basis of this information, decisions flow downward to influence the behavior of the components below. At each successive step of the way, the primary use of information is in justifying and then in controlling the per-

formance of the components the information is intended to characterize.

All variants of the rational manager's model and the evaluation systems that flow from it suffer in practice from an overriding constraint. Characteristically, systems do not behave as they are supposed to. Even the most bounded organized activities result in social systems that do not behave exclusively in terms of the rational purposes assigned to them. As distinct from the rational manager's model, there is always a real system of actors and agencies which interact with one another in the ways they are found to do and with the interests they are found to have. Their discovered interactions and

interests may have little to do with the interactions and interests imputed to them under the systems rationale.

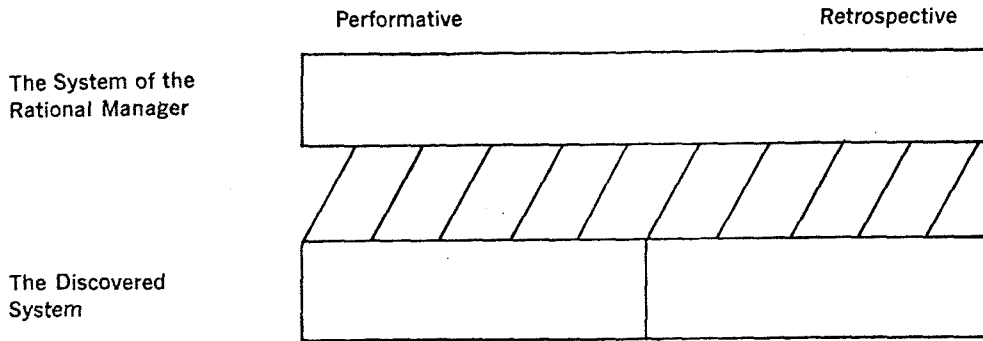
The "discovered systems" of organizations and programs tend to have certain features in common. Regardless of systems rationale, individuals tend to be interested in:

- their own survival in their positions;
- independence of action;
- local conditions and needs (as opposed to "central's" view of them);
- protecting and extending territory;
- maintaining stability.

These interests characterize the informal, homeostatic structure of organizations and programs. But discovered systems tend also to be open-ended, associated with emergent objectives and swift changes in goals which correspond to individual interests in creativity and responsiveness. Often the rational manager's model constrains creativity, responsiveness and freedom of action in ways that run directly counter to the interests of actors and agencies within the system.

Within any on-going program, the rational manager's model and the discovered system always co-exist. The state of their relationship critically determines the nature of evaluation.

When the two systems have little overlap and little interaction, evaluation is limited to retrospective justification.



In this condition, the evaluation system produces statements believed neither by the producer nor by the consumer, which are generated ritualistically in response to formal demand. Rational managers produce justifying statements at regular intervals, expressed in the language of the systems rationale, and resources continue to flow into the system. Evaluation processes have no other output than justification. They are used neither to modify the systems rationale nor to force the real social system to conform to it.

Where there is little overlap, but the rational manager seeks to impose a systems rationale on the discovered system, several things may happen:

1. The discovered system may respond verbally without other changes in behavior, by offering *pro forma* retrospective justification long on language but short on substance, a process generally known as "conning." The two systems operate substantially in parallel.

2. The discovered system may respond to the controls that the rational manager seeks to impose by adapting to the evaluation measures he prescribes but continuing to operate as much as possible as before. Measures of performance are always different from performance itself. For example, in an effort to control expenditures of the vocational rehabilitation system, Congress demanded to know how many "rehabilitations per year" the agency effected for a given investment. "Rehabilitations" were defined as job placements lasting three months or more. As a consequence, the vocational rehabilitation system began to "cream" its clientele for those most likely to graduate to job status leaving out those who were most in need and least able to qualify; to select low-level jobs for graduates so as to facilitate entry; systematically to avoid distinguishing between a "case" and a person, so that a graduate who had achieved job status, lost it and returned to training, could be counted as another "rehabilitation"; and systematically to avoid follow-up of clients after three months.

3. The discovered system and the rational manager's system may fight one another more or less openly until they reach a compromise. From the point of view of the discovered system, this is paying a price. Those in the system do some of what the rational manager wants in order to preserve considerable ability to satisfy the interests of the discovered system. From the point of view of the rational manager, the discovered system is merely distorting system objectives in the direction of its own interests; but he has to put up with it to get any response at all.

In none of these dissociated cases is there any interest in producing or using information that runs counter to the strategy of evaluation as justification. Where the systems are operating in parallel but without much contact, there is common interest in avoiding information that threatens dissociation. In the other two cases, there is common interest in information that supports the systems rationale; since justification rests on the systems rationale, and resource allocation rests on justification. The discovered system is content to generate information that conceals how great the discrepancy is between the goals of the rational system and the behavior of the discovered system in order to protect the resource allocation they need to continue doing more or less what it is they want to do.

However, where the whole activity is conceived as a learning system, then relationships between rational and discovered systems can be fundamentally different from those just sketched. The opportunity for learning is primarily in the discovered system. The discovered system offers the most vital basis for reformulating systems objectives and redesigning systems theory. Discrepancies between the rational manager's system and the discovered system as perceived by its inhabitants become the basis for progressive modification of the system's rationale, of modifying the real interests of



individual participants, and of developing relationships between the total activity and its constituencies.

*It is critical that an evaluation system aspiring to an important role in intelligent management recognize rather than bury discrepancies between systems rationale and the discovered system.* The evaluation system itself must become a vehicle for continuing interaction and continuing mutual influence of the two. Its ability to support intelligent, direct interactions between the rational manager's system and the discovered system becomes a central function and a central criterion of adequacy in an evaluation system oriented to learning. While these considerations are important at all times, they become essential in a period of development or instability, when new kinds of activity must be devised to meet established objectives more effectively and when program environment changes so as to lead to shifts in objectives, as well.

#### **Learning-Oriented Evaluation in Discovered Systems Hooked to Rational Systems**

When planning begins to incorporate a mutual modification of objectives and activities, evaluation includes much more than mere measurement of the extent to which activities conform to specification. The evaluation system that is oriented to learning has special features:

- The conceptual framework for evaluation has to include a description of the discovered system as well as the rational manager's statement of systems rationale. This includes a description of key actors and agencies, actual relationships and modes of interaction among them, and the several interests of all of them. It must include also a description of the real (if informal) evaluation system as discovered – the information that actors in the system in fact produce, are interested in producing, and how they use it.
- An analysis of discrepancies and overlaps between the systems rationale and the behavior of the discovered system. This analysis takes account of the differing perspectives of actors in the system.
- Strategies for responding to discrepancies between the discovered system and the rational manager's system. Mere analysis is not enough; learning must be capable of application.

These factors focus on gathering accurate information about the discovered system. The discrepancy between the rational system and the discovered system, or the response of the discovered system to the rational

manager's efforts to control it, may mean that the rational manager is simply precluded from learning what's actually happening in the discovered system. But the rational manager may be able to bargain for this information by exchanging information about resources and ongoing administrative changes to which he is privy for accurate information about what's really happening in the social system. Even more powerful, when central rational management gains some freedom to modify systems rationale to take account of real local interests and activities, the basis for withholding or distorting information may disappear. The way may then be clear for central rational management and local people to bargain effectively and directly over changes in systems rationale, local behavior modification, and information flow. As in all such cases, the bargaining will depend on establishing and maintaining good faith.

Several additional consequences for the evaluation system flow from these considerations:

- Information intended to modify behavior must flow upward to influence systems rationale as well as downward to bring the discovered system into line with pre-existing systems rationale.
- The evaluation information that is gathered should be limited to amounts, complexities, and precisions determined by the capability and willingness of actors within the system to learn from it, as experienced in actual practice. Nobody in the system should be presented with more information than he can handle, nor information laid out in more precision or complexity than he can respond to. Analyses should not present actors with a greater breadth of alternatives than are real for them. As a corollary, the evaluation system needs to be able to detect the changing capability and willingness of actors to use information, and should itself be capable of responsive modification in turn.
- The evaluation process should be structured to accommodate the different kinds of learning appropriate to different roles and levels within the system (rational managers, project pushers, evaluators, planners, etc.).
- The learning objective should also determine the content, extensiveness, duration, and accessibility of information in the evaluation system memory. This requirement places high priority on accessibility and retrieval capability on behalf of many different levels within the system in addition to that of the rational manager.

- Since the learning derived from evaluation may be applied to evaluation processes themselves, the conceptual framework for evaluation may itself be expected to change (sometimes rather rapidly); so information needs to be gathered and formulated in ways that make it more or less equally usable in terms of a broad range of systems rationales. Priorities should be given to those bits of information that are likely to retain high relevance across a range of manager's rationale and discovered systems.

### Cases in Which There is No Explicit Systems Rationale

What if the activity to be evaluated is itself recognized as so diverse, diffuse, swiftly changing, and open that no overall systems rationale is credible? This situation may occur with respect to public problems urgently requiring solution but for which there are no clear policy answers, where national willingness to devote resources to their solution is high, though the credibility of proposed rational solutions may be low. Agencies may be funded to work on such problems, constrained only within very broad limits as to what their work should be like. What are the implications here for evaluation systems?

- Each region or subregion (or other entity) saddled with a whole problem becomes a center of its own problem-solving process. The number and location will depend on the number of centers that turn out to be capable of functioning under their own individually developed systems rationales. In this situation the distance between information and analysis is minimized, and responsibility for designing and conducting the evaluation process is very close to the actors who are accountable for the activities under evaluation.
- In this case central management's evaluation function is changed with respect to that of the regions. Central management may now impose on the localities criteria for the *evaluation process*, but it is no longer in a position to impose criteria for *substantive evaluation* of concrete activities. For example, central management can still ask whether regional evaluation processes are differentiated in terms of justification, control, and learning; but the central evaluator will accord just as high marks to a region displaying one workable form of differentiation as to a region displaying another form. It is only the region that does not explicitly

attempt through its own evaluation processes to accomplish justification, control, and learning that is downgraded. Accordingly, the evaluation information flowing to central from the local regions normally reflects the nature of the *processes* developed for raising and answering evaluative questions in the localities rather than the *answers to any specific questions* thought up by central management.

- Central also takes on the role of building a network learning system, facilitating information-transfer from locality to locality and encouraging specific local experiments.

### Section 2 RMP in the Context of Evaluation Theory

To place the Regional Medical Program in the evaluation context developed in the previous section, some of RMP's principal characteristics should be recited.

1. There is no single organization corresponding to RMP. RMP is a broad-aimed Federal program concerned with introducing changes of various kinds into a number of more or less interconnected systems of actors and agencies involved in health care. Within these systems, RMP attempts to play a variety of related roles with respect to other actors and agencies; but for the most part it cannot directly control them. RMP does not, therefore, have to do with a single rational "system," in the sense used earlier, and its boundaries are vague and shifting.

From the point of view of evaluation, this assertion has several implications. RMP's scope and turf do not have sharp boundaries. We cannot go about analyzing RMP as though it were a unitary organization, like the Veterans' Administration, for example. And while RMP has formulated broad objectives for itself, its fundamental activity in relation to these objectives must be understood for the most part as "influencing" or "facilitating" rather than direct control.

2. There is no single, established systems rationale either for the health care system as a whole or for RMP in particular. There are various rationales, held at various times and in various contexts by different actors in the system.

3. The larger health care system and the RMP are changeable. They are not in a stable state. The character and functions of these systems are themselves in process of constant change. Within them, the key actors are often unsure of their principal functions or of how best

to carry them out, and they tend to shift behavior as they learn and as the system around them changes.

4. Nevertheless, as a federal program RMP is locked into a structure of controls and demands for justification. At the national level these include regular reviews by the Congress, the Bureau of the Budget, and the Department of HEW. These demands for justification and for controls over the expenditure of funds are, of course, passed on to the regional program level.

The problem of devising approaches to evaluation for RMP is essentially that of meeting what may well be conflicting requirements for learning, on the one hand, and for justification and control, on the other. The vagueness and changeableness of objectives, lack of program control over components to be influenced, and sources of methodological uncertainty all argue for a flexible, process-oriented approach to evaluation-as-learning; whereas the agents of rational administrative control tend to press for firm, quantitative measures of program impact.

Like most broad-gauged federal programs, the legislation establishing RMP represented a series of compromises among the diverse interests of various concerned groups. The authorizing legislation is, therefore, a kind of mosaic of objectives, values, and constraints. Among the more important elements of the mosaic are these:

- Emphasis on the provision of means to improve the treatment of the three "categorical" diseases — heart disease, cancer and stroke.
- Emphasis on the transmission of advanced techniques and knowledge relating to these diseases.
- Emphasis on the method of continuing education as a device for this transmission; and on the major academic medical center as the principal source of expertise.
- Emphasis on maintaining or improving the quality of medical care.
- Concern with the region as the principal unit of activity; concern, that is, that the program be a regional one, with regional centers of activity throughout the country; concern with recognition of regional diversity of problems and resources; and concern with "regionalization" as a process of knitting together or building regional resources to realize the purposes of the Act.
- Emphasis on the establishment of voluntary arrangements among regional institutions as the dominant mode of program activity.
- Specific warning against "interference in the interface between patient and doctor."

The authorizing legislation made no attempt to rationalize these elements or to resolve potential conflicts among them. It was understood by many of the key actors that, as the program matured, the specific meaning of its legislative provisions would develop and clarify.

It is not surprising, then, that there have been perceptible shifts over time in the dominant systems rationale for RMP, even though no element originally considered as the legislation evolved has altogether ceased to exert some influence.

Let us be explicit about an evaluation scheme that is generally accepted as appropriate to one of the simplest and accordingly most easily rationalized interpretations of RMP. We refer to the center-periphery regionalization model based on the diffusion of technology and information that is assumed to be stored in the great medical centers. In this instance, it is seen as desirable to judge the program initially, at both national and regional levels, by its effectiveness in reducing rates of mortality and morbidity for heart disease, cancer, stroke, and related diseases. Individual projects are seen as means to these ends, and fall basically into the following categories: deployment of new facilities (for example, coronary care units); establishment of new linkages between medical centers and peripheral care-providing centers (for example, exchange of personnel); the development of new working relationships (for example, changes in referral patterns); continuing education (for example, training of physicians and other medical personnel); and information dissemination (for example, DIAL access).

The major kinds of evaluative questions under this interpretation of the RMP system are these:

1. What are the kinds of baseline data and measures of performance by which the impact of diffusion projects on mortality and morbidity can be assessed?
2. What is the relative effectiveness and efficiency in relation to cost of the various technologies diffused, seen as means of achieving reductions in rates of morbidity and mortality?
3. What is the related effectiveness, for particular technologies and for particular regional situations, of the various methods of diffusion? This question leads, in turn, to questions about the optimal "regions" for diffusion, the forms of greatest "diffusion impact" for a given investment of dollars and other resources, patterns of utilization of new facilities and the like.

Other aspects of the activities within the center-periphery model of RMP — for example, the management of new institutional arrangements at the regional level — must be judged in terms of their effectiveness in

leading to enhancement of the quality of care through the more effective diffusion of advanced technology, with the ultimate effect, of course, of reducing mortality and morbidity from the categorically identified diseases.

In the minds of many key actors in Washington and in the regions, the DeBakey model came to dominate the conceptual climate of the early phases of RMP. But it was not always or everywhere the dominant view of RMP activity. *In the discovered systems of some of the regions, regional co-ordinators and other key actors took as primary the sorts of changes in institutional arrangements which, from the point of view of the DeBakey model, figured only as secondary means to an end.*

In this interpretation:

- *RMP's central concern may be expressed through categorical diseases or with the diffusion of advanced medical technology, but RMP consciously concerns itself with overall improvement in quality of care and equity of access to care.*
- *But these sorts of improvements require changes in the structure and modes of interactions of care-providing institutions which no single agency controls – changes that can be generally described as knitting together components of the system that are now fragmented so as to permit more effective and rationalized planning and action.*
- *These systems changes are necessary conditions for improvement in quality or equity of care. They must precede any significant improvement along these lines.*

In the past year, systems transformation\* has begun to dominate among competing systems rationales for RMP (without, of course, completely displacing other views) at national as well as some regional levels. While it is to some extent a subject for guesswork why this shift has occurred, certain factors suggest themselves.

There has been a movement into good currency of certain basic concerns about the national system for providing medical care – concerns about rising medical costs, about the effective exclusion from the health care system of large numbers of disadvantaged people, about shortages of medical manpower, about the difficulties of negotiating the medical care system even for ordinary middle class people.

---

\*“RMP as process,” “RMP as facilitator,” “RMP as opportunistic change agent” were expressions heard as early as 1967 and conveyed the underlying idea behind systems transformation before this rationale became as significant as it now is. Recent legislative proposals convey the idea even more explicitly.

The effects of substantial investment in Medicare and Medicaid have begun to convince observers that no amount of investment in payment for care will suffice to introduce necessary changes in the provider system. There is clearly need for some forms of intervention on the provider side as well.

There continue to appear to be overriding objections either to the development of nationalized systems of care or to such decentralized solutions as community-based group practice, on a large scale. Shortages of scarce resources of medical manpower suggest that changes in the system will have to work with existing personnel and, very largely, with existing institutions. This means, to a great extent, attempting to facilitate voluntary re-arrangements of existing institutions.

Of the available program instruments (Neighborhood Health Centers, Comprehensive Health Planning, Community Mental Health Centers), RMP presents itself as perhaps the most promising candidate for intervention of this kind. What RMP has been doing, initially en route to the DeBakey model in some regions or in other regions as a matter of primary though informal agenda, now is emerging as a more dominant (though not exclusive) rationale for the program as a whole. It must be added, of course, that by no means all regions regard themselves as primarily involved in systems transformation. Some RMP's still regard themselves as solicitors and screeners of proposals, and do not yet conceive of themselves as “programs” in any sense other than as clearing-houses for projects. And in nearly all regions, there is the residue of the view of RMP as a conglomerate of projects centering around continuing education, training, coronary care units, and the like. At the very least, then, co-ordinators face, as part of the task of systems transformation, the problem of what to make of and what to do with the projects initiated under earlier views of RMP.

Under a systems transformation model for RMP:

- The primary unit for evaluation becomes the program; and since RMP is conceived as an essentially regional enterprise, this means the regional program. It will be necessary to reach both “above” this level to the national program and “below” it to the project; but the regional program is primary.
- Every element of RMP takes on a dual aspect. As we seek to assess projects, regional program and national program, we must ask *both* about substantive changes in the provision of care – changes in the quality and configuration of services,

changes in access to services, changes in health – and about systems transformation. Seen as systems transformation, RMP functions in two ways: through the direct efforts of the regional coordinator and those he works with to knit together or otherwise influence elements of the medical care system of his region, and through the shaping and selection of projects which become *occasions* to effect systems transformation.

- Evaluation must take account of regional diversity. The starting conditions of the region, the array of resources, the problems to be attacked, the level of development, the regional strategy – there may be as many of these as there are regions. From the point of view of evaluation, therefore, the content of regional programs should be expected to be different. There is no “model” of a regional program to be applied to all regions, although we should be able to develop a conceptual framework which will allow assessment of diverse regional models.
- Evaluation must not only take account of this regional diversity; it must also take account of the fact that regional programs are in critical ways open-ended.

Regional programs undertake systems transformation by engaging the emerging issues of medical care in the region. These are only partly, if at all, within the coordinator’s control; to be effective he must use them and build on them. Evaluation must take account of the open-ended or existential character of regional activity; except within a very broad range, it cannot second-guess the issues to be encountered in a particular region at a particular time; and it must not impose on the region a model of sequential activities independent of the issues of medical care which in fact arise.

The central questions of evaluation now become these:

1. *How can we facilitate learning about systems transformation, at all three levels, but with emphasis on the regional program?*

2. *Given regional diversity and open-endedness, on what basis can we control regional activities or hold them to standard?*

3. *Given the several levels of change relevant to evaluation of RMP, how can we go about the justification of past or projected regional activity?*

The questions of justification demand separate treatment. Given the multiple impacts of RMP activity, justification requires methods for identifying baseline data, ends-in-view, and indicators of change at the several

levels of change in health, access to health care, quality of care, configuration of health resources, as well as changes in the institutional arrangements, interactions and attitudes characteristic of the health care system. The issue of justification raises sharply the problem of what it is possible to *know* about these matters, and at what level of generality it is possible to know it.

The remainder of this paper will be taken up with questions (1) and (2), above. We will focus on the view of RMP as systems transformation and will attempt to spell out the bases on which, in spite of regional diversity and open-endedness, judgments about regional performance may be made and learning about systems transformation may be fostered.

### Section 3

#### The Central-Regional Dialogue

There is a conceptual framework for systems transformation in RMP from which we can derive criteria and questions useful in undertaking and assessing systems transformation, without violating regional differences and without second-guessing particular regional answers to the substantive questions of medical care.

The essential elements to which attention must be paid are these:

- Starting conditions (What is to be changed?).
- Ends-in-view (Changed to what end?).
- Processes and techniques (How can change be accomplished?).

Broad regional *strategies* for systems transformation express directions for the process through which the region may be brought to move from its starting conditions (as they are conceived in a particular instance) to particular ends-in-view. Characteristically, such a process proceeds in *stages* of:

- Diagnosis (getting started, casing the region).
- Involvement (engaging these individuals and agencies whose interaction is taken to be critical).
- Planning and goal-clarification (discovering feasible processes and choosing and testing specific ends-in-view).

These stages are apt to be cyclical rather than sequential. The passage from diagnosis through implementation leads to a revised picture of starting conditions, and through the cycle again. Because several streams of activity often proceed concurrently, the region may at a given time engage simultaneously in all stages. As the region moves through stages of systems transformation, in its *developmental cycle*, it may extend the scope and depth of the issues it tackles.

Given this skeletal view of systems transformation, an evaluative process oriented toward learning must take the form of a dialogue—a continuing process of inquiry in which the regional co-ordinator\* and RMPs both raise and respond to questions. This is for several reasons. Given the open-endedness of “systems transformation,” prospective systems rationales for RMP must be inferred from (rather than imposed on) regional activities. Systems rationales and systems activities must modify one another. The evaluative process must detect discrepancies between systems rationales and discovered systems, and tactics for responding to those discrepancies. Moreover, project and program goals shift over time. That is often a sign of progress; and the evaluative process should help to discover whether it is, and in appropriate cases both reflect and encourage it.

In what follows, we list *guidelines* for the kinds of questions to be raised in such a dialogue; *criteria* for systems transformation, from which these questions flow; and, in some instances, *illustrations* of response.

One test of the dialogue is that both co-ordinator and “central” become able, on the basis of it, to form continuing, grounded judgments of regional program performance. A second test is that as a by-product of the dialogue the co-ordinator becomes more proficient at designing and carrying out the process of systems transformation. A third is that the national staff is enabled to formulate progressively more adequate “systems rationales” for RMP.

The dialogue follows what we have identified as the three main elements of systems transformation and, within them, the stages of development.

1. *Starting conditions. The co-ordinator should be capable of articulating a regional diagnosis which holds water, and which provides the basis for the formulation of directions of systems transformation.*

The subject has to be probed to the point that both participants in the dialogue are convinced that:

- The evaluator understands the spokesman’s view of the region and has stated enough of it clearly enough to reassure himself and the spokesman.
- The spokesman has stated whether he believes this particular array of starting conditions is tough, average, or a bit simpler to deal with than average (assuming for the moment the accuracy of what the spokesman has said).

---

\* We will use “regional co-ordinator” as shorthand for those agents involved in formulating and carrying out RMP strategy at the regional level.

- All likely emphases have been tried out by the evaluator in an effort to test and understand how the starting conditions fit together dynamically.

An adequate response constitutes a diagnosis of the regional health care system. It also furnishes the evaluator with some beginning hypotheses about how skillful the regional core staff is in casing the region.

When well explored and laid out in the dialogues, the diagnosis includes the data crucial to working out strategies of systems transformation, both those which define health issues and health needs and those which define the organizational and political character of the health care system.

- What is the character of the principal health problems of the region? What is their distribution?
- What is the character of the present configuration of health care facilities and resources? What is the nature of the health care delivery systems that are dominant in the region?
- What are the patterns of access to care among the principal population groups?

The foregoing questions aim at establishing starting conditions at the level of health, access to care, and configuration of care-providing resources.

- Who are the key actors and powers within the health care system of the region? How do they relate to the power structure and to the politics of the region as a whole?
- What is the nature of the linkages, the relationships, the patterns of referral, the tensions and conflicts, among these key actors?
- What do the central actors perceive as the major issues of health care for the region—whether these are identified in disease-specific terms, in terms of access to care, quality of care, or in terms of costs, manpower, patterns of dominance and distribution, or other facets of the health care system?

This next set of questions aims at an understanding of the “political” forces that can be used or that must be dealt with in any strategy for systems transformation.

Out of responses to these questions come regional diagnoses which provide the material for designing strategies of systems transformation for the region.

At the point of establishing agreement on starting conditions, the evaluative dialogue has to involve:

- Feed-back to a widening circle.
- Testing the perceptions of those who first describe starting conditions, strategies, or other aspects of RMP and the territory in which it functions.

- Some appraisal (i.e., development of a more or less acceptable description) of the way the local RMP went about data selection and gathering.
- Gradual clarification, through the dialogue itself, of the specifics on which detailed information is needed.

The following are excerpts from regional diagnoses which illustrate something of the variety of starting conditions to be discovered.

*X Region.* X is a prosperous, relatively homogenous society. Good medicine is practiced here, and the profession is in relatively good repute with the local political-social establishment. As yet medicine and the other health professions are facing only tentative questions about the "relevance" of where sub-specialization and bigger-better hospitals take us. But something very real is brewing in the state legislature's effort to force a "Family Practice" Department on the distinguished specialists of the University medical faculty. Additional intimations exist in the reluctance and opposition of the Academy of General Practice to the way the medical faculty had first planned to go about teaching family medicine.

Layer on layer of competent, skilled, devoted people working in hospitals and other health care institutions all over the state, all of which tend to emulate or somehow react or respond to the presence of the internationally famous institutions: the Central Clinic, the University, and Rehabilitation Foundation. There is an apparent shortage of manpower willing and able and wanting to perform health care services on the level of ordinary care for ordinary conditions. Town-gown issues are real, but because "gown" somehow includes Central City as well as "The U," and because "everybody" was trained at "The U," the issues take a special form. Centralization of the Clinic and decentralization of the University complicates their association, whenever joint commitments are required or contemplated. Good acute care general hospitals are a dime a dozen, and coming to view one another as competitive whether they are or not. Many are trying to become referral centers both in big specialist consulting staffs and many high technology services.

Generally the establishment, medical and non-medical, exhibits a tough-minded, "show me" conservatism, tempered by a very active consensus and willingness to try out credible ways of improving the situation (e.g., 40% of X-State private physicians have tried out group practice. They and their patients like it well enough to stick with it.)

RMP has to make its way among a number of giants, all zealous defenders of quality medical care, each with its own tradition of constructive innovation, each with its own considerable institutional inertia and sense of independence.

*Y Region.* In the region's largest city there is one large medical school and one large community hospital. The region consists of five quite different counties. Three counties made common cause with RMP from the outset. Two are left. In one, a private physician has his own comprehensive health plan; prepaid medical care has been attempted under his auspices; success is believed to be uncertain; critics prophesy failure. The other county is simply cut off and disinterested. It is difficult to get medical or consumer representatives from either county even to meet for reasons that pre-date RMP, but embrace it: several of the major counties are joined in uneasy alliance, with many rivalries, all felt particularly strongly in the smaller cities.

*Z Region.* The major hospitals and associated medical schools are all in the major city and dominate the region. These are set against the smaller community hospitals, each of which in turn is trying to be a medical center. Not surprisingly, there is relatively thin patient use of these expensive facilities in suburban hospitals. Not surprisingly, too, there are parochial and compartmentalized referral patterns disturbed by conflicts among the several large medical schools and hospitals. There tend to be economic and social distinctions drawn between the largest and the other medical school complexes, though these may be decreasing, and certainly keep changing. With all, the distribution of physicians to patients is highly inequitably spread over the region.

- ghetto areas: 1/3000 to 1/5000
- center city: 1/200
- suburban: 1/700 to 1/800
- rural: 1/1000 to 1/2000

The 5 medical centers have limited goals. All are under great financial pressure, pressure relative to income, to student load, and pressure to pay attention to the ghettos. They are beginning to believe that is where the money is. In the meantime, the cultural institutions of the major urban center continue to tend to turn inward, there is very little that can happen "unless you own it." So the tendency is rather stronger than average to want to turn RMP and training dollars to the enhancement of existing institutions and departments.

Rivalry conditions all attempts to regionalize or otherwise bring about constructive associations between people in the somewhat depressed cities of the North and the rich primary city.

2. *Preliminary strategies.* – Proponents of the regional diagnosis should be capable of meeting challenges as to the accuracy or relevance of their analysis. But the analysis need be neither exhaustive nor entirely accurate. It is of greater importance that it be capable of shifting in response to challenge and that there be, in the inquiry undertaken by the co-ordinator, a continual source of challenge to be met. In particular, it is important that judgments about major issues of health need, quality of care and access to care, facilities, manpower, cost of care, and the political and organizational structure of the health care system, all be subject to the continual test of the multiple perspectives of key actors in the health care system. Where important conflicts of perspective arise, they should be confronted explicitly and actively. Where they cannot be resolved, these conflicts of view themselves become issues for continuing work and inquiry.

*Based on the regional diagnosis, the co-ordinator should have formulated preliminary directions of strategy which reflect defensible judgments about crucial substantive issues of health care, issues relating to the political and organizational structure of the health care system, and key actors and initiators of innovation in the health care system.*

While the co-ordinator should be capable of arguing for these directions of movement, on the basis of the regional diagnosis, these preliminary views about strategy should remain developmental, in two senses. They should take account of the issues they do not address, and there should be some thought as to the means by which these other issues may come to be addressed. And they should be responsive to changes in the regional diagnosis which come to light in the course of RMP activity.

The basic question is “*How have you gone about formulating preliminary strategies for systems transformation?*”

- Through what process have you gone?
- What is the substance of the strategy as so far developed?
- Why this far, and no further – or why so far in this direction?

Often, the best way of getting at these issues in the dialogue is through questions such as these:

- Where are the outstanding strengths and weaknesses among key agencies and actors in the medical care system?
- What are the patterns of alliance and conflict, and how are these changing?
- For key actors in the system, and for the issues they regard as critical, what are the ends-in-view

both for changes in the delivery system and for changes in their own position within the system?

- What are the critical “starting issues,” and how might these be used to move toward systems transformation?

But the specific forms of these questions must come from the regional diagnoses, and must elicit the ways in which preliminary strategies address themselves, or fail to address themselves, to the issues raised in these diagnoses.

The following are examples of some of the preliminary strategies emergent from the fragments of diagnoses listed above, and questions that the evaluator can or should raise about these strategies, to push the dialogue a step further:

### *X Region*

The primary problem is the isolation of many small communities, especially rural communities from which physicians are slowly disappearing, and their disinclination to collaborate. Corollary to and underlying this is the past success of medical education in selecting and training physicians to want to work in sophisticated hospital settings, thus creating strong impetus for hospitals to compete, even within communities, and to attract physicians by offering ever more highly differentiated and costly services, without careful, credible investigation of community needs and how they are satisfied.

The function of RMP should be (and is) through projects, membership on advisory committees, and core-staff activity to facilitate connections and collaborations among elements of the medical care system, particularly among small communities and particularly among physicians. The connections and collaborations should be multiple and small-scale, so as not to ruffle too many feathers.

So RMP, for example, should serve as broker and supplier of seed money for the merger of hospitals in adjoining rural market towns; should support short-term in-residence programs for GPs at the Clinic; should dot coronary care programs around the State; should promote outreach programs from the Clinic and the University; should use the RAG and its committees to involve all elements of the medical care system and representatives of its consumers, in order to connect small communities with one another and with the centers.

The object is to build larger movements toward collaboration and more ambitious ends-in-view from the success and the fallout from many small-scale efforts, in



the process of learning what is feasible and helping the various interests and groups involved to assume as constructive leadership roles as possible.

Some questions:

- Will the small-scale collaborations ever get big enough to make an impact on medical care in X Region, and will they happen so slowly that one is forgotten before the next happens? What is the threshold level of scale and pace for facilitation if it is to have a building effect?
- Have you taken into account what needs to happen in order to get the Clinic and the University really involved in the medical problems of the smaller communities? How much "involvement" do you want and why? Can you do that without confronting the "family practice" issue, and helping to attain a viable resolution to the conflict among the Academy of General Practice, the University medical faculty department heads, and the legislature? Would sponsoring more activity within the allied health manpower field force or encourage a more valid solution to the general practice-family practice problem — or just convince the MD's that RMP is against doctors?
- How do you propose to respond to the conservative stand of many GPs, particularly in southern areas who don't see what RMP has in it for them, and who feel threatened by or disagree with what they hear?
- What stance will you take toward groups currently left out of the strategy — for example, hospital administrators, dentists, mental health practitioners? Are there parts of the State in which it would make sense to do so?
- Does the current mix of efforts respond, at the level required, to the serious problems you have identified — i.e., to the problems of rural medicine, isolated communities, care for the small but clustered populations of minorities, the deficiencies associated with the (otherwise desirable) proliferation of specialist physicians and the disappearance of family physicians, both in the central parts of the large cities and in rural areas? If you cannot envisage any adequate response in first-round activities, how do you plan to build toward such a response? If manpower shortages seem to you the central questions about the response, how do you plan to attack the question of manpower over time?

Often the formulation of preliminary strategies depends upon the *involvement* of key actors and agencies.

*The co-ordinator should have found ways of including actors and elements of the region's medical care system identified as key in the regional diagnosis; where some of these cannot be included at the outset, the problems about their inclusion should be explicitly confronted and strategies developed for overcoming these problems over time.*

"Inclusion" may be indicated by participation in a range of RMP-related activities, including involvement in RMP committees, in project work, or in ventures initiated or supported by RMP. The difference between significant and *pro forma* inclusion must be resolved by tests that vary from case to case.

What is to be appraised includes:

- Whether there has (or has not) been a real attempt to arrange for specific people to be included in RMP. (Was the labor union representative really invited to RAG meetings? Did he feel invited? Was there anything for him to do?)
- How well the attempt is related to the co-ordinator's sense of starting conditions and his strategy and objectives (which depends on having learned those things first).
- How explicit the co-ordinator can be about who is not to be included, and under what circumstances those persons would or should be included.
- How much the co-ordinator and core staff learns about the process of including people from the experience of doing it. (If they had it to do over would they do it another way? Are they increasingly imaginative and increasingly direct in their approaches to people?)
- The impact on others of the co-ordinator's attempts at including people (clumsy or skilled, relevant or irrelevant, useful or useless, well planned and well understood or otherwise).

A case in point is the following:

### *Y Region*

The RMP has taken the position that it is a clearing-house for projects; it solicits and processes applications from elements all over the region. RMP is, therefore, a conglomerate of projects; how can it have a program strategy for systems transformation or anything else?

But there is the sense of need to involve the two counties currently disengaged from the program. The preliminary strategy has impacted on the starting conditions in a way that permits, encourages, and partly specifies a revision in approach.

One county, medically under the leadership of a strong physician, has no involvement in the RMP program. And there are 250,000 people there. The belief in the county is that the big city always wins, and that's where the money is.

In spite of its apparent role as a "clearinghouse for projects," the RMP turns out to be operating on a strategy which says, "Get every major actor and every county active in RMP." Their tactics are based on this strategy.

The major physician in the isolated county is concerned about diagnosis of cancer, and about the 100-mile round trip required to get specialized diagnostic screening in the large city. He is encouraged, therefore, to propose the establishment of a diagnostic center in his county.

Some of the relevant questions, especially appropriate to early involvement phases:

1. Is the investment worth it? How much does it take to "purchase" involvement? as a percentage of the overall budget? compared to the costs of confronting other urgent health care issues? Are there other excluded or isolated elements of equal importance (geographical areas, professions, voluntary associations, health departments, medical societies, hospitals, or a combination)? What are the potential future consequences (enmity, retribution, etc.) of failing to try to involve somebody now? How does an effort to include Dr. H. relate to the regional diagnosis?

2. What are the signs that investment has been successful in involving Dr. H and his county? How do you distinguish *pro forma* from significant involvement? For example, visibility at RMP meetings? Attitudes of Dr. H. toward the proposals of others? Willingness to permit some "teaching days" in the area? Other projects coming out of the county? Willingness of Dr. H and others in the county to lend voices in support of RMP activities? Willingness of Dr. H. to share his emergent strategies for development of medical care system in his county, or to participate with others in formulating such strategies?

3. *Ends-in-View.* – *Out of interactions of key actors, ends-in-view should have been established. These must confront at least some of the key issues earlier identified as crucial in the region. On the level of substantive health care, they must confront at least some of the constant health problem themes, or emergent issues in health care.*

At a zone in time, attention shifts from the problem of "getting all the key actors active in RMP" to the problem of formulating the more specific ends-in-view and the strategies for achieving them which are to emerge

from the interaction, planning, bargaining and negotiating of the key actors.

These ends-in-view are the specific rearrangements sought in systems transformation. They, too, have many qualities that are subject to evaluation. The emphasis, again, is first to discover what attempt has been made to identify these qualities, and to deal with them. Evaluation of specific content makes sense only after its clear and more or less agreed what has been attempted, and the context for attempting it.

The following are examples of appropriate questions:

- Have the issues earlier identified as crucial in the region found their way into the formulation of ends-in-view?

This is an illustration of what such a list of issues might look like:

- Guidance to get people into the health professions.
- Coordination and involvement of the voluntary agencies.
- The urgent need for dental care in the north.
- The lack of out-patient care centers except for emergency rooms.
- Essentially no preventive medicine is done in the State.
- Too many community hospitals trying to become medical centers.
- There is no weekend and almost no night-time medical coverage now in a major rural county area."

Is the RMP engaging *some* of these issues through the deliberations and interactions stimulated among elements of the health care system? "Engaging" means, here, facilitating the formulation of ends-in-view and strategies adapted to them.

- Certain general criteria cut across regions and across possible activities within regions. Questions about "relevance" of particular activities apply not only to the match between ends-in-view and judgments about issues, but to the need for some attention to these criteria.
  - Costs of care, particularly for hospitalization, extended care, and costs as experienced by lower- and lower-middle income persons as well as others.
  - Quality of care, and the distribution of quality of care across the region.
  - Access to care, and equity of access to care, across socio-economic strata, minority and majority groups, and geographic subregions.
- Have the processes making for inclusion, discussed earlier, extended beyond formal membership in

RMP activities, to formulation of ends-in-view and strategies for achieving them?

- How are priorities formulated? Are priority issues being confronted explicitly at all? By whom? Do priority considerations enter explicitly into the deliberations and interactions of elements of the medical care system, or are they handled by the coordinator or core staff alone, or ostensibly or really left to Washington? If there are conflicts among elements judged to be crucial to the region – for example, conflicts between major hospitals and medical schools, between town and gown, between professional providers and representatives of users – are these conflicts allowed and encouraged to enter into the formulation of priorities? Does the coordinator intend to attempt to build clusters of these elements into working groups, through explicit confrontation of these questions? If he is not doing this, is it a matter of deliberate intent? Is he working – temporarily, or as a matter of continuing strategy – on a model of compartmentalization, in which conflicts over priorities and ends-in-view are not allowed to come up, except within limited subsets of elements? Is he “sub-regionalizing” in this sense? If so, does it make sense to do so?

Is conflict of ends-in-view being handled as a matter of “dividing up the pie” among competing actors, or is there also an attempt to relate such judgements to shared judgements about the urgency of health issues, or about the usefulness of issues as ways into systems transformation in the region?

- *Major themes of RMP activity should be developed and stated. These should be not merely a reflection of what is common to ongoing activities, but a source of guidance for the generation of new activities. Questions of priorities among ends-in-view should have been confronted, through a process in which key actors in the region work on their conflicting interests not only on the level of ownership of RMP resources but on the level of substantive health issues and strategies.*
- How appropriate, acceptable and feasible are the strategies being developed for achieving the ends-in-view adopted? For example,

---

\*This may be the first time that themes of RMP activity become explicit and that questions of priorities become real issues (often first stimulated by conflicts over ownership of limited funds).

– An outreach center, as a way of involving a major hospital and medical school in the problems of an adjacent ghetto? Who will make it work? Who wants it?

– A joint coronary care project as a way of encouraging collaboration and rationalization of planning among a set of community hospitals? What will make it transcend its original focus?

Questions about such strategies will focus on a number of dimensions:

- Adequacy of scale of the “solution” to the problem.
- Feasibility of the methods proposed.
- Appropriateness of the strategy to objectives on multiple levels of the activity (e.g., substantive health impact, as well as systems transformation ends-in-view; clarification of ends-in-view as well as involvement).
- Appropriateness of the strategy to the constraints and problems perceived to be underlying the issue. One of the questions to arise at this point is the question of “teeth.” Is the issue one that will yield best, or at all, to voluntary involvement on the part of the key actors concerned? Or does it require some forms of sanction and compulsion? This is a question of ideology, strategy and legislative mandate for RMP, as well as of propriety: possibly some other agency is more appropriate.

Where the focus is on learning, attention will go not only to questions of this kind but to questions about the ways in which the development of strategies is handled:

- Is there evidence of the active consideration of alternative ways of achieving the same ends-in-view?
- Does the deliberation over strategies carry with it consideration of effectiveness of the strategy in relation to the costs of carrying it out, and consideration of the cost/effectiveness characteristics of alternative strategies?
- Are there timetables for accomplishment? How realistic are they?
- Has there been consideration of ways of determining over time how effective strategies are in achieving ends-in-view? Tests for their achievement?

Where the focus is successfully placed on learning, the impact of such questions will not be to “grade” the strategies at this zone in time where emphasis is on the development of specific ends-in-view, but to influence their development positively, by “accelerating” and “enriching.”

4. *Implementation.* – *The process of implementation should be characterized by involvement of implementers in selection of ends-in-view and strategies for achieving them; and by a relationship of co-ordinator or core staff to implementers which permits continuing mutual modification of strategy and end-in-view and of implementing activity.*

The implementation of strategies toward ends-in-view may take the form of core staff activity, of the conduct of specific RMP projects, or of the activities of committees or *ad hoc* groups, under the aegis of RMP. The end-in-view and the strategy may be specific enough to lend themselves to only one of these kinds of activity, and to a well-defined unit of implementation, or they may lend themselves to a widespread cluster of activities.

For example,

<i>End-in-view</i>	<i>Implementation</i>
To foster collaboration and rationalization of planning among 13 community hospitals.	A coronary care project jointly granted to the 13 hospitals, requiring the use of common facilities.
To encourage multi-level collaboration between two hospitals in adjacent rural communities.	Brokerage functions by core staff, RMP support of one hospital staff member charged with working out details of the merger.
To increase the "power base" of the medical community "on the other side of the mountain."	A series of projects, funded in that area, linked to major medical institutions. Brokerage activities. Use of RMP committees to establish relationships crossing the mountains.

Some of the relevant questions are these:

- Are initiators and leaders of the activity aware of the ends-in-view, and the processes leading up to their formulation, on the basis of which the activity actually came to be undertaken by RMP?
- What are the patterns of access to resources required for implementation? Is there a basis for judgments to be made, on a continuing basis, as to the adequacy of resources to the task?
- Is attention given to the possibility of shifting definitions of ends-in-view as more of the reality of the discovered system comes to light? Is the project or activity leader locked into a potentially stultifying view of what constitutes "success"?
- What constitutes progress? Are there operational tests of performance, short of more nearly final

judgments of impact, which can help to guide performance in the course of the activity?

- What is the relation of the regional co-ordinator and his staff to the activity? If it is not their activity, do they have, in relation to it, a continuing monitoring, learning-evaluative contact which allows mutual modification of the ends-in-view and the strategies by which the attempt at implementation is being made?
- How compartmentalized is the activity? Is it connected to analogous activities in the region, or to activities which are parts of the same program strategy, so that both learning and concerted action may occur, where appropriate?
- What is the relationship of these processes of implementation to the overall strategies of systems change held by the coordinator and/or his collaborators? Has the coordinator attempted to be explicit about these? Is there an effort to relate them to particular strategies for achieving particular ends-in-view? For example, to connect a particular activity as a feature of a "master plan"; to identify a particular negotiation as part of an overall strategy which seeks to involve key actors in a process of negotiation over their interests and conflicts in relation to the system of medical care. Is the coordinator able to use the experience of particular activities to learn from or to influence his overall strategies of systems change?

There is one side of the question of impact which should be treated separately here, because it involves the impact of the *process* of implementation, which can reflect back both on the formulation of particular ends-in-view and on the region's capabilities for carrying out further systems transformation activities. This is the process through which the definition of accepted ends-in-view may shift.

- The connections established and reinforced in a particular activity may lay the groundwork for new forms of collaboration, e.g., the joint planning of a coronary care unit which leads to joint planning of a range of common facilities; the diagnostic screening project in a county previously cut off from the medical system of the region, which leads to a series of boundary-crossings. Are these things happening? Are there attempts to make them happen?
- Learning from an implementation process can lead to changes which facilitate new processes, e.g., the cumbersomeness of a process of review and monitoring can lead to simplifications which make

it easier and more attractive for others to enter the orbit of RMP activity.

- Processes of implementation can display or enable development of "role models" which influence the character of new activities undertaken, e.g., the impact of Jim Musser as broker-facilitator on other key actors in the North Carolina region, or of Paul Ward in California, e.g., the influence of the few emerging medical care corporations in California on similar, varying approaches to medical corporations.

Questions about impact of implementation, then, need also to be addressed to the impact of the process of implementation itself.

At this point, RMPS criteria for systems transformation in the region take the form of meta-criteria for the evaluation processes carried out in the region.

- Without specifying evaluative criteria to be used in assessing the impact of implementation on any of the levels of change, RMPS should require that such criteria be developed and that they be appropriate to the ends-in-view and strategies adopted.
- These criteria should not be limited to programmatic criteria (e.g., how many nurses trained? how many calls received?) but should attempt to assess change at one or more of the several levels of change in substantive health care.
- In each instance, consideration should have been given to the choice of level at which change is assessed, aiming at health outcomes, then at access to delivered care, and so on. There should have been review of the definitions, test-methods, and measures appropriate to the end-in-view and strategy involved.
- With respect to the process of evaluation, the evaluative framework should have been developed collaboratively between the regional center and the implementing agency. There should be an openness to modification, through the process of evaluation, both of the implementing activity and of the original choice of end-in-view and strategy. This openness should be evidenced in the demonstrated capacity of evaluative activity to influence the planning of the implementing process, and in the evolution of the concept of end-in-view and strategy during the course of implementation; and the frequency and pattern of contact between core staff and implementing agency should be such as to make that kind of mutual influence feasible.

- The evaluative processes adopted by co-ordinator and core staff should be conducive to learning across sub-regional boundaries, so that those engaged in analogous activities (continuing education for GP's, for example) can learn from one another's experience, and those whose activities are elements of a larger strategy can interact in the light of that strategy.

5. *The Developmental Cycle.* Regional programs develop iteratively, if at all. Cycle succeeds cycle, each growing out of, but resembling, its predecessor. A regional program, seen as systems transformation, moves through its cycle: casing the region, planning and implementing. Then through another cycle widening and deepening its rings of activity. The evaluative questions of any one phase continue to be relevant; only, new sets of questions are also relevant to established activities, and to other sets of activities. The process of bringing new elements into RMP, for example, continues even as the ends-in-view emerging from earlier processes of inclusion begin to be carried out.

The most relevant new questions help uncover the directions of change in the scope and purchase of the whole program as it moves through successive interactions of the process. These questions are of several kinds:

- Is the process increasing its scope?
  - Is it increasing in the overall volume of activity, as measured by actors involved, dollars mobilized, number of separate activities undertaken?
  - Is there a widening range of parties involved in interaction and negotiation? Is the level of aggregation of the parties increasing? For example, is the interaction beginning to involve clusters of community hospitals rather than individual community hospitals? Is the level of aggregation also decreasing? For example, are individual physicians as well as medical society representatives coming to be actively involved in a way that extends the scope of the program?
  - Is there an increase in the number of health issues engaged? Is there an increase in the coverage of the region represented by those issues and by the ends-in-view and activities generated? Within each phase, the map of the issues confronted and their location in the

region should reveal changes of the following kind:

		Regional Location									
Issues	X								X		
			X								

Phase 1

		Regional Location									
Issues	X	X									X
	X	X									
			X	X					X		
	X	X	X							X	

Phase 2

- Is the process increasing in depth and intensity?
  - Is there an increase over time in the perceived importance, urgency, and ambition of the issues engaged and the ends-in-view formulated?
  - Is there an increase in the connectedness and “clout” brought to bear on the issues engaged?
  - Is the level of aggregation of the parties decreasing? Are individual physicians as well as medical society representatives coming to be involved in a way that deepens the program?

We can provide an example of the development of ends-in-view and strategies in a regional program as it begins to go through a succession of cycles:

### The K Region

Dr. P., the coordinator, came from a program of continuing education in the one large medical school, a program of continuing education for GPs which, by his own present view, was not too successful. He began by seeing the creation of RMP as an opportunity to expand his own educational program, and obtained a planning grant to create K-RMP. He visited local medical societies over the region and with them set up a program around tumor registry, coronary care units, and continuing education. Boundaries of the region were set up by the

expression of interest of the parties approached who attended the meeting.

As the program has begun to expand, its emphasis has shifted away from the categorical approach. The RAG, which began with 30 physicians, has begun to change composition to include laymen. In view of the relative weakness of other institutions, including the State Health Department, KRMP has moved toward a controlling position for health planning for the State.

Concentration at the beginning has been on work with individual physicians and community hospitals, with an emphasis on education, viewed as the easiest and least threatening way in. At the same time, core staff became involved in project-writing for individual hospitals, KRMP has now withdrawn from CCU programs, except for continuing education. However, a similar effort based on the earlier experience (establishing facilities, loaning equipment to communities who could not afford to buy it) is now being carried out for respiratory programs.

Dr. P. now realizes that in his region, which is poor in physicians and clear in its referral patterns and which has one medical school and not much institutional rivalry, the provision of continuing education to physicians and others is not enough. What is needed is the provision of a system of care and appropriate facilities within which the fruits of education can be realized.

Here, since the structure of the program as a whole is built around the coordinator, the development of ends-in-view becomes very much the development of his own views of the issues that need to be confronted and the ends-in-view adopted. Is the process characterized by an evolution of issues, ends-in-view and strategies, which reflects learning?

The regional diagnosis of the coordinator, the issues he takes to be important, the ends-in-view and strategies to which he is committed – in short, his own systems rationale – may shift in response to new perceptions of the discovered system of the region, as regional activities bring that system into focus.

This learning may take the form of an explosion of “rational” plans for the building of the health care system, by contact with the political interests and powers of the real-world actors in the system. It may take the form of a shift in priorities about health issues, as previously “hidden issues” – for example, the depth of inadequacy of health care in ghettos – come to the surface. It may take the form of perceiving the extent to which the needs of physicians and community hospitals in “have not” areas are inadequately served by diffusion

of the technologies and research findings generated at the major medical center.

In each instance, the discrepancies between systems rationale and discovered system, at the regional level, may lead to the reformulation of regional diagnosis as well as of ends-in-view and the strategies corresponding to them.

It is not reasonable to set uniform standards for the periods of time within which regions should have reached certain levels of maturity in their developmental cycles, just as it is not reasonable to apply uniform standards across regions to the time periods within which the various stages of development should be completed. On both levels, the time intervals will vary with regional conditions. The key factors here are not so much the size of the region as its complexity, its internal connectedness or disconnectedness, the number of conflicting or disconnected elements within it, and the seriousness of their conflicts or isolation from one another.

Elements that affect the speed of motion include:

- simplicity of the politics of the medical care system. Few elements to be connected; few conflicts to be resolved.
- relative weakness of other elements of the system, permitting RMP to function from the beginning in dominant or unusually significant health planning role.
- relatively high degree of connectedness among elements of the medical care system.

It may be possible to establish a typology of RMP regions in terms of their potential for movement, similarities in strategy, and characteristic types of activities chosen to carry out the RMP program. There are, for example, many instances of efforts to stimulate collaboration among community hospitals through their joint involvement in some program of approach to categorical disease; to establish outreach arms of major medical centers; to reach isolated subregions through programs using paraprofessionals, continuing education, and the secondary support of specialists. Regions and subregions differ as to the constraints they put in the way of these kinds of activity, but they, too, can be grouped in terms of the seriousness of those constraints.

The purpose of such a typology would not be so much to permit judgements of the effectiveness of one region against another as to provide guidelines both for RMPS and for regional coordinators as to the rates of movement it is reasonable to expect in a given region and for a given kind of activity.

*Judgements about a region's progress in systems transformation may be made on the basis of its ability to meet criteria within any given stage of development; its rate of movement from stage to stage, given the constraints under which it is operating; and the level of scope, depth and learning evidenced by its overall cycle of development.*

In point of fact, most of the RMP regions are still primarily involved in the problems of inclusion of key elements of the medical care system in RMP activity and on the formulation of preliminary directions of movement and strategies. In spite of the number of operational projects, most regions are only beginning the work of fitting projects into strategies for achieving specific ends-in-view. Most are only now at the stage where the formulation of themes of RMP activity and the confrontation of questions of priority among ends-in-view become feasible tasks.

#### **Conditions for the Central-Regional Dialogue**

Having sketched out a national-regional dialogue aimed at fostering learning in relation to systems transformation, there remain questions about the particular vehicles through which such a dialogue may be brought to reality and the conditions under which it can be effective.

- The two parties to the dialogue must begin with some commitment to and understanding of the goals and methods of this kind of evaluative process. The requirements here relate both to the theory of the evaluative process and the role of the dialogue within it, and to the particular skills and techniques involved in carrying it out.

- Although we have used simple words like "central" or "RMPS" and "coordinator," the parties to the dialogues will be complex. On the regional side, the dialogue will be carried on by groups of varying kinds, depending on the makeup of those involved in carrying initiative at the regional level. In one region, it may be a "strong man coordinator," his key assistants, and from time to time others that he may wish to bring along in order to involve or educate them. In another region, it may be the team the coordinator has been trying to assemble out of core staff, certain RAG members, and certain key actors in the medical care system of the region.

- On the side of the national staff, there is a key requirement for continuity of involvement in the dialogue with region over long periods of time — ideally,

over the life of the region's development under RMP. The requirement for continuity becomes particularly critical, given the diversity and open-endedness of regional approaches to systems transformation; it is only out of intimate knowledge of the content of earlier stages of development that central can be effective in dialogue with the region.

But, given the realities of life in both central and regional bureaucracies, continuity of this kind is to be achieved not through one man but through small groups whose members overlap in the course of time.

From central's point of view, the small group permits the inclusion of the varieties of competence required to carry out effective dialogue with the region — competence to question and respond on issues of substantive medical care and on issues of systems transformation, and skills in the evaluative process of the dialogue itself.

There will be no need to distinguish the central-regional dialogue from funding decisions, and, concurrently, to move away from the usual mode of central-regional contact, in which the region displays its wares for central and central and the region then engage in a game of attack and defense. For the central-regional relation to be solely or primarily in this mode prohibits learning, in the senses outlined above, and makes it difficult or impossible for central even to gain information about regional activities.

On the other hand, the dialogue requires that the RMPS staff be capable of being tough with the region, raising issues hard enough to be heard and challenging the region in the light of findings and commitments which emerge from the dialogue over time.

In order to make these things feasible, there is first a need to model the roles involved and to set the tone for such a dialogue, and concurrently to set apart and formally distinguish the funding-justification process from the central-regional dialogue. The dialogue will surely feed into RMPS judgments about regional funding, but should be formally and operationally separate from the funding process.

Will such a distinction be feasible, given the tendency of the region to view central as monolithic and the region's knowledge that funding decisions will be made by central? This problem is comparable to the problem of the regional evaluator in establishing his "helping" role, in spite of the fact that his findings will be influential for decisions on project funding; indeed, the problem is central to any process of good management in which the manager seeks both to facilitate learning and to exercise control. The feasibility of the effort will depend ultimately on the good faith that central and the

region are able to establish with one another, and on the extent to which the dialogue is *found* to facilitate learning.

The dialogue requires a certain frequency of contact between central and regional groups. Given the rate of movement in most regions, once a year is not often enough. Within the interval of a year, too much happens, and too many decisions are made which lock the region into patterns of activity. Frequency of contact should be determined by the time required for the coordinator to take significant steps, or for the regional situation to shift in significant ways that mark important milestones in the stages of systems transformation. Intervals are likely to vary over the course of the region's cycle of development. For example, contacts might be established around key events such as the first formulation of regional diagnosis, the establishment of themes of RMP activities and the first effort at establishing priorities for specific ends-in-view, or the first phase of experience in implementing a specific strategy. Within the range of frequency indicated by "oftener than once a year," there should be provision for flexibility increases if a representative of central and the regional coordinator can maintain contact during intervals between meetings of central and regional groups.

The central-regional dialogue offers another perspective on the role and conduct of regional site visits, and on the proposed process of anniversary review.

The central-regional dialogue could become the main function of the site visit. The site visit team would then become central's party to the dialogue. Such a concept would answer some of the problems currently reflected in regional and central reactions to the conduct of site visits — for example, the pattern of regional display and of attack-and-defense which make it difficult or impossible to find out what is really happening in the region; lack of continuity in the site visit team; lack of feed-back to the region; inability of the site visit team to respond to the region by clarifying or modifying central's "signals." There are also significant potentials of the site visit as a vehicle which the central-regional dialogue may help to tap: the opportunity for on-site contact with regional actors and agencies, and the presence in the region of persons regarded as peers by many of those undertaking regional activities.

There is the further issue of the manpower requirements RMPS would experience if it took seriously the conduct of central-regional dialogues with all of its regions. The site visit team concept, in which outsiders are mobilized alongside central personnel, would provide a crucial extension of central staff. But the concept



would also require intensive efforts at internal training and team-building for the site visit teams.

With respect to Anniversary Review, that event would have a very different significance if it were to function as the yearly culmination of central-regional dialogue, rather than as an isolated contact which will tend to be

seen, whatever the intent, as a funding-justification process. The site visit team would then come to play a critical role in the anniversary review process, and the results of earlier phases of the central-regional dialogue would then provide the basis for the inquiry conducted and the judgments made in the course of anniversary review.

## HOW OTHERS SEE REGIONAL MEDICAL PROGRAMS AND EVALUATION

ALEXANDER M. SCHMIDT, M.D.  
Dean, Abraham Lincoln School  
of Medicine,  
University of Illinois

Having come in late, I was sitting in the back of the room, rather than here on the platform; and I am very pleased to have been able to hear the elegant discussion on Regional Medical Programs and systems change by Dr. Schon. I was late arriving this morning because our three upper classes are returning this morning, and I met with them - about 625 strong - for a re-orientation session. This is something new for us. Change now is so great, and the rate of change is so rapid that we are not only orienting our incoming freshman class of 225 students, but are re-orienting students who have been on vacation. The need for such sessions was made evident by their questions, I thought as I was driving to the meeting. Among the questions asked were:

"Is Cook County Hospital still alive and well?"

"How many medical schools are there in Illinois today?"

And finally, "How many people have you added to the university police force?"

And my answers were respectively:

"Not very."

"Ten."

and "Plenty."

I was also musing that it was only a couple of years ago that I gave a talk entitled, "Is Evaluation a Dirty Word?" The response from the audience then clearly indicated that they thought it was.

In the ensuing years, however, it has become apparent that the word "evaluation," like some other words we are now hearing almost daily, has had the shock value worn off, as more and more people have used the word in open public.

It is really too bad that evaluation got off to a rather shaky start in Regional Medical Programs. From time to time I have tried to figure out just why it happened. Certainly from the viewpoint of the administrator (who hopefully is a good manager) evaluation is a very powerful friend. Evaluation ranks along with cost accounting and program budgeting (two other dirty words), as one of the most powerful management tools we have. We all probably know this, and believe in at least the theory, yet our response to the word is too often less than

favorable. It has occurred to me there are three principal reasons for our aversion to the subject of evaluation.

First, there is the general feeling, expressed over and over to me, that "seat of the pants flying," if it gets you there, can't really be all that bad. Over the past decade, through trial and error, in both education and health service, we have evolved methods that we *think* we know to be both good and effective. It is my belief that we are far too content with this type of reasoning.

Secondly, evaluation turns out to be hard work, expensive, time-consuming and technically difficult.

Lastly, it is now apparent that evaluation is a discipline all by itself, and not many disciples are available. It seems also true that the discipline is, to some extent, quite backward in its development. Thus, application of the discipline is even more difficult.

The great importance to Regional Medical Programs of evaluation was recognized early by the National Advisory Council and Review Committee. Many of you will recall the numerous early messages from the Division about evaluation, and the resulting anguish, frustration and even outright hostility felt in some of the regions. In retrospect, I don't think anyone concerned fully appreciated the three reasons I have given for the initial negative feelings about evaluation.

During the early years of the programs, the case for evaluation was argued. A significant amount of research in evaluation techniques was supported by the Division (wisely, I think) - as well as training programs, conferences, seminars and the like - all designed to provide needed expertise. As a result, while we are much better off today than we were four or five years ago, the problem still remains. I'd like to discuss RMP evaluation as I now see it in 1970, from the perspective of a member of the Review Committee and a medical school dean.

To go back for a moment to the first of my three reasons for our aversion to evaluation, it seems obvious to me that the trouble with "seat of the pants" flying is simply that technology has rendered it totally obsolete except in bush country. Anyone flying a plane nowadays, almost anywhere, can pinpoint his location accurately in seconds. And, if he is approaching O'Hare Field and wants to survive, he must do so, and know how to use the proper technical devices.

In point of fact, the methods we have developed by trial and error over the past 50 years in both the educational and health service fields simply aren't doing the job, and we must now very accurately and scientifically determine our position, and plot a new course. We must assess our education and health service systems, and plan to make needed changes. I'm absolutely convinced that Regional Medical Programs are, as Dr. Schon has said, the best mechanism that now is available for doing so.

Since we are trying to make changes in a lot of "traffic" — when surrounded by agencies and organizations and individual citizens (often irate) trying to do similar things, I think the O'Hare Field analogy is quite appropriate. Anyone trying to get a program off the ground today had better know precisely and scientifically where he's going, how he's going to get there, and very importantly, when to land. "By guess and by gosh" isn't good enough anymore. And we should reject the argument that intuition tells us we're being good or successful in medicine as in flying airplanes.

The importance of regional capability in evaluation is made evident by the current efforts of the Division to decentralize authority and thus enhance regional autonomy. We are moving to the anniversary review system, to local project review and approval and to greatly increased overall regional autonomy. In theory, this is very, very good. In practice, there are definite dangers and problems.

Early in the program development, the Review Committee often found that regions were passing the buck to the Review Committee when theoretically they shouldn't have been doing so. Two reasons were commonly given for this avoidance of local responsibility:

First, regions were new, and local expertise simply wasn't available to allow local determination of the value of the proposed program. The Review Committee early on saw literally dozens of projects with no stated goals, no hope of evaluation and really no hope of accomplishment. Yet, this was the best the region could do at that time, in that particular field of endeavor. This was very understandable, and led to the establishment by the Division of the research and training programs mentioned earlier.

More bothersome, really, was to receive a proposal of much poorer quality than one might expect from a particular region. This was often justified by the region on the basis of political expedience: it would be better for the National Review Committee to turn a poor project down than for the local program to run the risk of alienating some faction. I'm sure that early in the program, many local fights and much hard feeling were

avoided by this ploy, but such tactics do delay decentralization of project review and approval. Happily, I think we are now rapidly overcoming these difficulties and using Dr. Schon's analogy, I would agree that the metaphase is upon us, and the diagram on the board to your right really is applicable now, if the nuclear chromatin represents the evaluation and review of most activities within Regional Medical Programs.

I recently have discovered that most regions realize that the National Review Committee is only a collection of individuals drawn from regions. Several regions have begun developing their own specialized review bodies which often for specific purposes are better than the National Review Committee. On two recent site visits, was provided with sounder, more detailed reviews and critiques of projects than the National Review Committee has had the time to develop. Some regions have mounted their own project site visits, using both their own experts and consultants from other regions. Several of these project reviews were so good that the Division sponsored site visits added little to the understanding of the project or activity. I'll add parenthetically that I have noted a regrettable reluctance by regions to respond to the criticisms of their own experts and review bodies, so that the same deficiencies existed, both at the time of the Division-sponsored site review and the subsequent Review Committee and Council meetings. But of great importance is the growing realization by regions of the value of a sound review process, of good project planning, and of good evaluation (of both program and projects), demonstrated by the willingness to hire or borrow the expertise necessary to do these jobs well.

As for the future, I agree almost completely with Dr. Schon's estimation of what will be important for us to accomplish. Anyone following developments in the health field today, for example, realizes the probability that private medicine is in danger of pricing itself out of existence. As one result, during the next few years a great effort will be made to control, however possible, the cost of medical care. This may well involve Regional Medical Programs. For example, there is currently a great rhubarb, which interestingly enough is pitting the American Hospital Association against the AMA and others, concerning the idea of creating "Professional Standards Review Organizations." They represent expanded, more powerful utilization review committees. It has been proposed that these organizations be established by local medical societies, which would then be charged with evaluating medical care and making decisions as to reimbursement for this care. The conflict arises over who should have this degree of power. But

more importantly, the question to be asked is whether or not county medical societies have available to them the expertise to do this job. If this legislation passes, I would imagine that at least in some areas, Regional Medical Programs will very suddenly be in the business of evaluating not only their own programs, but also the extent and quality of medical care delivered in their region. This should be a sobering thought to a good many of us here today. I believe that our traditional involvement with the providers of medical care will soon be put to very good use, indeed, as we get more and more directly involved in the problems of quality and availability of health care.

If you have also followed the life and hard times of medical education, you know that while we need many more physicians, simply graduating more of the same type of physicians we now have is not thought a solution to our health care problems. We are told that our current graduates are not able to solve the problems of our health care system, that our curricula are too narrow, and the training base, largely the urban specialized teaching hospital, is irrelevant to much of community medicine. Thus, there is now general agreement that medical education must be geographically distributed, for one thing. Also, medical schools must assume increasing responsibility for graduate and continuing medical education, and they must train a variety of types of physicians to practice the profession in totally new ways. Medical schools must engage more and more in health services research. Finally, the new physicians must stay in the state where they were trained, and be paragons of virtue and excellence. What is common to all these goals is the involvement of what is now called the "private sector" of medicine. Indeed, what we in medical education are looking for is some way to create a brand new education/medical care system out of the old separate systems of education and care.

In the past, some Regional Medical Programs have looked to medical schools to provide expertise for planning and for projects such as training programs for coronary care nurses. I'm convinced that medical schools should now be looking to Regional Medical Programs for help in creating the new education and service mix, incorporating most or all practicing physicians into a new system of teaching, learning and service. Our new graduates, like many physicians now, must all assume a lifelong responsibility for learning and teaching, for renewing their own talents and skills and those of others. If medical societies or the profession as a whole is given or assumes the responsibility for setting and keeping its own house in order, Regional Medical Programs will,

without question, be turned to for the process and the expertise to do this job.

An important key to success in all of these things is good evaluation. Regional Medical Programs are still the best instrument our society has created to do all these jobs, and we must develop the necessary capabilities. As the action moves to the regions, whether we succeed or fail will depend on how well we manage the tasks. If we know what we want to do, we also have to know how well we are doing it. And evaluation in these terms is the only possible way to manage our efforts. I believe that the climate is now favorable for evaluation. In recent years we have seen significant fractions of Federal agency budgets earmarked for evaluation. It has become accepted practice in Regional Medical Programs to budget specifically for the costs of evaluation. Thanks to Regional Medical Programs and other agencies such as the National Center for Health Services Research and Development, growing numbers have been trained in the science of evaluation. If these experts are not locally available, they usually can be brought in as consultants for a time.

I suspect that as we mature as a program, national conferences such as this will diminish in number, and we will have regional conferences on evaluation, regional training programs, and the emergence of the word "evaluation" as a very friendly, commonly used, everyday household word — safe even for young children.

PETER D. FOX, Ph.D.  
Senior Economist, Office of Management  
and Budget

I would like to begin by discussing some of the trends in Federal health expenditures as background to understanding the context in which all health programs, including Regional Medical Programs, are likely to be evaluated in the next few years. Federal health expenditures are large. They are expected to exceed \$20 billion for the first time this fiscal year and represent over 10 percent of the total Federal budget.

Many of the health programs were started during the 1960's and carry with them the potential for tremendous demand for increased funding. For example, some 80 comprehensive health centers, funded by the Office of Economic Opportunity and HEW, are now in operation, and each center receives an annual Federal contribution of roughly \$2 million. Few of these centers can be self-supporting without Federal project funds, and estimates of the number of centers required to meet health needs in poverty areas run as high as 800.

Similarly, the Federal Government has supported staff of community mental health centers on a seed money basis. Federally financed centers now in operation provide services to less than 20 percent of the country. Already, the authorizing legislation has been changed to extend the time limit on the grants from 51 months to 8 years because many of the centers have not become self-supporting. Whether these centers will be self-supporting after eight years is questionable, and in the meantime, increases in budgets are required merely to support existing commitments.

Similarly, pressures exist to expand the Medicare and Medicaid programs. Many medical schools, rightly or wrongly, say they face insolvency if they do not receive additional Federal support. The pressures for Federal support of health research are strong. Some people argue that health services research is underfunded. And, last but not least, I see estimates that Regional Medical Programs requires at least twice its current level of funding to be fully operative.

I will not attempt to project the actual size of the Federal health budget in the coming years. However, it is clear that we must do better with the funds that we are already spending. This is the environment in which we live, and it is a considerably tighter environment than the one to which we were accustomed during the last decade.

What, then, does the Office of Management and Budget expect RMP to contribute? The goals of Federal health programs in general include improving the health status of Americans, increasing the efficiency with which care is delivered, and fostering equity of access to medical care. RMP is expected to assist in achieving these goals, and in setting budget levels, OMB must assess whether the \$97 million currently spent on RMP could have higher payoff if spent on other programs such as Comprehensive Health Planning or the National Center for Health Services Research and Development. We also assess the alternative of not spending these funds at all.

Measuring directly the impact of RMP on the achievement of these objectives is difficult, and one must be content with proximate measures. These include changes in decisionmaking procedures, in decision outcomes, and in attitudes. For example, RMP should be able to demonstrate that it has promoted sharing of health resources in a manner that contributes to better care or increased efficiency. The commonly used argument that RMP has achieved better communication among those concerned with the health care system does not in itself justify the current level of expenditures.

Most of RMP expenditures are for three types of activities—support of the efforts of core staff, demonstration projects, and continuing education and training programs. Consequently, the questions that the Office of Management and Budget is likely to ask of RMP in future years will largely be directed towards the output of those activities.

First, with regard to core staff. Are their activities in fact promoting new patterns of medical care? Subsidiary to this, one can ask whether these activities are successful in rationalizing the relationships among the various organizations in the region that deliver health care or otherwise impact on the local health care system. RMI should prevent wasteful duplication in training programs and health care facilities. Core staff should both foster the acceptance of new technology and promote new approaches to health care delivery. For example, training programs for physicians assistants and other types of nonphysician manpower are now multiplying in an uncoordinated fashion. The problems of the location of training facilities, training content, career mobility, and physician acceptance of new forms of manpower should be concerns of RMP. Is RMP successful in achieving solutions to these problems? Similarly, is RMP bringing about proper coordination among health care facilities? Has it achieved an appropriate level of coordination with other government programs such as Neighborhood Health Centers and Comprehensive Health Planning? Is it providing a vehicle for physician acceptance of new forms of medical practice, such as prepaid group practice or improved referral patterns, that may lead to higher quality or less expensive care?

We also expect RMP to fund only those demonstration projects, continuing education courses, and training programs that are an integral part of a well-conceived strategy to satisfy the health care needs of the region rather than their essentially reflecting discrete and uncoordinated proposals that are simply related to the interests of the persons applying for funds. Much has been said about the diversity that exists among RMP's. Such diversity is commendable if it represents a response to local conditions and factors. It is less commendable if it devolves from confusion over objectives or how to carry out these objectives.

Core staff should also avoid funding projects or training activities that the market place is likely to undertake without Federal support. Nor should it engage in activities that do not result in efficiency increases or medical care improvements that are of sufficient magnitude to justify the related expenditures.

In evaluating demonstration projects, market criteria seem very appropriate. Funds for these projects are intended as seed money. This implies both that the funds serve to stimulate new activities that would not have been undertaken without RMP support and that the activities are sufficiently attractive that the medical market place is willing to support them after an initial trial period. The extent to which RMP-generated projects are sustained after RMP funds are withdrawn is an important measure of effectiveness.

Training programs should increase the ability of health professionals to deliver health services by bringing them up to date on recent technological developments. They should also increase the productivity of the medical sector. Health professionals should be trained to use new capital-intensive devices. Physicians should be trained to use new forms of medical manpower. Non-physicians should learn new functions so that they can substitute for physicians and thereby permit physicians to spend time on activities that only they can perform. As with projects, one might ask why individuals' courses require RMP funds and why they are not supported in the private market place.

Program evaluation has at least one function other than simply leading to decisions on whether program expenditures are justified. In particular, evaluation should result in redesign of the program. Thus, if certain program activities appear to be successful and others not, the successful activities should be emphasized. Similarly, one would hope for information to improve the functioning of even the most successful elements of the program. The Office of Management and Budget is interested in the quality of evaluation at all levels. We are interested in evaluation of the total program, of individual RMP's, and of individual projects and training efforts. We expect evaluation to lead to assisting or phasing out weak programs or projects. While an overall cost-benefit analysis of RMP may not be feasible at this time, we would like to see a few clear successes perhaps along with quite a few ambiguous ones. We recognize that there will always be some mistakes and failures in this type of program, although one would hope that these would be as few in number as possible.

The health care problems of this country will not be solved simply by expanding Federal programs to support health services or by increasing the supply of existing manpower and institutions. RMP should be at the forefront of promoting the changes required at the local level to make the health care system and its related technology more efficient, more effective, and more accessible to the American people.

RICHARD S. WILBUR, M.D.  
Deputy Executive Vice President,  
American Medical Association

Thank you very much for the kind introduction and the chance to be here before a group of people in whose work the AMA is so deeply concerned.

Now, when I speak for the AMA I should make it clear that I am speaking for an organization of practicing physicians, and as such we are concerned primarily with the problems of practicing physicians — in the manual aspects of the delivery of health care services — the people who actually touch the patient.

Our major problems are those general problems you know so well:

The manpower shortage. . . And I'd like to say this is mainly a shortage of front-line troops. If any of you have followed the development of armies over the last century or so, you know that, in years gone by, if a general had 100,000 troops, he could usually expect that most of them would get into the fight on the day of the battle.

As you now know, if a general has an army of 100,000, he's lucky if 10,000 of these men actually get into the fighting. . . Or maybe they're unlucky.

What we find in the medical care field is very much the same sort of thing. Everybody wants to be a consultant. There's little reward or recognition for the primary physician, and he sometimes gets a little lonely when he thinks of all those night house calls he has to make and all the people who are planning on how he should make still more of them.

So the manpower shortage to us is that of the practicing physician, although there are many other shortages as well.

Second, we have a concern about quality of care that has been expressed before today.

Third, is the much discussed problem of cost which needs no elaboration before this audience.

And, of course, we have the problem of remembering that we are dealing with human beings. Problems of cost and human factors are certainly widespread today. At least we know the feelings of our college students, who are well versed as to their educational institutions and the loss of human factors in some of the larger medical teaching institutions.

The problem pertinent to this meeting is how to get information to those doctors who, so to speak, are in the front lines.

We find that what the practicing physician needs most is help in solving the common problems of common diseases in common people.

Sidney Garfield of Kaiser Permanente, in writing for *Scientific American*, speaks of the "slightly sick and the worried well." These people make up the volume of patients that these doctors see.

We need help in knowing how to see them in the office, and possibly even more so, we need help in keeping these patients well so they don't have to come to the office. And what is even more important, we need help in keeping them out of institutions, particularly, of course, hospitals.

Being in an institution is not only bad for the budget, as Peter Fox has just stated, but it's bad for a patient, and he should avoid it if at all possible. Being in a hospital is bad for a patient's morale. And as many of you know, it's where most of the side effects of treatment occur.

The physician needs help in the prevention of disease. I don't mean by this just immunization, because we don't see many diseases these days that are preventable by immunization. Maybe it's because we have immunized people so well already.

We don't need help in delivering more physical exams. I won't bore you with the argument of whether a physical examination is worth the money spent or not, except to say it's a highly debatable subject — and that I intend to go on getting them. As a good internist, I could do no other.

But we do need help in the real problems that face us, the things that cause people to get sick and to come to the doctor's office — tobacco, automobile accidents, pollution, the lack of exercise, nutrition — in the inner city, too little nutrition, and in groups like this, too much nutrition — urban crowding, sanitation, alcohol, drugs, etc.

And then, of course, what causes us the most trouble, is the psychic stress of our day which drives the patient into the doctor's office.

This is where we need help. And as we look at and evaluate the ability of RMP to plan, it's not just how many coronary care units are set up, but by working on the causes of disease, how many people could be prevented from ever having to use a coronary care unit.

The value to the provider is in helping him to take better care of people. And, as I said before, he needs help in dealing with the common diseases which common people develop commonly.

Now, there is an historical problem that has developed with getting this help from the medical schools

and other institutions. In times gone by, clinical research was done by a clinician, who took an afternoon or evening off, or even went on a sabbatical and did research. He could then use this research in his practice.

As research became more complex, this evolved until there were two people, the clinician and the research man. They got together at lunch time or shared common meetings to exchange information.

It's often said now that we need a third man, a translator, who could tell the research man what the clinician was doing, and who, more particularly in recent years, was able to explain to the clinician just what it is that the researcher did and what it means to the clinician in terms of his practice.

I think we need a fourth person too, and he is in the field with which you are concerned — the communication of this information to the clinician *after* it is translated so that he can understand it and can use it. Just as important is the communication back to the medical school, of the kind of information that the clinician really needs, so that it can be translated, at least at the clinical level — not at the basic research level — into the kind of research that is going to help him do a better job.

We need a two-way street. Let me use an example. Many of you know John Hogness, a former Dean of the Medical School at Washington. He wrote a very good article and gave a superb speech about the time he spent a couple of weeks filling in for a general practitioner in the rural areas of Eastern Washington.

He's not quite sure how medical care in that community fared during the period he was there, or how much he helped it, but he is very sure that he learned a great deal that was of value to him in training more physicians.

It's a two-way street, with which we need your help, because we need practical planning — planning for people and not just for census tracts.

We need to avoid overspecialization in planning. We need to avoid the problem we run into when we solve one problem and, as the old saying goes, we cause two others. It's all well and good to solve the problems of uremia with renal dialysis and kidney transplants and to make these procedures generally available, but in doing so we diminish the budget available for housing and for pollution and for the other problems of health care which may be more important to more people.

As we solve the problems of keeping the elderly and the chronically ill alive, we build up the problems involved in the population explosion. As these people stay alive there is less for the rest of us. Or, if you believe in

the theory of population zero, the longer you keep an older person alive by new modern techniques, the longer it is before a new life may enter this world.

We must have overall planning, not single-problem planning.

The doctor himself must always be concerned with his own individual patient. He cannot be concerned with whether keeping his individual patient alive is a good thing or a bad thing. He is committed to keeping the patient alive.

And, therefore, you must be concerned with whether his success and your success will actually make this a poorer world for all of us to live in.

Our evaluation may not be as sophisticated as those which you have heard and which will be discussed further at this meeting, but we will certainly ask: Have you in your planning helped physicians to deliver better care and have you helped the people of this country to live a better life?

If you have done this, then our evaluation of your planning will be that it is a total success.

JOHN M. BLAMPHIN

Assistant Bureau Chief - Washington Office

*Medical World News*

For several years now, I've been covering medical meetings and listening to speakers get up and give their papers. It always seemed so easy. I figured all you had to do was to step up to the microphone and say with some degree of confidence — "First slide please." And the projectionist would do all the rest.

So here's my big chance and I didn't bring any slides. As a result I stand up here this morning with a bit of hesitation, knowing full well that of all present at this conference, I know the least about the intricacies of RMP and the science of evaluation.

But before I jump into the topic of evaluating Regional Medical Programs from the public side of the fence, I thought I should tell you something of what I do, and how I view my own relationship to what you are trying to do.

As you know, I work for *Medical World News*, McGraw-Hill's weekly news magazine for physicians. It is my job, simply put, to tell doctors what is going on in Washington that is important to their practices. My primary audience is about 200,000 physicians in private practice and on hospital staffs.

In addition, we go to about 5,000 people around the country who subscribe or are on our "Freebee" list. In Washington, the list includes dozens of Congressmen and

Senators and their personal and committee staffers, top officials in HEW and through the department's health agencies, and many representatives of voluntary and professional health association and consumer groups.

We also go to a select mailing list of medical and science writers on major newspapers across the country. Hardly a week goes by that MWN is not quoted in the press, or on radio or television news. So, the public I represent is far wider than the medical community.

In Washington, my beat is primarily the political and economic side of health. I regard my role as one of evaluator. I watch what is going on, attend hearings, read testimony, talk to dozens of policy makers both on the Hill and in the Administration, listen to the reaction of other groups, then when the time is ripe attempt to set events into some perspective for my readers.

As to Regional Medical Programs? I joined the magazine during the days of the DeBaKey Commission, and began to cover Capitol Hill during the House and Senate hearings on Regional Medical Complexes. That was the time of the mighty 89th Congress when passage of a new federal program was regarded as the answer to all the problems which plague mankind. You know — take one RMP, add water and stir. Voila! Instant health care for all. I believe that approach, incidentally, did you a great deal of harm. But more about that later.

So I watched what went into the Congressional mill and I saw what came out. I've been watching and evaluating, and reporting your progress ever since. Evaluation of RMP takes a simple format for me. I merely look to see what progress you're making toward a single goal — the delivery of high quality medical care — the latest medical science has to offer — to patients with heart disease, cancer, stroke and related diseases. I also look to see in which ways the means developed to deliver that specialized care are also used to cope with other more general health problems.

Over the past five years, I have performed this evaluation by reading your annual reports, by hearing your representatives before congressional committees, by talking to RMP officials in Washington, by visiting regions whenever I can find the time, and by listening and reading what others say about RMP — the usual routine a reporter goes through covering his beat.

In the course of this evaluation, I have formed some opinions about RMP and health care in general which I believe are shared by a great many people in Washington these days. To me, the quality of care and the way it is delivered go hand in hand. One is useless without the other. It does no good to tune an automobile engine with new points and plugs, and add a fancy fuel



injection system, if the car's transmission is shot, and the tires are bald. It's the same in the health biz. Tuning the skills of physicians and hospitals to a high degree of quality and efficiency is no good if the system through which those skills are passed on to patients has broken down.

It is my opinion and the opinion of others in Washington I spoke to about this before coming out here, that a federal program such as yours which is using the taxpayer's money, cannot stop at providing the physician, the hospital, and other providers with quality tools. It must also do what is necessary to see that these tools are applied to patients. Many of us have the uncomfortable feeling that there are those in Regional Medical Programs who feel their responsibility has ended at the conclusion of a continuing education course, or after the technicians have installed the coronary care equipment.

Nevertheless, I have seen evidence that you are moving — albeit slowly — toward a patient-centered goal. About a year and a half ago, for example, Dr. Robert Headly, a Bowman Gray cardiologist, took me on a tour of several small hospitals in the State of Franklin in Western North Carolina. During our visit, to the 50-bed C. J. Harris Community Hospital, the doctor showing us around asked Dr. Headly if he would look at one of his patients who was in the hospital's coronary care unit — staffed incidentally by nurses trained with RMP funds.

Dr. Headly readily agreed and a few moments later in the hallway I heard this exchange: "If you can you'd better send her on in," said Dr. Headly. "You've gotten her out of a failure this time. But if it happens again, she'll probably go fast. If we give her a valve, she's got an 85% chance." The local doctor pondered a moment, then asked "When can you take her?" "In a day or so, I'm sure," replied the younger man. He pulled a pad of paper from his inside pocket and began making a few notes. "You talk with her family and I'll let you know tomorrow, maybe tonight, when to bring her down to Winston-Salem."

I don't know the outcome of that case, but I suspect the exchange between rural physician and medical center specialist saved a life. I do know that it probably would never have happened were it not for the North Carolina Regional Medical Program.

I also understand that after the RMP helped put coronary care units in several of the small hospitals in Western North Carolina and also established a mobile coronary unit staffed with rescue squad workers and aided by local physicians — that the mortality rate from

heart attack has dropped better than 60 percent. I call that delivering health care to people.

In region after region RMP has successfully brought the normally fragmented elements of the health community together to talk about the state of health care in this region, to admit that gaps and weaknesses exist, to identify them and then to plan ways to improve the situation.

For the private sector of medicine this is a tremendous accomplishment. For the first time, in many sections of the country there are evolving systems of health care which are more than the sum total of their parts. This advance may well have laid the groundwork for the development of new health care systems such as the Health Maintenance Organizations now being touted by HEW, which could never come about without the change in atmosphere which RMP's have created.

But lest you think I have been completely snowed by RMP, I must also say that I believe this success has been spotty, and has worked better in rural areas and smaller communities than it has in highly complex metropolitan areas. As a colleague of mine said to me at lunch the other day, "The real test of the RMPs is not in being able to organize care where it is *unorganized*, but to organize care where it is *disorganized*. And this," he added, "just hasn't happened."

Since I have followed the RMP for five years, perhaps more closely than other reporters in Washington, I understand the significance of what you *have* achieved. I know also how difficult it is to evaluate this type of groundwork in terms of the usual morbidity and mortality indices.

But let's speak about the public for a moment. By public I mean just about everyone outside of RMP and the health professions. Included might be the present administration in Washington, the Congress and voluntary and professional health organizations. How do they evaluate RMP?

The present Administration evaluates RMP in terms of national goals. And so far as health is concerned, this means using federal money first and foremost to influence changes in the organization, and delivery and financing of medical care. It also means spending money in a more flexible, non-categorical way. One needs only to read the Administration's Health Services Improvement Act of 1970 to get the Administration's present evaluation of RMP.

I must also say in passing that the Administration's commitment to health care so far hasn't manifested itself in much more than rhetoric. If the President is

serious about solving the so-called health care crisis, he is going to have to substitute money for talk.

Now, to the Congress: Here the evaluation has been more simplistic. It also demonstrates how RMP got off to a bad start. It goes like this: "If you spend all the money we appropriate, you're doing a good job. If you don't you must be dragging your feet." As Rep. William Springer of Illinois said during the hearings on extension of RMP two years ago, "The initial legislative testimony was presented before this committee to justify a program for a billion dollars, which turned out three years later to have spent \$85 million." But he also shows a great deal of insight into the nature of new programs, and especially of RMP's when he says, "I think it ought to be brought out here that what we get in the way of landslide testimony here is a selling job and snow job claiming that something can be done immediately."

It has been my observation that the promises which were made for health care at the beginning were made in terms of the original DeBakey Commission report. They were not significantly modified as the program itself was modified by the Congress. As a result, the evaluation of the promises and potential was made before the RMP even began. And when you try to evaluate what you can't even define — and who in 1965 could define a region in understandable language — you get into trouble.

Only recently have members of the legislative and appropriations committees begun to understand the subtleties of RMP. But they too, like the administration, expect RMP to pay more attention to problems of delivering health care. As Sen. Ralph Yarborough said earlier this year in introducing his bill to extend the RMP program: "Explicitly, the extended legislation provides that Regional Medical Programs concern itself with improving the organization and delivery of all health services, and strengthening our primary health care system."

In the meantime, RMP still has to prove itself on the Hill. One Capital Hill staff member told me the other day that "Regional Medical Programs have failed to take on broadened responsibility in health care and have hung tenaciously to heart disease, cancer and stroke labels."

True, you have pointed out that without the disease categories, medical center specialists may be less likely to participate. There may be some truth to that. But many on the Hill read it as a cop out and as an attempt to maintain the status quo.

Individuals within health associations in Washington are also skeptical of RMP progress. But these, of course represent vest interest groups — hospitals, medical

schools, voluntary health associations — all of whom are looking for a piece of the action themselves.

Now why is there so much cynicism surrounding RMPs? I think my friend Ed Friedlander would boil it down to a problem of communication, of providing the facts from which others can make evaluations.

Certainly, not everyone can spend time looking over the operating projects within one or more regions to learn first hand what is going on. So, I would suggest that you consider very carefully how you justify your existence to your publics. In a nutshell they — and I'm talking about those in Washington — want to know what you are doing for people, for patients, for constituents. When they hear you talk in your own jargon of regionalization, of cooperative arrangements, of closed-circuit TV and other gadgetry, they are going to go away shaking their heads. Maybe *you* can translate those matters into improved delivery of patient care, but *they* can't. As far as they are concerned, you're off in some other world.

Let me give you an example of what I mean. About a year ago during the budget crunch for RMP, the Illinois program, like a lot of others, wrote letters to its Senators pleading for reinstatement of RMP funds cut out by the House, and describing what would happen to the program if the money is not put back. The letter was well-written, telling how the region had been organized, about the progress toward achieving regionalization, and how after months of planning, grant applications were pending to put the program into gear. The letter said that if the money wasn't forthcoming that those grants could not be awarded.

But nowhere did it say what the money would be used for in terms of helping the people of Illinois who just happen to be Senators' Percy and Smith's constituents.

You must remember that members of the House and Senate get dozens of letters a day pleading one cause or another. Most of these are handled by aides and only get a cursory review by the boss. If you had been sitting in a busy office on Capitol Hill reading that letter, knowing nothing about the concept of regionalization and caring less, what would your evaluation have been?

To me it boils down to this: If you communicate the proper information, and by proper I also mean that it be honest, and communicate it in the context which your audience can understand, that is patient care, the evaluations you get will more likely approximate the true state of RMP. You will, of course, still have critics who will say that only total federal control of health

services can eliminate the gaps in delivery and quality we now face.

In the long run, you will be judged by the changes that occur in the quality and delivery of health care which result from your activities. And in the future, for Regional Medical Programs to survive as a major federal

program in health care in the eyes of the administration, of the Congress and of the public, those actions in my estimation are going to have to be directed more and more toward improvement in the quality of the system through which the best medicine known to science reaches the patient where and when he needs it. At least that is going to be my yardstick.

## LUNCHEON REMARKS

HAROLD MARGULIES, M.D.

Acting Director, Regional Medical Programs and Service

When Dr. Glasgow was introducing me, he talked about my task of drawing together the threads of the morning session as a "herculean task." I remembered some of the labors of Hercules. One of them was to clean the Aegean stables. I guess you do remember, don't you?

I was also thinking this morning as I listened to the descriptions of evaluation and of the Regional Medical Programs, of something that I had almost forgotten about -- a flashback to my early youth where I once saw a picture of a man standing on the deck of a ship. He was the greatest archer in the world, sort of the modern Robin Hood. He was standing on the ship's deck with the deck going up and down. There was an empty keg floating in the water with a little cork in the bung, and his job was to hit that cork with the arrow.

I had the same sensation when someone was talking about evaluating the Regional Medical Programs, and it gave me the opportunity as I sat there to decide what the title of my talk should be, because I frequently give talks without titles and then somebody wants one.

I have selected one for this one. It is as follows: "By the Time I Get to Where It's At, It's Always Where it Was." Which seems highly reasonable.

Before I comment on the general discussion this morning, I must say it was superb from every point of view. The thanks should go not only to the participants but to Pete and the people who have helped him put the program together. It's off to an awfully good start.

I would also like to say a few things about the general atmosphere, sort of overall environment in which we are thinking about evaluation, whether in Regional Medical Programs or in other areas. I was particularly charmed by the sense of determination to deal rationally with systems which have often been dealt with intuitively, and the expressed preference for the rational over the intuitive.

At the same time, I had to realize that there is a drift in this nation, a preference for mysticism over thoughtfulness, which expresses itself in interesting ways. The Knight newspapers did a survey not long ago of some 1,700 readers to see what they thought about people landing on the moon. Some of them had interesting comments which give you a sense of what at least part of the country feels at the present time. One lady said, "I

can't see how they could have been on the moon. My TV set can't pull in New York. How could it pull in the moon?" They talked to a man in North Carolina, and he said, "You know, if you got on an airplane and went to Ashville and then came back and I saw you again, how would I know you had been in Ashville?"

I'd like to also point out the fact that 1,200 of the nation's 1,700 newspapers carry daily horoscopes -- and a few years ago 90 did. And last year there were 2 million ouija boards sold -- which is the greatest bonanza in the history of the business. Now, those are just casual observations, but they are, at the same time, symptomatic of a drift toward the mystic, toward the intuitive, toward the doubtful, toward the seeking for solutions which are non-rational, at the same time that we are trying to look very strongly in rational directions.

When I looked in the *New York Times* this morning the present status of important legislation was listed, but as always the Regional Medical Programs were not mentioned. I think this also helps you to appreciate the environment in which we are functioning.

Aside from these general statements, I think we must recognize that in looking at the Regional Medical Program or any other health activity from the evaluative point of view, we have to enter into a game to which we are generally unaccustomed. The health profession does not characteristically evaluate its own practices or its own institutions. It may do so on an individual basis, but on a broad basis, little or not at all. If you doubt that, try to look at your own program sometime, or look nationally at what you have available if you want to measure the influence of some health event or the community. And look very hard to see if you can find any information that will allow you to say, "Here is where we stand and here, as a consequence of what we are doing, is where we are going, and here in retrospect is where we have been." It's an astounding fact that those kinds of data bases do not exist.

We do not, generally speaking, relate our institutions to the processes which we have been discussing today. We do not relate general health problems to the efforts in which we invest. And we allow ourselves little managing room to set up a conceptual basis for future planning. To ask Regional Medical Programs, as a consequence, to enter into this kind of a process is to be

asking for perhaps even more than any of you here realize.

I think at this point it might be useful for me to recapitulate as best I can what was said this morning, which, I believe has set the tone for this afternoon. discussions. There were some general themes, which were sustained throughout the entire morning. They were things I have said to you and you have said to me during the past several months. They included the determination that this health program, like any health program which has the backing of the federal government — and probably those with other kinds of backing — must be involved with improving the organization and delivery of health care.

Words like heart disease, cancer and stroke were brought up either on a pro forma or on a somewhat wistful basis. This by no means suggests that people are no longer concerned with the specificities of these diseases, nor does it suggest that our legislation or our legislative mandate has suddenly changed. What it does say is that priorities emerge in a society, and the priorities which have emerged in this society are being enunciated from several points of view - from the public point of view, from the evaluative point of view, from the fiscal point of view, from the provider's point of view. We were told that if we are to look at Regional Medical Programs and evaluate their usefulness, it must be with the prior determination that we have set some value structures, that we have a clear statement of what it is we wish to be in the health care system.

I thought the major speaker of the morning described magnificently the problems involved. He told us that we must have some goals which are clearly stated, that are determined locally by the regions, and that these goals cannot be controlled with meaning and with purpose on some kind of a central basis. He described to us a pattern of Regional Medical Programs which have their own special knowledge and their own special issues which must be introduced into their planning flexibility, and which must design well defined goal-related programmatic efforts that can be evaluated. He said, as did others, that some of this may have to be retrospective, and that makes me a little nervous, because I think very little of what we can evaluate retrospectively is going to have much meaning when we get to the prospective end point, and a reconsideration of what Regional Medical Programs are all about. In fact, the issue of which direction RMP will take is either an evaluative, a political issue, or a social issue which must be looked at very attentively. And perhaps this was a little bit vague in the presentations this morning or there wasn't time to get

into it explicitly. If it is true — and I believe it is — that RMP has changed its purposes in the very process of serving its initial requirements, then to carry out evaluative activities on what it has been doing runs the risk of being archival rather than programmatic. We must very carefully distinguish between what will be of value historically and what will be of value for the purpose of building a new kind of a program.

In fact, if I were to make a generalization about RMP — and we all know that evaluation and generalization are dangerous companions — I would say that certainly one of the great potentialities of RMP is to create an environment among health care providers that will allow it to be as well sensitized as possible to the indeterminable changes that are coming — a kind of preparation for the apocalyptic — which may be a rather tall order. And since RMP reaches that far, it has to be thought of perhaps not too loudly, but at least above a murmur.

I also heard this morning that we must be careful not to ruffle too many feathers. I suspect my emphasis would be to ruffle the feathers but don't smash the bird in the process, because I think some feathers will indeed have to be ruffled.

I heard something else — the need to look at various aspects of the evaluative process, the justification aspect for instance, which from my point of view is a very urgent one because I have to defend and justify the Regional Medical Programs at all times in Rockville, Washington, or wherever I may be representing them.

I heard about evaluation. It has something to do with methods of controlling what happens, of conformation to standards - standards which are established according to the requirements of the program and which are determined locally - and of a process which is called learning in which we perceive new ways of doing those things which we do, and their relationship to where we wish to go. I was especially pleased to hear references to the need to avoid replacing the objectives which are being sought with the measures utilized to reach a goal. It is so easy to establish measures - and we in the bureaucracy are fully capable of doing that — so easy to establish measures that describe how to get there, and then concentrate on meeting those measures so fully that why we established them disappears.

I used to see that a great deal when I was working in Asia, where the great game was to have a five-year plan. The five-year plan would contribute - whether it was in education or in agriculture or in health - what was to be done from point zero to the end of that point. One could be very sure that at the end of the first year, 20 percent of the goal would have been reached, at the end

of the second year 40 percent, and so forth. Any variation from that was easily corrected by replacing the fellow who was doing the reporting. This is one way of getting where you want to go, but it does seem to emphasize the measures more than the goals.

There also appeared during the course of the discussion some reference to the need to examine all alternatives instead of simply taking advantage of the opportunities. And again I was thinking a little of what might be called a kind of Mae West approach to this thing. A young lady asked Mae West how she could get out of the particular dilemma she was in. She said, "I'm sure I can never do what you have done, which is to find a man who loves me and has \$10,000 and who would buy me the beautiful kind of a mink that you're wearing." And Mae said, "Well, honey, you could think about getting 10,000 men with a dollar each." You see, she understood the discovery system and she understood the ways in which you do develop alternatives.

I suspect that most of us were thinking during the course of the morning what all of this discussion about evaluation meant with reference to the process that we are going through at the present time in Regional Medical Programs. The process is something which is called anniversary review, and will obviously place a very different kind of burden and emphasis on the Regional Medical Programs. If I do nothing else in the process, I would like to say that I am convinced that what you heard this morning is so highly consistent with what we anticipate in the process of decentralizing the RMP's, that it could easily be played back again to you every morning for several days to make sure that the message is clear.

There can be no question not only that RMP's will be given the prerogative, but it will be demanded that they establish programmatic directions, and within those programmatic concepts, establish projects which specifically fit those programs. The core activities will all have to move in that direction. There's no question that this will be the way in which we will have to go. There is also no question that there will be a need to evaluate the effectiveness of that whole process and that the way in which you evaluate it will have to be based upon your understanding of where you wish to go and what your senses of value are.

If there is anything to add to what was said this morning, it is that there was probably less emphasis on a sense of expediency than I would have liked to have heard. It came out. It was mentioned. It is that part of the evaluative process that had to do with how rapidly things are to be accomplished. But at the present time, I

am confident that whatever Regional Medical Programs must do, they will have to do it more rapidly than seems at all reasonable.

There are two other aspects of the evaluative process that I would like to speak about. If RMP, as you have heard this morning, is to be as diversified as it should be, and if it is to maintain the flexibility which is one of its great assets, and if it is to mobilize those providers who are always going to have to be involved with the delivery of health care services, it is going to do it in a variegated fashion. And that's fine. But this presents a great difficulty for us in Regional Medical Programs Service. Because while this kind of an activity is going on, there must also be a sense of coherence, which if not maintained, will make the RMP look like another process in fragmentation and in activities going off in a variety of directions. As a consequence, I think it is essential that we establish more effectively within RMPS and among the Regional Medical Programs an understanding in the process of programmatic development and in the process of pursuing programmatic goals, a communication network which lets everybody know what is going on and which gives a better understanding of the expectations in RMPS, with reference to what represent HEW overall goals.

For me to pretend to you that this government or any government can support activities without our own concept of what those goals should be, and without at least a broad kind of framework in which we will function, is to be misleading.

Now, it is not likely that at any point we will be so foolish as to direct the RMP's to do a specific number of things. We would fail in that effort. But I think you need to join with us in the interpretation of what really matters in this country in the health care system. And this you have heard over and over again. You heard that people are concerned about the costs of medical care. You heard that people are concerned about access to medical care. You heard that they are concerned — and I'm not sure in what way this is true — with the quality of medical care. In the public mind quality has a lower order of priority. I think access and cost are far and away the greater considerations.

You also heard from the people who are looking at the evaluative system, where we will have to go and what will have to be done, on the basis of what we have. Simply flooding more into the system is no longer going to be the answer.

You heard a very strong inference, which I join in, from the Office of Management and Budget, that there will have to be greater selectivity in what is supported

and a readiness on your part to abandon what doesn't really seem to be working very well. This will entail some risks, but careful risks. You will have to eliminate what appears to be ambiguous, and give heavy support to what appears to be a strong direction in which to move toward the kind of goals which we have embraced. Now, if we can manage this variety of activities in such a way that we can interpret them coherently in our own defense of Regional Medical Programs, I think we can do well.

Now, mind you, I'm saying this at a time when our legislation has not yet appeared. We are really living on borrowed time — and we're used to that. We still do not have appropriations. We are living on borrowed money — and we're used to that. But regardless of how these events emerge, and even if the definition of our legislation is fairly narrow, you as individuals responsible for RMP's would be most foolish to overlook the elements of evaluation that have been discussed today, and therefore the elements of purpose in Regional Medical Programs.

If we came out with legislation that says: "Confine yourself to categorical programs and within those programs to continuing education," - which for the most part consists of what I now describe as episodic information transfer - if we indeed are to move, are mandated to move in that direction and we do indeed respond by isolated categorical projects, there will no longer be a Regional Medical Program.

There are times to use judgment. There are times to exercise your own knowledge of what is going on and what needs to go on. And what you will have to do is establish evaluative techniques which anticipate even and then be ready to prove, when you arrive at the event, that you have done what is necessary in the process of projection. Anything that depends entirely upon what is here and now is likely to fail. Anything that is purely retrospective will surely fail.

If you have difficulty in deciding where you need to go or what matters, I think a careful scrutiny of the daily newspaper is very, very helpful. If you need to go beyond that, it helps considerably to go where some of the problems are, to talk with some of the people who are not getting the kind of medical care they wish, to consider the fact that the quality of care is not merely a matter of considering the exchange between the provider and the lucky person who enters the system and gets quality care. You must also consider that the final measurement of quality care is diluted by factors such as those who do not get care, or where the quality is so bad that it is a very large minus.

These broad considerations can probably be resolved by a sense of societal concern which has been expressed wherever I have gone in the Regional Medical Programs. But there is a difference between one RMP and the next in the determination that a bold direction is a good direction. In fact, at the present time, the bold ones have been the wise ones, and in a kind of paradoxical way the bold have been cautious.

## HSMHA — AN INSTRUMENT FOR IMPROVEMENT IN HEALTH SERVICES

VERNON E. WILSON, M.D.

Administrator, Health Services and Mental Health Administration

Thank you, Harold. There are enough of you who have heard me speak before, that no one will expect a flowing, eloquent speech that is "snake charming" in capability.

I'm delighted to be evaluated by this kind of group. It seems to me if anyone will do this in an objective way, you will.

As many of you know, the Regional Medical Programs have probably been closer to my heart than any of the other new movements of the federal scene in recent years. This is not a maudlin sentiment. It is my evaluation of the promise of this program.

A substantial portion of that promise arose out of the opportunity to allow the grass-roots mobilization of innovation and the grass-roots decisionmaking process to take hold.

In that context there were several kinds of problems with which you've been struggling over the past few days. I'm not going to treat anew the things you have been talking about. But among them, of course, has been the continuous struggle between the two polar tugs. One of these is to give clearcut guidelines so that people know specifically what to do in order to assure a "good" performance. The other is to wait patiently, Rogerian style, until from all of the massed intelligence, the discomfort of silence brings forth the new idea.

This has been an extraordinarily challenging sort of process following the Rogerian style, because Congress, which votes money on the strength of local support, has had some difficulty understanding why there was a strong movement.

Some of you need to keep this set of complex variables in mind as you look at the way we are trying to explain to the Congress how extraordinarily important we think it is that we let the grass roots make the decisions.

If one characteristic of RMP can be set forth, it is that RMP has not had a distinct public. There hasn't been one particular group, external to the organization itself, which has gone to Congress and said, "We must have this." Instead, there have been several publics who have gone to Congress, each with its own image of Regional Medical Programs, and therein lies part of the problem which I hope we are beginning to resolve.

If you say HSMHA or Health Services and Mental Health Administration, the usual reaction is, "What is that?" I understand that reaction because it was my own when they first talked to me about HSMHA last May.

Let me give just a precis of the Health Services and Mental Health Administration for those of you who may not know what it is.

Regional Medical Programs is one part of HSMHA, as you well know. The Community Health Service is another substantial part. It is a program with a budget of some half billion dollars a year. Incidentally, this is where Comprehensive Health Planning fits in. The National Center for Health Services Research and Development, which some of you have contacted, is another component of HSMHA. The National Center for Health Statistics is another.

In the newly established family planning endeavor, Dr. Louis Hellman is setting policy for the Department. Dr. Frank Beckles, as Director of the National Center for Family Planning Services, has most of the administrative responsibilities in HSMHA.

The Indian Health Service is another HSMHA program, as is the hospital program providing care to merchant seamen and other beneficiaries. These direct care activities account for a substantial number of our employees. The National Institute of Mental Health, HSMHA's largest single component, has a wide variety of programs in research, training, and service. The Hill-Burton program, Maternal and Child Health, and the Center for Disease Control are other constituents of HSMHA.

To present it in simplistic terms, in the organizational structure there is a director for each of these major HSMHA programs who has a direct responsibility for our legal and fiscal relationship to the Department and to Congress.

In addition, included in our programs are some guidance responsibilities that we assume for other agencies. These include, for instance, the Federal Employee Health Service, the medical portion of the Appalachia programs, the foreign programs under P.L. 480. And more recently we have been asked to have a look at the design of the Health Maintenance Organization.



In each of the ten regions, which recently have been slightly reoriented, there is a Regional Health Director. Roughly one-third of HSMHA's resources are now being expended at the prerogative and under the administrative authority of the regional health director.

I answer for these responsibilities directly to Roger Egeberg who answers to the Secretary.

It's an interesting and complex organization. I'll not go further into this, other than to say that the authorities for all of our programs are vested in the Secretary. And most of them, with other than policy impact, are then delegated to the Office of the Administrator and, in turn, to the program directors and regional health directors.

I hope this outline of HSMHA's organization will give you some idea of the perspective from which I will talk about RMP this morning.

The RMP concept has always attracted stimulating and innovative people. This conference is simply another manifestation of this fortunate tendency.

We are at a critical juncture, a decisionmaking point in the health care field generally. There are a substantial number of evaluations going on at all levels and with all degrees of sophistication. Currently, there isn't an effort in the health care field, public or private, that is escaping scrutiny; and apparently no assumption is going to be taken for granted in the foreseeable future.

The Executive Branch itself is engaged in a fundamental reexamination of both the appropriateness and the effectiveness of its health care programs. The Congress itself is entertaining proposals that are enormous in their scope and diversity. And all across the country groups of health professionals, such as this, and individual patients themselves, are weighing the options available to them in choosing courses of action that are now beginning to determine our health care system of the future.

Some of these evaluations, like the three-day session which you have had, are objective and as thorough as the state of the art will permit. Others are very subjective and based only on anecdotes or fragments of evidence.

It's important for you, I think, to remember that sophistication carries no guarantee of acceptance unless we make sure that our input is registered. The naive assessments may be the ones that are crucial to our future.

In the Health Services and Mental Health Administration, we too are deeply engaged in self-evaluation. Roughly one percent of our total expenditures, which are in the nature of \$1.5 billion a year are set aside for evaluative purposes. We are trying to find out what the

scope of our agency's role should be as a responsible Federal agency in health services delivery. And we're looking at HSMHA primarily as an agent for working with the systems of health delivery.

The Federal role in health care has been moving recently from a passive to a more active involvement. As you know, in the past the Federal functions have emphasized limited direct responsibility and considerable use of various kinds of stimulating mechanisms. It is my impression that even when it is stimulating private initiatives, the Federal role in the future will tend increasingly toward setting the terms and conditions under which those initiatives will be carried out.

Now, we in HSMHA have made some assumptions in the early deliberations during my 85 days of this incarnation; and I should like to share some of them with you for your thoughts. These are not dicta but are, I think, bases for departure in the analysis process.

The Federal role should always be complementary to the private sector insofar as possible.

The Federal role must, however, protect the common good where inadequacies or inequities appear in the system.

Maximum effectiveness must be assured when federal dollars are used either as an expenditure or as an investment.

And Federal leadership must assure coordination of efforts and common communication among health activities in both the public and private sectors.

None of these is new. You have all thought of them. Perhaps the difference is that we intend to act on this set of assumptions; and that might make a difference.

In our opinion, the government should help the energy in the health care system to flow where it will do the most good. Viewed in terms of energy flow, we have to look at ourselves in proper perspective. For instance, we have 25,000 employees who contribute 25,000 man-years to the health care system; whereas the practicing physicians of this country account for roughly 300,000 man-years. And the 7,000 hospitals across the nation have a combined energy input of several million man-years.

While there is no such thing as a common unit of health energy, it is evident that our problem is how to make the contribution we have effective in a very large system.

Ideally, the government activity should concentrate, as we have already said, in the areas where the free market is unable, for one reason or another, to fulfill the public need. Such areas tend to occur when the anticipated private return provides insufficient motivation or

Child Health, the Center for Family Planning, and the others.

I'm aware that the relationship among these programs and particularly between RMP, Comprehensive Health Planning, and the R&D Center has been the subject of endless debate since these programs began. Almost everybody in the health field has had a piece of this action. We have had advice from everybody, but the subject still remains.

I'm sure this is one of your concerns. It is one of my highest priority items; so much so, in fact, that we have initiated an intensive administrative study that is targeted to the specific mission of defining separate, distinct identities for these three major programs.

There is an extra special group of consultants who will be functioning in various ways. The work will be coordinated by Dr. William Willard of Kentucky. Dr. Willard will be spending about eight days a month with us over the next several months. His efforts are going to be augmented in various ways by Dr. Monty Duval, Dr. Ed Pellegrino, Dr. John Hogeness, Mr. Nathan Stark, Dr. Julius Richmond, and Dr. Ward Darley. I think those of you who know some of the stalwarts in the field recognize that we really have pulled out the biggest guns we know of to get some administrative discussion of how we can do this constructively and preserve the tremendous promise of each of the programs.

In a generalized way, the shape of the distinctions can be deduced from the terms in which these programs were originally framed, at least as I understand them.

RMP was originally conceived as a bridge between human need and scientific advance, if you put it in simplistic terms. It represented in a sense a practical attempt to link C.P. Snow's "two worlds," which may be somewhat out of date now, but, nevertheless is what was in mind.

The requirements of the individual patient were to be better served by creating arrangements that would enhance the flow of greatest expertise to the patient's bedside through an effective linkage of the providers—and I should like to emphasize the effective linkage of providers.

It did, as we have already said, give providers an opportunity to innovate from grass-roots ideas.

Comprehensive Health Planning approached the same ultimate objective from a different angle of attack. Here, by fostering planning processes at the State and community level, the intent or at least the greatest promise seemed to be to encourage a political consensus, in the broadest sense of a political consensus, as to health goals and the use of health resources.

The planning agency had its greatest promise as a voice of the people in the political sense, enunciating the providers the public determination of needs and priorities.

It has a geopolitical responsibility to assure to constituents equity of care at the highest possible level of quality through the instrument of planning.

The R & D Center, the newest member of the triad was envisioned as an experimental instrument applying scientific disciplines to the model of the health service delivery system in the community. Hopefully it was to be a generator, tester, and evaluator of innovative approaches in the system, addressing itself to such things as cost containment, equity of access, and efficiency of resource utilization.

These philosophical differences, however satisfactory or unsatisfactory they may be in the intellectual sense, haven't provided adequate guidelines for practical distinctions in the health delivery system of the real world.

It is imperative that this situation be clarified in such a way that we maximize cooperation and minimize the overlap and confusion among these programs. Unnecessary duplication, with its resultant waste of effort and money is intolerable. In fact, it is destructive in the face of a limited budget and an unlimited need for improvement in health care.

The effort at clarification to which I am assigning top priority is not to be construed as competitive. We are not talking about one program versus another. I view it as essential if we are to justify and obtain continuing and productive support for all of the efforts of HSMHA, wherever they may be.

If you are concerned about administrative arrangements at HSMHA headquarters that have an important bearing on the conduct of your RMP activities across the country, I am sure that you will make it known to us in whatever unrestrained or restrained manner you have in the past. We solicit that kind of interest and input.

I have made a fairly fast attempt in a short period of time here to sketch out for you in broad strokes some of the dimensions of the broader stage on which your individual programs are enacting an important role.

RMP is an integral and extremely important part of HSMHA. HSMHA, in turn, is the agency charged with exercising appropriate federal stewardship in DHEW for health care delivery. It is not simply a collection of programs; it is a composite. And each of its components is to contribute to a common mission.

The test of our performance, yours and mine, will be whether or not we can apply our combined leverage so

where the scale of investment required is beyond the capability of the private resources.

I'd like to give you a few illustrations of fields where HSMHA believes it can perform a valuable service to the system as a whole.

As a health care delivery agency of the federal Government, we can meet one urgent national need for a central source of valid and creditable information on the delivery of health care.

Now, I don't know how many of you have really gone at the business of looking at our performance in the health care field. But it turns out that any set of data that you pursue far enough seems to come back to the National Center for Health Statistics, often somewhat mangled in the process, and discouragingly far from anything that tells us about the health status of the nation.

It tells us a little about the absence of disease, but not a great deal about the health status of the nation.

Somewhere there needs to be a competent group of experts to sift through the diverse health care activities that are conducted in our many localities under many different auspices, and to analyze and summarize the national experience.

To the best of our knowledge, no such source really exists that is capable of providing validated information on multidimensional and multidisciplinary questions on health care delivery.

In my opinion, efficiency dictates a single central source, and logically this is the role which the federal Government should do as it has already done in agriculture, commerce, and to a substantial extent in biomedical research through the NIH.

It is our intent to become the locus of such an activity. Interestingly enough, we have the mandate. In a substantial number of our programs, including Regional Medical Programs, Comprehensive Health Planning, National Center for Health Statistics, and National Center for Health Services Research and Development, we even have the models. We have the instruments, and we have the capacity.

The role represents one way in which this small energy input can help direct the flow of a larger system.

A second role to which we might aspire will be characterized as a kind of guardian of the nation's standards in health affairs. And I'm not thinking here of regulation as much, although this may apply in a few cases, such as in quarantine; rather I am thinking of an evaluator of performance—which is exactly what you're doing here today—and an activator of the public conscience.

When it becomes apparent that a given segment of the population—for example, expectant mothers or migrant workers—is not receiving the kind of health care it has a right to expect, someone has to be responsible for setting this forth in clear terms and making it a part of community thought.

Someone has to begin a stimulation process to a point that the system will respond. This does not imply direct action in terms of meeting the need, although we are involved in some of that, but I think more importantly involved is getting the selected endeavor into realistic discussion.

This function will have in it at least two phases. The first is a continuing systematic and sophisticated overview of what the health care system is doing, projected against two grids—what it could do and what the needs are. At this time we don't have really adequate surveillance of performance, capability, or need.

The second phase involves getting something done about it. And certainly from our relatively small fiscal base, we will have to look at communication and persuasion rather than direct entry into meeting the need itself.

The tools at our disposal then are going to be communication, persuasion, selective encouragement of innovation. And that's the name of RMP as far as I'm concerned.

We need to use that instrument well. One instrument of stimulus for improvement can come from RMP's functioning as a center of expertise; and this is the instrument of information display in which we hope you will join us.

I think all of you are aware of the fact that when a company's stock is performing badly, this is made pretty clearly visible in the daily listing on every financial page. And general knowledge is a powerful spur to self-examination and change in those whose stock is not doing so well.

In health care performance, the criteria may be a little harder to define and the comparative information harder to acquire, but once acquired and displayed, it could and should have a similar effect.

We have some other instruments for change, programs that are explicitly designed to stimulate innovation in health care delivery and effective synthesis of health resources for the benefit of the patient. This, of course, again includes RMP.

It also includes the planning and project support activities of Community Health Services and the other activities we have talked about such as Maternal and

the total impact of our efforts is greater than the sum of the parts.

As you go through this evaluation conference and your further evaluative efforts in RMP, you need to do so in this broader context. Your evaluation efforts should be an important input for ours. The results of our evaluations, in turn, must be factored into the equations which involve the total health care system.

For this reason, it is important that we make our findings and your findings widely and freely available. Just as communication is one of the strongest instruments for change within the system as a whole, it is almost our only instrument for change. It is going to be effective in proportion to our use of the evaluation process. If we don't make known what we know, we will have no cause for complaint if we are not a part of the future.

Finally, in all of our evaluative activities it is imperative we keep in mind the ultimate objective of our endeavors—that what happens to the patient or pre-patient is really what we are supposed to be concerned about. That's the hardest evaluation of all.

The one thing we still lack is the measure of health as an ultimate yardstick.

In the same area, we're dealing with the health care system which is still a crisis-oriented system. It pays least attention to first things—health maintenance and disease avoidance—the greatest attention to illness after it has occurred.

We need to be sure, if we are thinking truly about serving the public both present and future, that we are not similarly distracted in the planning process.

## WORKSHOP SESSIONS

### WORKSHOP ON DATA

#### Participants

Arthur R. Jacobs, M.D., M.P.H. -- Moderator  
Director of Statistical and Evaluation Unit  
Rochester Regional Medical Program

Jarah J. Peterson -- Associate Moderator  
Biostatistician  
Fabius, New York

Charles A. Metzner, Ph.D.  
Professor of Medical Care Organization  
University of Michigan

J. William Gavett, Ph.D.  
Associate Professor, College of  
Management and Department of  
Preventive Medicine and Community  
Health  
University of Rochester School of Medicine

David E. Reed, M.D.  
Assistant Director for Evaluation  
Western Pennsylvania Regional Medical Program

Katherine G. Bauer  
Research Associate  
Harvard Center for Community Health  
and Medical Care

John E. Wennberg, M.D.  
Coordinator, Northern New England Regional  
Medical Program

---

#### The Values and Limitations of National Data

##### CHARLES A. METZNER

A short presentation on this large subject can only sketch the topics and arguments. The attempt to be short results in more direct and unconditional statements than are strictly warranted, but this may be the basis of the discussion to follow, although my aim is not to be deliberately argumentative to stimulate controversy. I shall try to elucidate problems and lead toward some useful conclusions. Explanations are not complete, either, but questions, if necessary, can elicit more. What I am trying to do is to stimulate thoughtful consideration.

Censuses are not new. In fact, there is Biblical mention of a census and the ideological response to it then still has repercussions among fundamentalists. One is reminded also that total counts are sufficient - Gideon became famous by applying a behavioral test to select a subset with characteristics he wanted. Now many characteristics are incorporated into census data. The attempt even at health data is not utterly recent, however. In 1870, the United States Census became very ambitious and, among many other data, tried to obtain information on illness. The procedures were somewhat crude, but the amount of data of all kinds was so voluminous as to threaten the decennial census by taking over ten years to process. This is the point at which the mother of

invention enticed Herman Hollerith to father punched card procedures for mechanical data processing, which now make possible, particularly since electronic procedures have been substituted, the derivation of so many tables that it is hard to find our way around in them. One additional historical point may be interesting to you. It was in 1942, a relatively recent date, that "A New Sample of the Population" was developed, which embodied the first practicable methods for probability samples of human populations. It may be worth recalling that these area sampling methods were a product of the WPA, and later incorporated into the Bureau of the Census. Sampling enabled many more data to be generated at much lower cost when estimates are sufficient.

##### Sources

There are some guides to data that are useful. The Statistical Abstract of the United States presents an annual overview of data, with references to sources. It is a good index to availability. It is mentioned (on page five) in what should become a basic reference, the National Center for Health Statistics' short pamphlet, "A State Center for Health Statistics: An aid in planning comprehensive health statistics". (Revised October 1969.) It is available from the Center or the U. S. Government Printing Office. Among other items, the chart on page

11 on input-processing-output relations describes roles that may conflict and useful advice is given for handling these. In particular a number of user-designer problems are considered. On page 13 is a discussion of the use of computers and the necessity for thought that makes all else commentary. A rich passage, deserving expansion, occurs on page 14 with respect to cooperative relations between users and suppliers of data.

Some state agencies have been developed, and many health and planning departments generate data that should be looked into. As mentioned later, the more local the data the more specific the estimates that may be derived.

As an introduction to problems, another publication of NCHS is valuable: The 1970 Census and Vital and Health Statistics. A Study Group Report of the Public Health Conference on Records and Statistics. Documents and Committee Reports. PHS Publication No. 1000 - Series 4 - No. 10. Government Printing Office, April 1969. This is a planning volume for the 1970 Census, still useful on issues.

### Problems

National data involve many kinds of problems. In common with other data, becoming knowledgeable involves not only the names of variates but definitions, particularly embodied in a questionnaire, the instructions, and codes - in short - all processes which shape the final product.

There are some special issues concerning terms and definitions that arise in a nation like ours. Some of this may be easy to see nationally, but you should not be too certain that this applies only to someone else. Ours is a pluralistic, individualistic society, with plural health care systems. A single basis of definition does not encompass all. Ordinary classifications, such as the "International List", assume an M.D. etiological base, largely micro-biological. Because we have not recognized the ways in which other people live and think, we are being tested again concerning some accommodation to multiplicity. How far are we willing (or able) to go?

Would we accept a voodoo health center? The question is put in this form to test associations. Since the audience is more or less white, and more or less Christian (although perhaps not up to the standards of Dr. Martin Luther King), we are inclined to be shocked but accept the racist implication that this would be a black enterprise. We should examine our readiness to accept the implication. California does pay faith healers, Christian Scientists do not get the diseases of the International List, and Jehovah's Witnesses do not accept all

of the ordinary beliefs concerning blood transfusion. Demoniacal possession is not included in many disease classifications, but you should check with your own religious leader as to what he thinks your beliefs ought to include. It is my understanding from nutritionists that the usual concept of what constitutes a "bland" diet has little more to recommend it than the idea of "hot" or "cold" foods.

The point should now be made that we have varied subgroups in our population. The discussion may have seemed farfetched and rather distant from data. I believe I can make multiple use of the ideas, however, in trying to follow the implications of a pluralistic society for a) the generality of findings, b) the generality of concepts used, and c) the necessity, flowing from these for greater freedom of research, particularly in government.

### Implications

#### *The Generality of Findings*

From the fact of high variation within our society, it follows that national data have little specificity and high variation. Certainly, the mean outlay per person for health care times the number of persons equals the economic load on the private sector for health care. But the mean is not very representative. Not very many people get the mean income. The standard deviation is so vast as to encompass most of the information and render the mean almost meaningless. If you look at any distributions in the health field, you can see this. This is one reason why insurance is so important a mechanism for achieving a mean value.

There are several ways in which this is directly important. The variations in national data can and do spread between Census regions, states, and regions in the RMP sense. National data do not necessarily represent your area, to the extent that your area is distinct. National data must be sorted to yield data for your area, which may be done, and cognizance should be taken of the fact that the Census is willing to do this, and that one large study is cheaper than several small studies. However, for sample data, the results may well be unreliable for small areas. And equally important, while sampling distributions for the estimates may be derived, in general the common statistical tests (t, chi-square) are not valid for these statistics.

A frequent procedure to "adapt" or "derive estimates" for a particular area from data for another is to analyze the data by other variables (age, sex, previous condition of servitude) and use regression or standardiza-

tion techniques to estimate values for our specific distribution of the analytic variables. Unfortunately, the residual variation also applies, because the analytic variables can only transfer as much of the variation as they absorb, and what remains is error, in both cases. For health data, our analyses so far do not account for much variation, and the estimates are correspondingly poor. The fact that it is done with mathematical statistical formulae on a computer will not improve it. More complete statistical analyses or a mathematical presentation would verify the logical argument above. Any statistical text dealing with multivariate techniques will explain this.

### *The Generality of Concepts*

Again, there are several limitations encompassed in the generality of concepts. The first involves communicability and response. Respondents understand and report only those terms they know. And much knowledge of what is or was wrong with us comes from the health care system. This is a feedback process of some importance. We learn as we use, both in terms of diagnosis, which must be given us to be at all reliable, and even in terms of recognizing what symptoms are important by how health professionals respond. What this implies is a strong bias against reporting by those not habitual or economically enabled users. The problems of non-users are not reported by a system assuming use, and the resulting confounding conceals problems of the system. When a symptom list is used, it will help those who recognize the symptoms, but it will not elicit a misery or a devil bothering a respondent unless it includes these. It is much easier to adapt to this locally, since many terms are regional, although they may be ethnic or status-related also. To check, find out from your friends from different parts of the country when evening begins for them, and you will get some idea of the problem. At any rate, the reliability and validity of data are high only for those using the system generating the concepts, and may cause serious under-reporting of illness and the unorthodox treatments engaged in by those uninvolved with orthodoxy. Of course, only a national study can demonstrate all of the variability, but then only if they are prepared for it. If one wants to find out about problems, one must be receptive.

A second issue is the problem of "general purpose" data to be used by many. Of course, agreement on what information to get is a political and economic necessity, but however valuable compromise may be in politics, it

does not settle conceptual problems, at least correctly. To settle an issue of the best, which is to say most predictive or homogeneous, definition of what is an epidemic or what constitutes group practice, we have to try them all and find what difficulties ensue or what utilities be in each. Frequently, we are forced by circumstance into premature definition which is copied and standardized. Sometimes we just pick up a handy classification, as in the case of health studies using the International List, reflecting etiology. This classification is no doubt valuable for the practice of medicine, but it does not resolve (or predict) the use of services which forms the basis for manpower and cost studies. At least, some concept such as the seriousness of the illness must be added when the fiscal or personal impact is what we really desire. Much more must be done to develop concepts suited to purposes. And this leads to the concluding issue.

### *Freedom of Research*

The argument thus far culminates, I believe, in a plea for greater support for many kinds of data and for research more nearly directed toward well specified problems. Much of this may best be done in the locality of a problem where the distinctive character of the situation may be seen, although without effort and receptivity there is nothing to warrant a belief that being next to a problem ensures noticing it. Most people with glasses do not report any disability, and it is hard to convince people that they are deaf.

Mainly, I believe, it is necessary for our national policy to incorporate the fact that to encompass the variety and subtlety of our national life entails independent thought and effort and the development of queer and unpopular ideas, and mistakes. Our affluent society does have people suffering from hunger. We must acknowledge that we do not with any certainty know how to interrupt the transmission of poverty from generation to generation. Uncovering the hunger implies allowing studies, and particularly analyses. Discovering procedures for bringing ghetto dwellers fully into the society or organization into the health system necessitates evaluated experimentation. But we all too frequently constrain those with the information from using it for analytic monographs, and insist that a problem be fitted with a single agreed-upon solution. Diversity in the society must be matched by diversity in approach, conceptually and operationally.

Our national agencies are producing many good and useful data. If they themselves, who know a number of

the weaknesses better than those with second-hand acquaintance, were allowed to use them to draw conclusions, we might do better. They are willing to meet us more than half way, though. There are many special analyses that may be obtained, if we ask, and although they will not be free, they are less costly than special purpose studies. The Census Bureau will, for example, design samples for us using their rich data base. Within the limits of confidentiality, information on special groups may be obtained.

National data can be exceedingly useful, but they are no panacea. They are not universally applicable, they are not fully analyzed, and they do not serve all purposes. We must consider the limitations in the light of our objectives, and we may thereby help to eliminate some limitations.

### Data for Ambulatory Care Planning

J. WILLIAM GAVETT

American communities are concerned with the inadequacies of existing primary ambulatory services, but do not have quantitative data necessary to plan alternative systems for the delivery of ambulatory care. Studies of primary ambulatory care are relatively new compared to studies of hospital care. New techniques are needed to evaluate existing, as well as proposed facilities, for the delivery of primary ambulatory health care. The existing facilities include: private practice (solo and various forms of group practice), occupational health services, school health services, hospital emergency departments, hospital out-patient departments, neighborhood health centers, health department clinics, as well as various state and federal primary ambulatory services. Proposed models include facilities differing in manpower, financing, and utilization patterns located in different areas under private, voluntary or government auspices.

Variables to be studied might include legal, contractual, and business arrangements; availability, accessibility; degree of specialization vs. generalization of services; consumer payment mechanisms; reimbursement for services mechanisms; manpower configuration; equipment; ancillary services; capital, financing arrangements; and characteristics of services rendered. Within the context of defining the basic characteristics of primary ambulatory care organizations, consumer attitudes, outcome of care, and design characteristics such as: working spaces, procedure and communication systems, etc. are less important.

The purpose of this article is to consider the relationships between patient classification, data collection, facilities and manpower utilization for ambulatory care. Before doing so three caveats are offered:

1. The design and implementation of a data system for ambulatory care should proceed concurrently with the development of hypotheses about the planning and organization of such care. The data will provide the decision maker with the necessary statistical information for evaluation of alternative ambulatory care proposals. Unless hypotheses about changes are offered prior to simultaneously with the design of the data system, the latter effort may be extremely costly for what information is required and used.

2. There is no single decision maker in the typical community ambulatory care system. The community system is typically fractioned and consists of a variety of independent organizations (listed previously). Changes in individual and independent organizations can make sense from a community systems point of view only if there is coordinated and cooperative community planning unless severe legal constraints are imposed in such a manner as to force the consideration of community-wide objectives. The purpose of voluntary "community planning is to provide the independent decision maker (administrators, physicians, etc.) with information that permits a rational evaluation of their decisions relevant to community objectives.

3. Data and information collection must be related both to the consumer of health care (the patient) and to the processes (methods) of ambulatory health care. The efficacy of a data and information system in aiding planning will depend on the manner in which both patients and processes are *classified*, described, and measured. It is on this issue that the remainder of the discussion will focus.

It is suggested that a classification scheme that will relate health care demand to the manpower, equipment and facility requirement is needed. Traditional classification methods include *patient characteristics* (maternal and child health, pediatrics, veterans services), *path physiologic processes* (tuberculosis, cancer, hemophilia), *services rendered* (radiology, medical, surgical), and *organizational setting* (neighborhood health center), as well as *organizational rendering the care* (private group practice, hospital out-patient department, occupational health services). Each of these classifications has some use in health care planning. For primary ambulatory care planning, a classification system is needed to categorize, compare, and project patient utilization for different types of primary ambulatory care delivery units. A measurement



demand in a given ambulatory unit is the first step in the conceptualizing of alternate ways of satisfying demand.

The demand or load placed upon a primary care organization at any given time can be expressed in terms of the number of cases (patients), the episodes (specific medical problems requiring management), and the visit (interface between the patient and the health care system). A classification system based upon a set of criteria related to the characteristics of the case and the visit as they relate to the services rendered is proposed under the assumption that such a classification scheme will facilitate the conception of alternate care organizational designs. This classification method focuses on the complexity of the case and visit.

The primary care setting encompasses a continuous scale of case complexities. At one extreme is the medically simple case in which modest resources (man-

power and equipment) are involved. The other extreme is the critical, medically complex case in which extensive resources are used often within a short period of time. It is suggested that cases might be classified into categories such as *A*, *B*, and *C*, where *A* is the urgent, complex, resource-intensive case; and *C* is the simple case, involving minimal resources. The *B* cases would include those involving long term episodic illnesses where diagnostic skills, continuity of care, complex therapeutic measures, and extended support and observation are required.

Figure 1 represents a definition of each class in terms of specific attributes. These include manpower and facility requirements, frequency of visits for the episodic illness, diagnostic problems and disposition of the case visit. A fourth class might be developed for psychosomatic cases and minor psychiatric cases.

FIGURE 1.—Case Classification Table

	<i>Manpower</i>	<i>Facilities</i>	<i>Frequency of visit for episode</i>	<i>Diagnosis</i>	<i>Disposition of visit</i>	<i>Comments</i>
<i>Class A Case.</i>	MD required	ICU	Not applicable	Extensive Dx	Hospital, home care, or Long term care facility	Acute
	Likely team effort	ED Hospital facility required		Skills required or not required		Life-threatening Totally interrupts normal living
<i>Class B Case.</i>	MD required	MD office or more extensive Dx or Rx facilities Possible hospital	Revisits required	± Difficult	Home Possible hospital Long term care facility	Non-acute
	Possibly referral or consultation			± Obscure Chronic		Ongoing comprehensive and continuity care important ± Interrupts normal living
<i>Class C Case.</i>	MD or nurse	Average MD office Dispensary	One or two short visits	Relatively simple, obvious, self-limited (URI or minor injury)	Home	Acute or not Accessibility and availability of service important to patient Interrupts normal living to minor degree Support and reassurance may be a major attribute

The ABC classification, and further refinements of it to include subclassifications, provides a basis for considering the questions of ambulatory care organization. For example:

1. For a *given community*, what proportion of case visits by *A*, *B*, or *C* type are made to which organizations; are resources allocated among organizations in an intelligent manner, e.g. perhaps demand for type *A* services should be consolidated at one or more hospitals?

2. For a *given ambulatory care organization*, what proportion of case visits are in each of the *A*, *B*, and *C* classes; are the unit's resources intelligently related to the given proportions?

3. How can a rural community, that cannot attract or hold an MD, benefit by an *A-B-C* classification of its ambulatory patient load? e.g.:

a) Could class *A* cases be serviced by volunteer community-supported emergency units, highly trained to provide on-the-scene first aid and transportation service to the nearest intensive care unit (presumably located centrally)?

b) Could class *C* cases be treated at a private or community-supported convenience clinic, manned by paramedic personnel, and organizationally linked to the nearest community hospital or group practice?

c) Could class *B* patients be provided with long term episodic care by nearest physician (patient's choice) but with routine and non-complex class *C* visits serviced by the convenience clinic?

4. Where in the management of *A*, *B*, and *C* cases are community-sponsored facilities advantageous in the larger community? For example, what community organizations should sponsor multiphasic screening, convenience clinics, special preventive medicine clinics, etc.

5. Can a clinic for the treatment of *C* cases be effectively used also for the purposes of triaging non-*C* cases to other community health care organizations for those individuals who do not have access to other primary care organizations?

6. What proportion of ambulatory case visits classified as *C* cases involve mainly support and reassurance?

7. How are the concepts of family medicine and comprehensive care relevant to the class *A*, *B*, and *C* cases?

8. How is the question of the use of paramedic personnel specifically related to *A-B-C* case care? Does the *C* case and *C* visit load on the community consume significant physician resources such that extended use of paramedic personnel is justified?

9. The case classification technique may reveal which variables are important and should be incorporated into

ambulatory care data systems for patient care, institutional management, and for community planning.

10. What is the role of the hospital in ambulatory care? Analysis of the Emergency Department and Outpatient Department by case classification characteristics may provide quantitative data for reorganization of the hospital's ambulatory services.

11. Does the measurement of low income urban ambulatory care demand by the case classification technique provide insights as to how to organize ambulatory services for the urban poor?

### Information Systems to Meet Common Data Needs of Health Agencies

KATHARINE G. BAUER

It has been observed that information is to the decision-making process what oil is to the internal combustion machine. It does not itself make the process work, but without it there is considerable wasted effort, misdirected motion, and eventual breakdown.<sup>1</sup> Those who are at the wheel in making health policy decisions usually find themselves in the position of the motorist with a dry engine in the middle of a Texas oilfield. The million barrel output of raw material surrounding him is useless to meet his urgent need for a mere two quarts which have been suitably processed to meet his engine requirements. We would all agree that health data gush more freely than oil — and that for the most part we haven't yet found very satisfactory ways of tapping and refining it for the particular uses of those who make health decisions — whether these involve expenditures of thousands of dollars, or of millions of dollars.

My assignment today is to discuss the organization of a health information system as a means of meeting such important needs, particularly those of RMP evaluators and their opposite numbers in other agencies. Can such a system be designed to supply, link and refine the main streams of health data that are routinely being generated from diverse independent sources — such as the facilities manpower, and vital statistics compiled by State agencies, the various utilization and patient origin records from hospitals and other service providers? And can these be more usefully related to the basic demographic and health statistics from the U.S. Census and the National Center for Health Statistics? I was asked to lead off this discussion by virtue of my association with a two-year study of this question at the Joint Center for Urban Studies of M.I.T. and Harvard — a study largely inspired by Dr. Osler Peterson, Director of Research for

the Tri-State RMP, who made major inputs to it at every stage.<sup>2</sup>

What do we mean by the term "health information system?" I suspect this is one of those in-terms that people have come to use quite widely without benefit of definition. For our purposes today, I'm simply going to be generic and talk about a systematic approach to producing, storing and gaining access to many kinds of health data produced from many sources, for multiple uses, by multiple users. Also, in an effort to put first things first, this paper focuses primarily on the organizational process for systematizing this access to data — only incidentally on computers. As an analogue — if we didn't yet have libraries but many writers were producing books which many potential readers needed to read, a first step would be to organize some system to identify the books of interest and to decide on policies for their acquisition, storage and circulation. Only as a second step would one commission a computerized index. The health information broker system recommended by our Boston study naturally differs in many respects from conventional library organization — yet there are underlying similarities of function.

The broker system predicates that it would be mutually beneficial to a region's major public and private health programs and agencies, such as RMP, Blue Cross, State Health Departments, comprehensive and area facility planning agencies, to join forces to obtain and share the kinds of data they need in common for their separate research and planning activities. At the same time, the study warned against constructing unworkable multi-million dollar data banks. Before describing this model, and telling you some of the considerations that influenced its design, let us briefly review some of the reasons it seems particularly timely right now to promote this or some other type of cooperative organization for improving health statistics.

#### *Why a Health Information System?*

As budgets in every sector of the health system get tighter in the face of medical price inflation, it seems certain that in every type of program, public or private, the big questions of accountability raised to you yesterday in the plenary session will be increasingly posed: what benefits are patients actually receiving for the money spent? How can the program policies be modified and adapted to improve these cost-benefits? Obviously the day is almost over when those who pay the bills will be satisfied by simple tallies of patient days and O.P.D. visits juxtaposed with total dollars expended and a request for a 15% budget increase next year.

This means that throughout almost all health programs, not just RMP, researchers will be trying to construct various types of performance indicators — to permit comparisons of past and present experience within a program. To measure the impact of their program on specific target populations over time, and to compare their program results with those of other programs which use other techniques or methods. However, as we all know, the right kinds and quality of data are rarely available to permit this crucially important research to be carried out. One can make a safe guess that not only throughout our concurrent workshops now, but in similar health evaluation research meetings everywhere, the identical complaints are being voiced: "The 1960 census data were obviously useless for computing 1969 rates — we simply can't tell the trend so far . . ." or "Unfortunately the reporting system changed, so it's impossible to compare past and present performance" or "we can't compare our results with those of program x because they used entirely different age breaks — and besides we have no way to get comparable unit costs." One concludes that all concerned have an enormous stake in improving the kind and quality of the data base.

To provide the denominators of the rates they need for their various purposes, researchers in all major health programs seem almost universally to require certain common types of data — the demographic, health status, vital records, facility and manpower and the kinds of utilization data reviewed here earlier. Some of this simply isn't now available — such as disability rates of populations in cities or small geographic areas. Other widely needed data, however, such as about health facilities, are being routinely generated for their own operating or management purposes by some one agency which, in turn, may need management — generated data from other agencies for its own evaluation research.

Finally, staffs in different agencies quite often duplicate their research efforts, both in their separate quests for identical source materials, and in time-consuming activities such as constructing S.M.S.A. profiles, or population projections. This costs everyone money.

Given such common needs and problems it would seem that major health organizations have everything to gain by joining forces at least for the limited goals of:

- improving the quality and comparability of existing data commonly shared,
- identifying commonly needed data now unavailable, and finding means to secure them,
- eliminating duplications of research effort,

- arriving at agreements for specific types of data sharing.

Although funds were not available last year for a proposed demonstration of our Boston model, it seems possible that within the next few weeks Congress will authorize federal support for experimental health information systems of this kind as part of the Health Services Improvement Act of 1970.\*

Funding is only one aspect of the problem of data sharing among independent organizations. Given the realities of the operating environment, can a satisfactory means be found to promote inter-agency cooperation? When one looks at the activities of the Bureau of the Census and of the N.C.H.S. and other important information centers at the federal level one can feel hopeful. But further down the line at the regional, state and sub-state levels where mixes of various public and private data sources are sought, issues of agency confidentiality and of inter-agency power struggles inject a host of complexities. Whether organizational forms can be devised during the next few years to circumvent the problems while fulfilling the need remains to be seen. Our Boston study's recommendations represent one possible approach — Dr. Wennberg will tell you about another, and I know that several other people here have been wrestling with these problems in their own regions — from New Mexico to Ohio.

#### *The Broker System Model*

The Boston study concluded that (and I quote):

“The needs of health planning and research in this area at the present time will best be served not by a new prime data processing computer system, but by a mechanism designed to interface between several newly developing hospital, public health, mental health, and social welfare information systems at regional, state, metropolitan area and municipal levels. Such a mechanism should promote compatibility between the subsystems and thereby maximize the possibilities for mutually beneficial information spin-offs, now and in the future. A consortium of health planners, major health agencies and research organizations should establish a health information system to serve this broker function, to facilitate the

development, sharing and use of information pertinent to their common needs. Such an information system should be planned at the outset as the first step in a more complex communication network should future expansion seem indicated.”

A broker function between independent health information subsystems rather than a centralized data bank was recommended because it would:

- adapt better to the predictably ever-changing data needs of its users,
- provide better quality information over the long run,
- avoid direct confrontation of the issues of agency confidentiality and of individual patient privacy,
- function better within the present limitations of computer software-yet permit adaptation to future technological advances expected there.

Finally, a consortium of users was recommended as the policy-making body for the broker system, with administration temporarily vested in a university. This structure was put forth in order to avoid threatening the existing power relationships among agencies sufficiently to foreclose their participation.

Before going on to elaborate on some of these points I'd like briefly to mention some activities proposed in the Boston model.

#### *Some Possible Functions and Activities of a Shared Information System*

1. Making data more available for secondary analysis by:
  - inventorying and cataloging data sources and files
  - furnishing detailed descriptions of data files to guide the user — such as dates and methods of data collection and up-date; sample size; format in which preserved (file folders, magnetic tape, etc.) person responsible for maintaining files; conditions of access, etc.,
  - guiding and helping the user select and use computer programs best suited to his needs.
2. Improving the utility of available data by actively encouraging data generating agencies to arrive at:
  - compatibility of key items on report forms — such as age, residence, condition, service, etc.,
  - compatible definitions of terms used in such reporting.
3. Identifying common unmet data needs, and helping meet them by:
  - promoting addition of new categories of information in existing data sources — such as finer age breaks in a State census,

\*“The Secretary is authorized, directly or by contract, to undertake research, development, demonstration and evaluation, relating to the design and implementation of a cooperative system for producing comparable and uniform health information and statistics at the Federal, State and local levels.”

- developing directly, or contracting to develop, new sources of information – such as population health surveys.
4. Helping users find the computer resources most appropriate to their needs by:
    - organizing local conferences and workshops – such as are now conducted by the census,
    - inventorying and brokering use of agencies' partially idle computer hardware,
    - evaluating software packages, and purchasing for joint use,
    - demonstrating, through case examples, the uses and benefits of new advances in computer science.
  5. Developing policies regarding privacy:
    - formulating policies governing agreements for sharing data,
    - promoting codes of ethics; specific legal safeguards.
  6. Furnishing routine monitoring and special status reports such as:
    - trends in the locality's death rates, health facilities, manpower, utilization, etc.
    - comparisons with other regions, states, etc.
  7. Promoting the integration of separate streams of data by:
    - negotiating agreements between agencies for data sharing,
    - advising on legal matters and computer locks to safeguard privacy,
    - conducting file merging operations and providing tables and maps – such as county, city or census tract profiles showing health status, mortality, the population's use of hospitals and health resources – and utilization profiles according to service, patient characteristics, conditions, and proportion of community served.

It is assumed that any such information system would build in its own evaluation process and would continuously re-cycle on the basis of experience, new health research and planning needs, and new computer technology.

You will note the heavy emphasis placed on staff activities. One important thrust of their work would be to inventory and catalogue data sources and computer hardware resources in the region, and to evaluate computer software packages. Another set of functions would relate to improving the quality of the data, by negotiating format compatibilities, and promoting adoption of common definitions. Again, staff would help negotiate inter-agency agreements for data sharing,

and promote common efforts to contract for or in other ways gain access to commonly needed new data, such as from small area population health surveys. Finally, the broker system staff would provide direct research services, such as file-merging operations, and would furnish regular monitoring reports on health and social indicators requested by users. However, it was assumed that the system would not require its own computer facilities at least in the foreseeable future, but would contract for the use of the necessary resources.

#### *Why a System of Sub-Systems?*

A coordinating mechanism between independently organized information sub-systems rather than a central data bank was dictated by users' requirements for flexibility, quality, and privacy – as well as by the state of computer art. I will touch briefly on these points.

*Flexibility.*—Health researchers need to tap data flowing from many sources. Although much of it comes from the operational and management reporting systems of institutions and programs, it is important to remember that despite the overlaps between the specific types of information required for good research and good program management, there are usually marked differences in the characteristics of the data required for these different purposes. For example, instant on-line inputs and retrieval are hardly necessary to provide data for studying the effectiveness of appointment systems in following no-show cancer patients, yet can be invaluable for actual appointment scheduling. Above all, the particular characteristics of the data a researcher needs usually changes with every new problem he addresses. For each study he may need not only different types of data, but different geographic breaks, frequencies of data updating, degrees of individual patient identification, etc.

At an even higher level of generality, maximum flexibility is imperative in a system designed to serve the information needs of policy makers. There can be no fixed solutions to the problem of providing health information since both needs and solutions are dynamic and ever-changing. Many methods of care and facilities for treatment we regard as essential today will be obsolete or unnecessary ten years from now. New methods of payment will be adopted. New health professions will emerge. Information to serve research and policy makers must therefore, above all, be designed to anticipate and to accommodate to change. A network of sub-systems permits this.

*Quality.*—In view of the massive data base required and the large number of files that might need to be

tapped for all the various types of health delivery system evaluation that might be desired now and in the future it would be sheer fantasy to expect that any one centralized system could incorporate them and manage their updating and quality control. Nor would that be progress. It is far more desirable that each organization have a genuine and active concern within itself to continually improve its own information management, while taking due cognizance of the needs and requirements of others.

*Privacy.*—The privacy issue was another major factor in recommending a broker system where every agency would maintain custody of its own files. Clearly, data sharing is an area fraught with fears and ambiguities — where the power of information can be used on individuals and institutions alike for good or ill. And where the conclusions as to what is good and what is ill depend very much on who is making them, and under what circumstances. Or, more succinctly, whose ox is being gored. Confidentiality of information about institutions and organizations relates clearly to issues of the confidentiality of their actions and effectiveness. The Boston study, as the better part of valor, resolved these issues by recommending data be limited to that which could be used in aggregated form, and by promoting specific inter-agency agreements on data sharing designed with appropriate legal consultation. After the system had proved itself and appropriate controls designed, moves might be made towards more specific sharing of fine-grained data.

*Computer Limitations.*—A huge, centralized data system incorporating many files presents problems not as yet adequately solved by computer science. With long lead times for design, by the time such a system goes into operation it is apt already to be behind the state of the rapidly changing computer art — to become a vastly expensive antique. Such disasters have occurred regularly in the urban information systems so hopefully installed in the late 60s. The M.I.T. computer scientists on the Boston study recommended instead, careful development towards a network structure among participating programs, where hopefully in the future a variety of computers of different types and sizes, with different hardware and software configurations might be able to talk to each other under the control of appropriate permissions. They expect that the next decade may well witness revolutionary software and hardware breakthroughs to make this possible.

#### *Who Plays the Role of the Broker?*

Undoubtedly this is at once the most sensitive and the most crucial question to be faced in implementing

this model or any other cooperative health information system. The answer will determine whether the system ever actually gets funded and into operation; whether those who generate the needed data will, in fact, contribute to it, and finally, whether it will truly serve the purposes of research and policy guidance for which it was designed.

The National Center for Health Statistics, which as you may know has recently published a description of a model for state centers for health statistics, states as a cardinal premise the absolute need for information that is completely unbiased and authoritative. I quote: "The inevitable disagreements on how to deal with health problems must not be confounded by controversy over the basic facts of the situation. . . This also means that no pertinent facts be suppressed. . . In effect, the statistical function must be discharged with high competence and cannot be captive to a particular point of view." Thus the N.C.H.S. model calls for the information system to be administratively independent of any one planning agency, though with strong working relations with all.

But how does one identify an administering agency which will command the trust and respect of all, in an environment where knowledge is indeed power — and where, in almost all programs, worry about loss of power is the name of the game? If there is an answer at all to this auspices question, I suspect it will be a different one in each region or state. Some possibilities to be considered are:

- a generalized state statistical center,
- some other state agency (possibly the university),
- a regional commission or center,
- a quasi-public information authority.

In addition to auspices, many important questions of staffing and function, of cost, and of the cost-benefit of such information systems remain to be explored. If Congress does now authorize the funding of experiments in cooperative health information activities perhaps you can all soon begin learning by doing. Certainly the failure to develop satisfactory efforts along these lines can only mean the continued burden of handicap to those who try to measure the successes and failures of our operating programs and thus to give the public the most value for its health dollar.

---

<sup>1</sup>Peter J. Henriot, "Political Questions about Social Indicators," *Western Political Quarterly*, Vol. XXIII, No. 2, June 1970.

<sup>2</sup>Moynihan, Beshers, and Cydel. *Problems and Perspectives in the Design of a Community Health Information System* (U.S. Public Health Service Contract PH 110-234), Joint Center for Urban Studies of M.I.T. and Harvard, Feb. 1969.

## The Northern New England Regional Medical Program Health Planning Data Base

JOHN E. WENNBERG, M.D.

Those of us involved in Regional Medical Programs are continually reminded that health planning, without an adequate data base, is more of a visceral than a cerebral process. We are often asked to support solutions which clearly are proposed without proper identification of the problems involved and are usually made without reference to other priorities. To a very large extent, our informational base for planning decisions in health is limited to impressionistic, non-verifiable opinion commonly arranged or provided by parties advocating a particular solution. Under these circumstances planning decisions run a high risk of being - at best-irrelevant.

A prospective population-based health information system appears as a particularly attractive solution to our data problems. As you have heard today, the necessary technology is not obscure. In fact, it has been available for some time. Why, then, does it remain generally unimplemented? The reason for this cannot be simply cost; it can be convincingly argued that health information systems could more than pay for themselves by providing the informational base for wise decisions. A more probable obstacle to establishing prospective information systems derives from the direct lack of utility of the system to the provider of the data. To provide accurate data is a bother and the effort to produce it must be either rewarded or required by law. Under our existing pluralistic planning and management systems, good planning is neither strongly rewarded nor required by law. Under these circumstances the establishment and maintenance of prospective health information systems are expensive - probably intolerably expensive - in terms of currently available management and persuasive energies.

If the current Regional Medical Program and Comprehensive Health Planning legislation does not contain the mandate necessary to promote prospective health information systems, is there an alternative approach which can begin to achieve the data base necessary for planning and management systems? I think the answer is a qualified Yes: under certain circumstances, Regional Medical Programs can establish an ad hoc but systematic data base which minimizes administrative inconvenience to participating institutions and is at the same time highly useful to its own planning and evaluation purposes, to Comprehensive Health Planning and other health agencies. In addition to the immediate utility of the data, establishment of ad hoc systems affords the

opportunity to accumulate experience with the technical and management problems of developing large data systems. It also allows one to evaluate the utility of components of the system. This should be of value to the future development of prospective, population based health information systems which, I think, clearly will be given central roles as part of the management structure of a national health insurance system.

The immediate purposes of the Northern New England Regional Medical Program data base are to provide information for health problem identification and program planning, evaluation and management. It supports planning efforts at the areawide and state health planning levels. A primary customer of the system is therefore the Vermont Comprehensive Health Planning Agency. Contractual arrangements have been made with that agency to supply them with necessary information. The data base also supports planning and operating activities of the Regional Medical Program, including primary care activities and disease control and continuing education programs. Finally, certain features of the system have been of use to operating health agencies and in some instances to planning agencies outside of our area. For example, aspects have been utilized by Vermont Planned Parenthood, The Province of New Brunswick in Canada, The Maine Regional Medical Program and the Maine Facilities Planning Council.

Basically the data system provides a characterization of the health system in terms of:

1. the communities being served in demographic, socio-economic environmental terms;
2. the manpower, facility and dollar resources of the health delivery system;
3. utilization supply and distribution aspects of the health care system;
4. outcome, as measured by morbidity, mortality and patient satisfaction.

The major products are planning documents and status reports covering the above mentioned areas. Examples are available from the Program office on request.

Establishing the data base has required a major effort which cannot be systematically reported at this time. However, I would like to elaborate on five important features of our approach: (1) choice of the New England town as the geographic base; (2) strategy governing collection of data; (3) resume of the contents of our data file; (4) approach to data processing; (5) approach to data analysis.

The geographic region covered by the data base includes the entire service area of the Northern New

England Regional Medical Program. However, in designing our approach, we wished to use the smallest feasible geographic unit that was available. The New England town turned out to be nearly ideal - for the following reasons: (1) it appears in the census; (2) most unit records in the region (for example hospitals and vital records) contain the individual's town of residence - thus utilization rates can be calculated on a town basis; (3) there are a total of 356 distinct towns in the region - 251 towns and gores in the State of Vermont - 51 in the three counties of upper New York - and 54 in contiguous portions of New Hampshire. Populations in each town vary between 35,000 and 10, with a median value of about 1500. Thus, using the town as a population base allows for a large number of discrete geographic units in the system. This in turn provides great analytic flexibility.

Strategy governing collection of data: all effort has been made to avoid duplication of existing data. Whenever possible, we have used existing sources of data, either published or existing in unit record files collected by cooperating agencies:

Existing data includes those collected, processed, and published by local, state, federal and national agencies: for example, reports of the Bureau of Census, National Center for Health Statistics, American Hospital Association, Blue Cross/Blue Shield, State Health Department and The State Planning Agency.

Existing unit record files include those collected by operating agencies and made available to the Program by special cooperative arrangements: by way of example, a three year file of 200,000 patient discharge abstracts obtained from the hospitals participating in Professional Activity Survey (PAS), the decade files of the Vermont-New Hampshire Vital Records and the individual tax returns from the State of Vermont Tax Department for 1967.

Special collection protocols have been established for "missing" data. This includes surveys, conducted by the staff, of hospitals, nursing homes and home health agency records. It also includes a household survey capability.\*

The avoidance of administrative inconvenience to institutions in providing data is fundamental to success of an ad hoc data system. When data collection has required staff time - such as reviewing unit hospital

records - we have used part-time Regional Medical Program personnel under close core staff supervision. Under these circumstances cooperation has been near universal.

While much of our data base spans more than one year and is updated periodically, the costs involved in fielding special utilization surveys led to a decision to restrict (at least initially) the complete utilization file to the calendar year 1969. Informational items corrected through special protocols have been kept to a minimum. These include patient record number, age, sex, diagnoses, procedures performed, length of stay, date of admission, type of insurance, referring and attending physicians.

*Resume of Content of Data Files:* Currently, our data files contain the following information:

1. Utilization review: hospitals, nursing homes and home health agencies.

A complete review - based on unit records - of all area hospitals for the year 1969. 68,000 records were taken from PAS and 29,000 collected by staff review of the hospital records. In addition, referral hospitals in Harver, Albany and Montreal have been reviewed.

A complete 1969 review of all area nursing homes (85 homes, and 4,000 records).

A complete review of area home health agencies (4 agencies, and 8,000 records).

2. Vital records:

Through excellent cooperation with the Vermont N.H. and N.Y. health departments, decade files of birth and death records have been established. Mortality data is a particularly useful source for defining major health problem areas for measuring outcome.

3. Manpower file:

Hospitals staff listings have been obtained from all institutions of the region. Health Department and A.M.A. registries are being utilized to classify physicians in the region by locality of practice, specialty training, age, board certification etc., both on a current and an historical basis.

4. Facilities:

In cooperation with the Tri-State Regional Medical Program, special inventories of hospital facilities throughout the region have been completed with the following areas being stressed; coronary and intensive care, emergency care, stroke care, radiotherapy and chronic pulmonary care.

In addition, published data encompassing facility staffing, size and location as well as cost data have been compiled from a variety of secondary sources for hos-

---

\*While an integral part of the "data base", this paper does not discuss the NNE/RMP social survey capability.



pitals, home health agencies and nursing homes. Sources include Blue Cross, American Hospital Association and state agencies.

5. Socio-economic and environmental:

Arrangements have been made to secure 1970 census tapes containing available processed tables. This will establish age-specific population rates on a town basis. Of particular importance is intercensal estimates on a town and other small area basis. Work has been completed in conjunction with the State Health Planning Agency personnel to construct inter-censal population age-sex structure for towns and counties.

Indicators of economic status are being constructed through the use of income data. In conjunction with the State Tax Department individual income tax returns have been analyzed by town, occupation and industry to provide an economic profile of the State and its subdivisions.

6. Published data:

For example, complete set of reports from the National Center for Health Statistics.

*Approach to Data Processing.* Routine reports prepared by agencies and organizations in the health field rarely provide direct answers to specific questions as they arise in planning, management and evaluation activities within a local or regional context. Processed data, organized and tabulated according to external dictates, is often irrelevant to immediate concerns. The limited utility of reports furnished hospitals by the Commission of Professional and Hospital Activities (PAS) and of publications of state and federal health departments reflect the series of compromises that must be made in developing multi-purpose reports. From several standpoints, the most effective method of information storage is raw data on individual cases. This is particularly true when efficient storage and retrieval methods are available.

Accordingly, the RMP has devoted a significant effort to the development of individual case files. Because accessible data derives from diverse sources, a number of compatibility problems have been encountered. These range from differences in coding of such items as sex and

age to problems in format design and basic character configurations. As an example, sources of data include magnetic tape obtained from PAS (Minneapolis Honeywell), Vermont State Government (General Electric) and New York Health Department (Burroughs). To solve these problems generalized recoding and formatting programs have been developed.

*Approaches to Data Analysis.*—The usefulness of the data base relates to 1) the completeness of each file (for example, one year of hospital experience for the total population) and, 2) the inclusiveness of the system in terms of the large numbers of separate data files containing relevant health data. This enables (for example) correlations between demographic and environmental factors in health status. Much of the analysis undertaken by the RMP has been computer based and allows for the study of complex relationships between “input” and “output” variables. Examples of correlation analyses that are possible include relationships between per capita income, admission rates, death rates, infant mortality rates, expenditures for medical care, procedure rates, etc.

While a number of general statistical programs have been adapted, we have also developed a series of new innovative types of health system analysis. Of particular note is a program designed to characterize total utilization and allocation of medical resources relative to the patient service areas of particular institutions. This includes resource allocation rates in terms of admissions, patient days, beds, dollars or skilled manpower. Because virtually all utilization experience for each town in the region is known, these rates describe the total experience of the population. Thus, for the first time, an accurate estimate, based on a small population, is possible: this includes total cost for institutional care, procedure rates, bed utilization and beds available rates, etc.

During the next year, the NNE/RMP will complete a number of reports for areawide and state planning purposes. I hope that the next time I report to you on the data base we will have much more to say about the effect the data has had on the planning process.

# WORKSHOP ON MEASURING CHANGES IN BEHAVIOR

## Participants

John S. Lloyd, Ph.D. - Moderator  
Assistant Coordinator, Evaluation  
California Regional Medical Program -  
Area V

William R. Crawford, Ed.D.  
Associate, Evaluation Studies Section  
Center for Study of Medical Education  
University of Illinois

Mitchell Schorow  
Assistant Coordinator for Educational  
Planning and Evaluation  
Intermountain Regional Medical Program

Barbara J. Andrew, Ph.D.  
Assistant Professor Medical Education Research  
Division of Research in Medical Education  
University of Southern California

---

## Measuring Changes in Knowledge

WILLIAM R. CRAWFORD

Sometimes the measurement of knowledge seems to be a straightforward procedure. Perhaps that is true when one is interested in measurement of simple recall of basic information which has been memorized. However, simple recall of basic information is usually not sufficient for measuring the achievement of educational objectives in areas as conceptually complex as medicine and the allied health professions. In most cases we are interested in assessing changes which are related to the ability to apply principles, solve problems, and interpret data, to name only a few. Clearly, these complex intellectual functions cannot be assessed with instruments designed to provide an estimate of the number and kind of memorized facts which can be recalled.

How, then, can we approach the greater problem of measuring the ability to engage in more complex intellectual functions? The obvious first step is to define what those functions are, why they are important, and how they relate to specific tasks which must be performed on the job. Defining these functions is a major operation, and an essential step before specific measurement instruments can be developed. The second step is to take these definitions and translate them into instruments which can validly and reliably measure the functions, and which will produce meaningful data. Concurrent with the development of the instruments one must develop a procedure for scoring and a plan for reporting and interpreting the scores.

Following is a brief outline of the topics covered in this session of the workshop, each of which was considered in more depth in the working session.

## Multiple Choice Items

- A. Advantages
  - 1. Some task clearly defined for each examinee
  - 2. Large sample of items permissible
  - 3. Scoring keys are standardized
  - 4. Easy to score
- B. Disadvantages
  - 1. Requires recognition of correct response, not production of it
  - 2. Permits guessing
  - 3. Difficult to construct
  - 4. Task is completely structured

## Measuring Changes in Clinical Performance

BARBARA J. ANDREW, Ph.D.

The health professional's ability to solve clinical problems has long been regarded as one of the most important dimensions of quality health care delivery. Yet because of its complexity and the challenges which it presents for quantitative measurement, clinical performance has not been as widely used as a criteria for evaluation as its importance would suggest.

Clinical performance is essentially a *problem solving process* which involves:

- 1. knowing what data are relevant;
- 2. gathering the data;
- 3. analyzing the data and evaluating their relative importance and significance;
- 4. synthesizing the data into conclusions;
- 5. knowing about available health care strategies;
- 6. selecting and applying the most appropriate strategies;
- 7. evaluating the effectiveness of the strategies;

8. making whatever changes in health care strategies which are needed.

Specific clinical problem solving activities can be classified as primarily diagnostic or therapeutic in nature. That is, while diagnosis and therapy are interdependent components of clinical problem solving, some health professionals have primary responsibility for diagnosis, while others are concerned with suggesting or administering therapeutic procedures. Still other health professionals, such as the physician, are responsible for diagnosis as well as therapy.

The measurement of clinical performance can focus either upon the entire problem solving process employed by a specific health professional or solely upon the frequency with which certain behaviors within the process are observed. In measuring changes in clinical performance to determine the effectiveness of particular experimental treatments, the decision to observe the entire problem solving process or only some specific behaviors within the process will be a function of the purposes of the study and the hypotheses which have been stated.

The validity of clinical performance measurement will, of course, rest upon the quality of the instruments which are devised to record the problem solving behavior. The following *procedures* should be followed in the *development* of such *instruments*:

1. the clinical skills to be measured are identified;
2. criteria for evaluating these skills are developed;
3. the criteria are stated in terms of specific clinical behaviors;
4. a method of scoring is developed which is logically appropriate to the skills being measured;
  - a. the assignment of differential scores to various levels of performance should be clearly defined and require as little subjective judgment of the rater as possible;
  - b. scoring intervals need to be sufficiently sensitive to permit the discrimination of different levels of clinical performance;
5. prior to establishing the validity and reliability of the instrument, extensive pre-testing is undertaken to determine its usability and capacity to measure all relevant aspects of the specific clinical skills;
6. if the instrument is to be used by a rater who observes an actual or simulated clinical setting, it should not attempt to measure more than can reasonably be observed and recorded by a single individual. [If two or more simultaneous dimensions of clinical performance are to be observed, additional instruments can be developed

and used by different raters (e.g., non-verbal as well as verbal interaction during history taking)];

7. finally, the validity and reliability of the instrument are estimated. (In instances where the instrument has been designed for use by raters to observe clinical performance, sufficient training to improve inter-rater reliability should be undertaken).

The *selection of appropriate validity and reliability estimates* depends upon the nature of the measuring instrument itself and upon the purposes for which testing data are gathered (3).

In estimating the reliability of observation devices one needs to determine the correlation among the evaluations of several raters of the same clinical performance. This procedure necessitates the refinement and careful definition of the skills to be measured and categories for recording performance, as well as the training of observers so that acceptable inter-rater reliability can be achieved.

When the measuring device consists of a paper and pencil test of clinical performance or the simulation of a clinical situation, comparability of forms and comparisons over time offer the best estimates of reliability. Estimates of the test-retest reliability of simulated clinical performance test are complicated, however, by the fact that these simulation tests permit the examinee to receive feedback from his selections and, hence, to some extent constitute a learning situation. Even if the time interval between test administrations is lengthened to enhance forgetting, one cannot control intervening variables which might improve the subjects' problem solving skills.

Since in measuring changes in clinical performance one is primarily interested in determining the degree to which the health professional possesses certain clinical problem solving skills, the use of criterion-related validity is somewhat less pertinent than is construct validity.

The establishment of construct validity can be undertaken by hypothesizing outcomes of performance for various groups on the problem solving test, and subsequently administering the test to determine whether the hypothesized outcomes occur. In instances where other tests of the same clinical performance exist, the correlations between the test being developed and these other measures should be estimated.

Regardless of the kinds of validity and reliability which are considered appropriate for a specific measure of clinical performance, the subjects on whom validity and reliability studies are conducted should closely resemble the population for whom the test has been

designed, in terms of their composition and relevant characteristics.

Two general approaches to the measurement of clinical performance may be taken: 1) the direct observation and measurement of actual or simulated clinical situations; 2) the indirect measurement of actual or simulated clinical situations.

The advantages of evaluating actual clinical situations result primarily from the difficulties in simulating some of the complexities and spontaneous aspects of actual problem solving settings. For example, the clinical performance of some medical technologists requires the use of actual specimens, thus rendering observations under simulated conditions considerably distorted and of limited value. This same difficulty is posed by the use of simulated patients from whom the physician could take a history and perform, in some instances, a physical examination, but on whom it would be impossible to perform laboratory procedures not only because the obtained data would be inconsistent, but because of the understandable unwillingness of subjects to undergo such experiences. Thus, the use of simulated clinical settings restricts to some extent the range of skills which can be measured.

However, since the measurement of clinical performance is generally for the purpose of assessing the effects of an independent variable upon clinical problem solving behavior, or to make comparisons among individuals regarding their clinical competence, the use of actual clinical settings may pose difficulties in obtaining uniform testing conditions and in securing adequate numbers of subjects. Thus, if one wanted to measure the effects of an instructional film on the management of hypertensive patients in a hospital clinic one would need a sufficiently large patient population randomly assigned to clinic physicians in order to permit valid conclusions to be drawn.

The *decision to employ either direct or indirect measurement* of clinical performance in actual or simulated situations will usually be based upon a number of considerations such as: 1) the kind of clinical skills to be measured; 2) the availability of subjects and observers; 3) the number and extensiveness of the clinical skills to be measured; and 4) the amount of time required for observation.

Peterson's study of North Carolina general practitioners (28) represents perhaps the most comprehensive attempt to measure physicians' clinical problem solving skills by direct observation of an actual situation. The observation forms developed by Peterson and his colleagues measure the physician's skills in history taking,

physical examination, laboratory procedures, and therapy. Particularly relevant to measurement of this kind is that the evaluation of clinical performance be a function of specific disease entities and the diagnostic and therapeutic procedures which are indicated for each. Thus, the validity of the conclusions drawn from this kind of measurement depends not only upon the appropriateness and sensitivity of the observation forms, but upon the clinical competence of the observer who in an actual situation must not only record physician behavior, but must develop his own diagnosis in order to evaluate the appropriateness of that behavior to the particular clinical problem.

While the direct measurement of a physician's complete management of a clinical case results in a more comprehensive evaluation of physician performance, some studies have focused upon specific components of patient management (2, 14). Foster and Lass (14) will soon be reporting procedures for the measurement and evaluation of patient interviewing. The measurement of patient interviewing skills can emphasize content (how much and what kinds of information are elicited) and/or process (the techniques used to elicit information). In order to measure the process of patient interviewing one needs to: 1) identify those dimensions which will account for all possible aspects of interaction; 2) determine whether these dimensions are essentially verbal or non-verbal; 3) develop observation forms which provide sufficient scope and flexibility to permit the recording of relevant aspects of communication and interaction.

Barrows and Abrahamson (4) have reported the use of trained actors to simulate patients with neurological disorders in order to measure history taking and physical examination skills. Although the use of the programmed patient imposes limitations upon the kinds of disorders which can be simulated, the pre-determined nature of the medical setting permits more accurate evaluation of the extent to which pertinent data have been uncovered by the examinee.

In a somewhat different approach to measuring competence in data gathering and analysis, Cline (7), Langsley (19), and Levit (22) have reported the use of motion pictures to assess observation and interpretive skills. The films which consist of a history and physical examination show a wide range of signs and symptoms which are both pertinent and non-pertinent to the formulation of a correct diagnosis. The data is presented with equal emphasis and in such a manner that the examinee must analyze all data, make judgments about

their relative significance, and draw conclusions concerning the nature of the patient's illness.

The medical audit, which in essence is an *a posteriori* evaluation of the clinical management of an actual case, has been the subject of numerous articles (5,6,20,21,26,27,29,30). Such a process requires the careful establishment of criteria by which the medical record is evaluated and the training of medical specialists who will serve as auditors of the medical record. There is the danger, however, that one may be measuring the accuracy and completeness of the medical records themselves, rather than the clinical performance of physicians.

Yet another indirect evaluation of clinical problem solving is the so-called "patient management problem" - a written simulation of a clinical case which measures data gathering and interpretive as well as decision-making skills (10,16,25,31,33). Although its use has been reported primarily with physicians and nurses, its applicability to other health professionals appears feasible. The problem-solving exercise is initiated by a brief description of the patient and consists of "a series of sequential, interdependent decisions representing the various stages in the management of the patient" (25:1) in which the results of each decision are given in the form in which the health professional would receive them in an actual clinical setting. Moreover, the problem not only allows the examinee to make a wide range of decisions from very harmful to very helpful, but forces him to deal with the consequences of his decisions by presenting additional choices through which the examinee can either correct or further compound his mistakes. Allowances are also made, where applicable, for the use of more than one acceptable diagnostic or therapeutic procedure.

The following selected bibliography has been included so that individuals wishing to do so may further explore the literature on clinical performance measurement.

#### Selected Bibliography

1. Abrahamson, Stephen. "Evaluation in Continuing Medical Education." *Journal of the American Medical Association*, 206:625-628, October 14, 1968.
2. Adler, Leta McKinney, and Enelow, Allen J. "An Instrument to Measure Skill in Diagnostic Interviewing: A Teaching and Evaluation Tool." *Journal of Medical Education*, 41:281-288, March, 1966.
3. American Psychological Association, et al. *Standards for Educational and Psychological Tests and Manuals*. Washington, D.C.: American Psychological Association, 1966.
4. Barrows, Harold S., and Abrahamson, Stephen. "The Programmed Patient: A Technique for Appraising Student Performance in Clinical Neurology." *Journal of Medical Education*, 39:802-805, August, 1964.
5. Beaumont, Graham, et al. "Medical Auditing in a Comprehensive Clinic Program." *Journal of Medical Education*, 42:359-367, April, 1967.
6. Butler, John J., and Quinlan, J. William. "Internal Audit of the Department of Medicine of a Community Hospital." *Journal of the American Medical Association*, 167:567-570, May 31, 1958.
7. Cline, Marvin. "A Film Test of Clinical Skills in Medical Students." *Journal of Medical Education*, 36:908-910, August, 1961.
8. Colton, Theodore, and Peterson, Osler L. "An Essay on the Abilities of Medical Students by Oral Examination." *Journal of Medical Education*, 42:1005-1014, November, 1967.
9. Cowles, John T. "A Critical Comments Approach to the Rating of Medical Students' Clinical Performance." *Journal of Medical Education*, 40:188-198, February, 1965.
10. De Tornay, Rheba. "Measuring Problem-Solving Skills: Means of the Simulated Clinical Nursing Problem Test." *Journal of Nursing Education*, 7:3-8, August, 1968.
11. Donabedian, Avedis. "Promoting Quality Through Evaluation of the Process of Patient Care." *Medical Care*, 6:181-202, June, 1968.
12. Evans, Lloyd R., Ingersoll, Ralph W., and Smith, Edwin J. "The Reliability, Validity, and Taxonomic Structure of the Oral Examination." *Journal of Medical Education*, 41:651-657, July, 1966.
13. Foster, Judilynn T., et al. "Analysis of an Oral Examination Used in Specialty Board Certification." *Journal of Medical Education*, 44:951-954, October, 1969.
14. \_\_\_\_\_, and Lass, Sandra L. "The Identification of Interaction Patterns in Student-Patient Communication: Abstract of a Paper to be presented at the 1970 Conference on Medical Education, Association of American Medical Colleges.
15. Hinz, Carl F., Jr., "Direct Observation as a Means of Teaching and Evaluating Clinical Skills." *Journal of Medical Education*, 41:150-160, February, 1966.
16. Hubbard, John P. "Programmed Testing in the Examination of the National Board of Medical Examiners." *Education Testing Service, Proceedings of the 1963 Invitational Conference on Testing Problems*.
17. \_\_\_\_\_, et al. "An Objective Evaluation of Clinical Competence: New Techniques Used by the National Board of Medical Examiners." *New England Journal of Medicine*, 272:1321-1328, June 24, 1965.
18. Kilpatrick, G.S. "Observer Error in Medicine," *Journal of Medical Education*, 38:38-43, January, 1963.
19. Langsley, Donald G. "Filmed Interviews for Testing Clinical Skills." *Journal of Medical Education*, 45:52-58, January 1970.
20. Lembcke, Paul A. "Evolution of the Medical Audit." *Journal of the American Medical Association*, 199:111-118, February 20, 1967.
21. \_\_\_\_\_. "Medical Auditing by Scientific Methods." *Journal of the American Medical Association*; 162:646-651, October 13, 1956.
22. Levit, Edythe, J. "The Use of Motion Pictures in Testing the Clinical Competence of Physicians." *Annals New York Academy of Sciences*, 142:449-454, March 31, 1967.
23. McGuire, Christine H. "Medical Education, Part I: A Scientific Approach to Problems of Professional Assessment.

- Canadian Medical Association Journal*, 100:593-598, April 5, 1969.
24. \_\_\_\_\_. "The Oral Examination as a Measure of Professional Competence." *Journal of Medical Education*, 41:267-274, March, 1966.
  25. \_\_\_\_\_. "Simulation Technique in the Measurement of Problem-Solving Skills." *Journal of Educational Measurement*, 4:1-10, Spring, 1967.
  26. Morehead, Mildred A. "The Medical Audit as an Operational Tool." *American Journal of Public Health*, 57:1643-1656, September, 1967.
  27. Payne, Beverly C. "Continued Evaluation of a System of Medical Care Appraisal." *Journal of the American Medical Association*, 201:126-130, August 14, 1967.
  28. Peterson, Osler, Andrews, Leon P., Spain, L.P., and Greenberg, R.S. "An Analytical Study of North Carolina General Practice, 1953-1954." Evanston, Illinois: Association of American Medical Colleges, 1956.
  29. Phancuf, Maria C. "Analysis of a Nursing Audit." *Nursing Outlook*, 16:57-60, January, 1968.
  30. \_\_\_\_\_. "The Nursing Audit for Evaluation of Patient Care." *Nursing Outlook*, 14:51-54, June, 1966.
  31. Rimoldi, H.J.A. "The Test of Diagnostic Skills." *Journal of Medical Education*, 36:73-79, January, 1961.
  32. Vigliano, Aldo, and Gaitonde, Mangesh. "Evaluation of Student Performance in a Clinical Psychiatry Clerkship." *Journal of Medical Education*, 40:205-213, February, 1965.
  33. Williamson, John W. "Assessing Clinical Judgment." *Journal of Medical Education*, 40:180-187, February, 1965.
  34. Wilson, G.M. et.al. "Examination of Clinical Examiners." *Lancet*, 1:37-40, January 4, 1969.

## WORKSHOP ON THE EVALUATION OF CHANGING HEALTH STATUS

### Participants

Robert R. Carpenter, M.D. - Moderator  
Director, Western Pennsylvania Regional  
Medical Program

Sam Shapiro  
Director of Research and Statistics  
Health Insurance Plan of Greater New York

Maureen M. Henderson, M.D.  
Professor, Preventive Medicine  
Department of Preventive Medicine and  
Rehabilitation  
University of Maryland

Howard R. Kelman, Ph.D.  
Department of Preventive Medicine  
and Public Health  
New York Medical College

Charles E. Lewis, M.D.  
Professor and Head  
Health Administration Division  
School of Public Health  
University of California

---

### Transcript of Workshop—Remarks by Moderator

ROBERT R. CARPENTER, M.D.

DR. CARPENTER: Thanks to Mitch Schorow, I found an interesting book, published in Boston in 1917 by E. A. Codman. It's called "A STUDY IN HOSPITAL EFFICIENCY, A DEMONSTRATION BY THE CASE REPORT METHOD OF THE FIRST FIVE YEARS IN A PRIVATE HOSPITAL."

It says by way of foreword that this hospital has for sale a product of a standard which is to be described on pages 12 through 63. It aims to be a \$100 hospital with a \$100 surgeon.

The volume is dedicated to Richard Cabot because Dr. Codman respected his motives and admired his courage and energy though he heartily disapproved of some of his opinions and methods. "He seems to want to reform the profession from the bottom whereas I think the blame belongs at the top," says Dr. Codman.

The case report is subtitled "A Practical Illustration of the Fact that It's Possible to Use the End Result System in a Hospital."

And the first page I think suggests how little progress we have made since 1917: "The trustees of our charitable hospital do not consider it their duty to see that good results are obtained in the treatment of patients. They see to it that their financial accounts are audited and they take no inventory of the product for which their money is expended."

"It is against the individual interests of the medical and surgical staffs of hospitals to follow up, compare, analyze and standardize all their results because (1) it is

seldom that any individual's results are sufficiently better than those of his colleagues so that he would desire such comparison. Perhaps the results as a whole would not be good enough to impress the public very favorably. (2) An effort to thus analyze is difficult, time-consuming, troublesome and would lead, by pointing out lines for improvement, to such onerous committee work by members of the staff."

"Neither trustees of the hospital nor the public are as yet willing to pay for this effort."

"Although the staff would admit that such follow-up analysis was a good thing for all, yet each practical man — and the practical men always hold the power — would wait for somebody else to do the work."

And he goes on to point out that the superintendent would be the last one to undertake this task because he surely would lose his job.

I enjoyed that 1917 description of what we are trying to do in Western Pennsylvania in 1970 and I don't know that we have come terribly far in our ability to measure health status and particularly any change in health status attributable to any of our efforts to improve what we are doing.

Yesterday we heard the public and the Bureau of Budget — good morning, Dr. Fox — ask for health status outcome measurements. I think RMPS asked for end results but not really health status end results. They were asking for lower cost and better distribution of care which is significantly different than outcome analysis going to end results.

I was interested in this workshop above all of the others. Since I am interested in the Regional Program as

a way to improve health, I want to know how to measure health and its improvement.

I think we have an unusually talented group with us this morning to help us do this. The speakers who will enter into discussions with you off and on during the morning have spent a good many years measuring health status; I look forward to learning a great deal from them.

They are Dr. Henderson, Mr. Shapiro, Dr. Kelman and Dr. Lewis.

I want to show you just four numbers as an example of the problem and promise of end result evaluation. We looked at the hospitals that serve a community of 200,000 and at the mortality from stroke in those hospitals. We were surprised to find that patients with heavy paresis when they were cared for by generalists died more frequently whether they were male or female than did patients with the same reported neurologic signs if they were cared for by internists.

I hoped from this that we could attract the medical staffs' interest to more careful care of stroke, attract interest in helping to understand these results.

We identified some cases that the medical staffs were particularly interested in reviewing: The patient who died with a diagnosis of cerebral vascular disease without any neurologic signs, for instance.

I hope as the morning goes on that we can learn the value of such measures of outcome as others who have made them more frequently have seen this value. I hope that we can find out when it's worthwhile to make such measures and how to make them. I hope we can learn how to interpret them once we have them. I hope that we can learn, particularly from Dr. Kelman, the key data bits that help us to measure and talk about outcome. And, finally, I hope we can learn how outcome and process analysis interrelate.

Mr. Sam Shapiro will begin the discussion. He will describe some of his studies and discuss why and when it is worth measuring health status.

### The Value of Health Status Measures

SAM SHAPIRO

I'm not sure that I'm going to be dealing with the questions precisely the way you have outlined them. What I thought might be useful is for me to give you some general considerations that underlie the concern with measurement of health status changes and then use a few samples principally from my own experience to illustrate what's really at stake when you get involved with health status measures.

The acceptance of the desirability to determine the effect of health programs on the well being of a population is quite general not only among researchers, but also among planners, administrators, and among those responsible for allocation of resources.

This acceptance moves from a state of passivity to worried preoccupation when change is contemplated or alternatives weighed in circumstances ranging from a highly specific component of health care to the broad design of organization and financing of health services.

It usually slips back to an uneasy but quiescent state when the complexities of end result measurement, costs, and time requirements become apparent.

This is not in the nature of a sharp criticism of the past. The difficulties of assessing the impact of particular actions on health status were and still are great. Further, the introduction of changes affecting the availability, delivery and economics of health care often could not and will not in the future wait for hard evidence from studies of impact.

Similarly for the introduction of some programs aimed at modification of primary and secondary prevention of specific diseases.

However, many of the problems and issues we face are stubborn, and courses of action are not at all certain. Because selection of an available alternative often involves commitment of scarce manpower, equipment and financial resources for which there is sharp competition, implementation faces serious obstacles.

As we all know, these are the considerations that force many of us to think in terms of demonstration projects or R & D projects in which operational effectiveness related to costs and manpower is a central concern.

Now, this is fine, but often the question that will remain even after a project has been well executed is whether any health benefits have resulted.

Bypassing the issue can compromise the potential for moving from demonstration to general acceptance. In fact, where the effort required for the extension is great, absence of evidence of impact on health status may well prevent such extension.

Conversely, availability of evidence of a program's health benefit can stimulate widespread consideration of early implementation.

I want to emphasize that many programs cannot be nor need be tested for health benefits although there are programs under active consideration today that will be plagued by doubts and challenges until the issue of health benefits is dealt with effectively. Just to mention a few: early disease detection through automated multi-



breast and mammography. The measure is change in mortality from breast cancer.

This is the situation. It's generally acknowledged that screening will lead to earlier diagnosis of breast cancer, but there has been no evidence that this results in lower mortality. Costs for including breast examination with mammography, in particular, are high. And, in fact, in automated multiphasic health testing programs where this procedure is used, mammography is the most costly single test.

In short, a national effort to screen women for breast cancer would require massive expenditures and diversion of equipment and manpower from other health care activities.

Clearly, to acquire a high priority, breast cancer screening should justify its value in the most rigorous manner possible. And as many of you know, a randomized clinical trial directed to this issue has been underway since 1963 in HIP under a contract with the National Cancer Institute.

The main objective is to establish whether breast cancer screening using mammography and clinical examinations results in a reduction in breast cancer mortality. Other objectives relate to the epidemiology of breast cancer and the search for high-risk factors that might be useful in future screening programs.

I don't want to go into the details of methodology. These have appeared elsewhere. But a few key points are important for me to touch on in this discussion.

Thirty-one thousand women aged 40 to 64 enrolled in HIP have been assigned randomly to a study group and a similar random sample to a control group. Only study group women have been invited for screening examinations. About 65 percent appeared for the initial screenings.

Three additional screening examinations at annual intervals were scheduled, and large proportions of the women with an initial examination have returned for these.

Control group women continue to receive their regular medical care.

Screening examinations have been performed at 23 of the HIP medical group centers. The clinician and radiologist record their examination findings and recommendations independently. Later their findings are reviewed jointly by a physician for final recommendations. Intensive follow-up to identify breast cancers diagnosed and mortality is carried out with equal rigor for women screened, women who refused screening, and control group women.

All screening examinations have been completed. And at every stage of the investigation when findings were reviewed it was clear that mammography and clinical examinations contributed independently to the detection of breast cancer. If mammography had been excluded, 31 percent of the cancers would have been missed during screening. If the clinical examination had been omitted, 44 percent would have been missed.

Further, screening did lead to detection of large proportions of breast cancer with no evidence of axillary nodal involvement — 70 percent — than among the control group — 45 percent.

Preliminary results on mortality are now beginning to be collected and will shortly appear in an article in JAMA. The findings are highly encouraging. There are 52 deaths due to breast cancer in the control group as compared with 31 breast cancer deaths in the total study group in the period available for follow-up.

The case fatality rates for cases with histologically confirmed breast cancers reinforce the impression that, in the short run at least, screening leads to lowered mortality.

These observations are preliminary, and more time is needed to establish whether the effect of the screening program is short-term or long-term.

However, the findings do provide grounds for cautious optimism and it would appear prudent to accelerate efforts to develop and test methods capable of dealing with the broad demand for periodic breast examinations that might emerge within a few years.

What I'm describing is a progression from very intensive study involving huge resources, a long period of time, and dedication of large numbers of personnel to achieve a result which if sustained can significantly affect the approach that medical care might be taking to the whole issue of screening for breast cancer.

Those of you who have been close to this field over the years know how much disappointment there has been in dealing with the problem of breast cancer, and how widespread is the pessimism about the effectiveness of breast cancer screening.

There is a great deal at stake in this study, and as I see it, these preliminary results are placing high on the agenda a new set of concerns, mainly related to the question, "what kind of screening program would be required to reach effectively large numbers of women if the present findings persist?"

I want to turn now to a much broader type of effort in the field of preventive care, that of automated multiphasic health testing. There are many justifications advanced for introducing AMHT, and I don't want to

make out a case for or against such programs. However, everyone will agree that AMHT is a costly addition to the spectrum of health services and most will agree that it is important to seek out opportunities to assess AMHT's effect on health status and on behavior that might be expected to have a desirable effect on health and well being.

Two projects are now directed to this issue. The one of longer duration is being conducted by the Permanente Medical Group in California. One phase of that program is very well known, probably much better known than the second phase which deals with the end result issue. This is to demonstrate how automation and computers can be applied to improve — and I'm now quoting — “speed, efficiency, and quality control in multiphasic screening techniques so that not only more tests, but more accurate and quantitative measurements can be performed, and at a lower cost.” All very important operational objectives.

The other phase of the program includes a set of end result criteria in the evaluation. Two randomly selected samples of the plan's member have been designated study and control groups. Efforts are made to have the study group appear for the examination. The control group is not approached, but those who request an examination are accommodated.

Morbidity, disability, and medical utilization patterns are to be determined over a long period of follow-up through periodic questionnaires and medical records.

This is an ambitious undertaking. But it has the potential of providing decisive information on the value of periodic health examinations generally and of selected components of it particularly.

Anyone who questions the time requirement for reaching an answer really has to look very hard at other issues that have come up in the past which have been plagued by doubts and questions long after the point in time when it would have been possible to initiate an end result investigation.

One of the outstanding examples is the Pap smear. It is no longer possible to carry out a control study in this country on Pap smear as an effective measure for reducing mortality from cervical cancer. There are very few people on the firing line who really raise any questions about Pap smear. But if you look at the scientific literature, there are some very serious questions being raised about the Pap smear.

The second end result study in multiphasic health testing recently started at HIP. This project is utilizing repetitive health testing to define the health status, practices and attitudes of a defined poverty population

covering a broad age range—12 years and older—from an absolute standpoint and relative to a nonpoverty group that will also have AMHT.

Action to modify adverse aspects of these health components among the poor is to be instituted, and evaluation is in terms of change as compared with what occurs in the nonpoverty group.

An underlying question is whether through the AMHT program, and activities generated by it, the anticipated gaps between the two groups can be narrowed.

A broad spectrum of measures are being developed to measure health impact. These include changes in impairment of function, immunization status of children, and complications of disease.

The last project I want to describe is in the proposal stage and is now being reviewed for possible funding. It concerns sudden death from coronary heart disease.

There is general agreement that until effective primary prevention methods can be identified and implemented, significant progress in reducing the incidence of this cause of death will depend on changes in community practice which bring advances in coronary care to patients who under present circumstances do not survive to reach a hospital.

It is estimated that about 60 percent of deaths due to acute myocardial infarction occur outside the hospital, and a great effort is being made to cope with the problem of rapid response to requests for medical care when a heart attack is suspected. Also, increasing attention is being given to finding out how patients and their families behave when faced with prodromal symptoms.

The proposed project is designed to incorporate these approaches in a comprehensive action program. It represents a combined effort of HIP and two of its Queens medical groups with a population of about 50,000, aged 35 to 74, and HIP's LaGuardia Hospital which serves both medical groups.

The goal is to effect more rapid requests for medical care after the onset of a heart attack or suspected heart attack and to institute a system capable at all times of a rapid and appropriate response which fully utilizes current medical knowledge.

The end result sought is a reduction in the present high rate of sudden death from coronary heart disease.

Basic changes to be made in the health services system consist of the following main elements:

- Patient education. Varied educational approaches will be made to the entire adult population of the two participating medical groups and their

physicians with the aim of reducing delays generated by patients or their families in seeking medical care for possible acute coronary episodes. A special target will be individuals at relatively high risk for sudden death (those with prior CHD, hypertension, hypercholesterolemia, etc., as identified through the HIP centralized medical record system).

- Centralized telephone screening at LaGuardia Hospital by physicians of calls from all possible coronary suspects in the population will take place 24 hours a day, 7 days a week. The aim is to reduce communication delays in bringing the patient's symptoms to trained medical attention.
- Operation of a special pre-coronary care area (PCA) at LaGuardia Hospital for observation of patients in defined categories, one of which consists of persons who do not meet usual current criteria for hospitalization, but who may be in an early stage of an acute MI not yet recognizable. The other consists of patients who might be experiencing an ischemic episode not destined to lead to MI but capable of inducing a fatal arrhythmia.

For purposes of this meeting I think what is of particular importance is that two types of evaluation have been planned for. The first is directed at those aspects of the project that bear on generalizing experience for possible use by other organized providers of medical care in Queens and the New York area. Information will become available regarding the operational effectiveness of the educational program, communication procedures for rapid response to patient's call, training of paramedical personnel, and the operation of the pre-coronary care area.

This information would be related to manpower requirements and costs.

By itself, this would represent an important advance in knowledge concerning the modification of health care systems to reach a patient early when a heart attack occurs. However, we would still be left with the unanswered question as to whether the effort involved does have payoff in reducing mortality.

A second type of evaluation has been included which is aimed at answering this question. The approach is to compare the rate of sudden coronary heart disease deaths in the demonstration groups with the rate in other HIP medical groups, and also provision has been made to compare the mortality situation in the demonstration groups before and after start of the program.

Each of the studies I have described contains an end result criterion. In the breast cancer study we probably have the hardest type of evidence. It's a single measure mortality. A randomized clinical trial approach has been used, and it takes a very unusual set of circumstances to make this type of approach a practical one.

The other investigations shade off in hardness, but an effort is made to maximize the opportunity within the medical care setting where the programs are to be carried out, to reach sufficiently hard conclusions about the effectiveness of the programs, from the standpoint of health status measures, to serve as a basis for future action.

One question that often comes up is whether all or most demonstration programs should attempt to incorporate an end result criterion? I don't believe so. Costs are high. Technical requirements are great. Are frequently the kinds of issues that are being faced are not susceptible to the inclusion of an end result measure.

But in the field of medical care, with all the changes that people hope will take place over the next decade maybe they're being optimistic about the next decade but let's say the next generation — there are very large issues with very large stakes associated with them. My point is that it is essential to seek out those few situations where such issues can be investigated effectively utilizing end result measures and thereby provide the basis for making judgments that have regional or national implications.

DR. CARPENTER: Thank you very much. I gather that hard work makes a man cautious. It seems that you are an enthusiast for health status measurement under proper conditions.

I heard you saying that the detailed effort required to carry out a study significant enough to be generalized from one hospital to another is very great indeed. You alluded to an indirect method of end result analysis which sounded as though it might be more often applicable to the problems faced by Regional Medical Program evaluators. In just a moment we will have a chance to discuss these and some other issues with you.

Before we go on, it's worth noting that the man who is evaluating the evaluation conference is with us—Glen Hastings from Nassau-Suffolk. Welcome, Glen; we'll be very careful what we say from here on.

Before we discuss Mr. Shapiro's paper, Dr. Hendershot will speak to closely related issues. Maureen will discuss some of the problems of end-results analysis, particularly as she experienced these within the framework of the Maryland Regional Medical Program where she is Associate Director of the Epidemiology and Statistics

Center. She will also discuss the relationship between evaluation and medical care research.

#### MAUREEN M. HENDERSON, M.D.

I propose to review a very different level of research from Sam Shapiro. One major value of this workshop is the way it illustrates the need for many different disciplines and approaches in the evaluation of health services. Mine is a very limited approach within the total context of health services research and evaluation. As an epidemiologist, I am most interested and only competent to deal with biological measurements. The end results I have been looking at in relation to the Maryland Program have therefore been measurements of morbidity.

I believe it is important to talk about ways of making these particular measurements because non-epidemiologists are not always aware of the series of confounding issues and problems related to their observations.

I trust those present who are sophisticated in epidemiologic techniques will bear with the fundamental levels I am going to discuss.

Let us first consider biologic outcome measurements in relation to overall evaluation of regional services. The two types of measurements consistently used are those of death and morbidity. We have made very little use of death records.

In looking at the picture of our total region, we have been using case fatality rates. The latter are of limited use now for two reasons: (1) there have been great changes in denominators—the census population from 1960-1970 and (2) death rates have been at a standstill for the past few years.

In terms of disease or morbidity we are looking at the prevalence and severity—that is, the distribution of the frequency and severity of disease as we see it in the region. We are also looking at aggregations of disease—that is, multiplicity of disease problems in patients who are found at points on each disease spectrum.

One good example is the presence of cardiovascular diseases and diabetes mellitus in stroke patients with mild or serious neurological deficits.

The greatest amount of the data we are currently studying comes from hospital in-patient records. I think it is appropriate to speak mostly about hospital in-patient records today because I am sure that most regional programs use these as their major source of morbidity data.

Let me briefly describe the collection of the information I am going to show. We took a random sample of admissions to every short-term general hospital in the region during a 12-month period just before the regional medical program began. The data, therefore, describe patients and procedures in every "acute" hospital whether or not it prepares its own statistics or has easily accessible records. In most other morbidity surveys, information is collected only from hospitals with viable (for research) record systems.

The Maryland Region includes all of Maryland except Montgomery County and includes York County, Pennsylvania.

The specific medical records reviewed in our sample were identified by our own staff and abstracted by trained medical abstractors under constant quality control and surveillance. Standardized abstracting forms and procedures were used.

The measurement data collected were specifically selected to:

1. get estimates of need;
2. look at the secular effects of the total program and of individual programs;
3. insure proper comparisons in assessing needs or effects.

The last purpose is one Sam Shapiro spoke about very briefly and one on which I would like to enlarge. Whenever you examine an effect or an end result in different time periods or between different groups of people or different geographic areas, you must be sure that you are comparing like with like. The original numbers that Bob Carpenter presented draw attention to this point, and because he mentioned that he was going to show those particular stroke data, I brought some of our own stroke data to illustrate and amplify this point.

This slide describes short-term general hospital discharges in the region of Maryland before the Regional Medical Program began. It shows annual case fatality rates from four hospitals. The rates are estimated from our sample. They vary enormously from 16 percent to 60 percent between the four hospitals. Just looking at the total numbers, you might infer that the hospitals seeing the most stroke patients give the best care and have the lowest case fatality rate. But in Maryland there is a great difference in the patients admitted into different hospitals. The easiest and quickest way to describe patient differences is to look at the racial distribution. The next slide shows how proportions of black and white patients differ from one hospital to the next.

The next slide shows one of a whole series of analyses to identify truly comparable groups of patients. With comparable groups of patients we can begin to look at the outcome of care in different groups of hospitals.

In this analysis we divided all "immediately admitted" stroke patients according to the severity of their condition on admission. Classes of severity are in ranking order and are exclusive. The worst class included all patients who were not conscious; the second identified those who were conscious but had swallowing difficulty. The third identified those with speech problems who were conscious and could swallow. The fourth category includes those with none of the three more severe conditions. Looking separately at the data for white males, white females, non-white males, and non-white females, you will see that 20 per cent of the white males were unconscious when they were admitted; fifty per cent of the non-white males were unconscious when they were admitted; thirty per cent of the white females were unconscious; and fifty per cent of the non-white females.

If you go to the other extreme and look at patients with no severe conditions, you will see 50 per cent of the white males; 30 per cent of the white females; none of the Negro males; and 20 per cent of the Negro females. These data may, of course, mean that blacks and whites have different diseases; that we are dealing with different age groups in the two races or that the two races choose to go to a hospital when they have different manifestations of disease. Hospital admission policy is another possible explanation. Whatever the explanation of racial differences, you cannot compare admission outcomes unless you adjust for at least the severity of disease at the time of admission.

One other point I mentioned, that of aggregations of disease, is also well illustrated in stroke patients.

In all Baltimore surveys looking for conditions predisposing to strokes, we have observed heart disease and general vascular disease behind a majority of pre-stroke symptoms. This association shows up again in our hospital admission survey. The numbers you see in the slide are from the reviewed records before total sample estimates were reconstituted. The slide shows data from approximately 4,000 stroke patient records. Different stroke diagnoses are listed across the top of the table and down the side are listed other major chronic diseases. The numbers and per cent of stroke patients with these other diagnoses are shown in the cells of the table.

The heart disease category shows the most obvious relationship. For every stroke diagnosis, a high proportion of discharge records have a secondary diagnosis of heart disease. More than 50 per cent of stroke

patients had at least one heart disease diagnosis. So, if you are looking at the outcome of stroke care, do not ignore the fact that a lot of patients have concurrent disorders such as heart disease which affect the likelihood of survival and recovery. Once more, we have to adjust for the presence of other diseases before we can say whether outcomes of different treatment programs are more or less successful. All these examples illustrate why the first question epidemiologists raise when they look at any evaluation is: Are the patients comparable?

The second question is: Have the physicians taken equal pains to make the diagnosis? That is, are we comparing the same diagnosis with the same diagnosis. In this example you see the frequency of a very usual diagnostic technique, E.K.G., in patients with a primary diagnosis of heart disease.

In this slide, I want you to look at the overall frequency with which the test was used and also its patterns of use in these patients. We have divided the regional hospitals into four groups according to their size. We have used the annual numbers of discharges as our measure of size. These E.K.G. frequencies are, therefore, tabulated from the smallest to the biggest hospitals. Remember we are only talking about patients admitted with a primary diagnosis of heart disease. In the smallest hospitals, only three quarters of the patients had records of an E.K.G. examination. There was evidence for 92 per cent in the largest. In all except the largest hospitals, the proportions of patients with E.K.G. examinations were lower in black (mostly service patients) than in the white (a sizeable percentage of private patients).

In our data, therefore, the degree of certainty that a heart disease diagnosis is correct is going to vary from one kind of hospital to another and from one kind of patient to another.

To repeat: the second question in the epidemiologists' mind is: "What was the extent of the effort that went into making a diagnosis and were efforts sufficiently alike in different hospitals and among different patients that outcome measurements can be compared?"

Next we consider the recording of the physician's diagnosis and medical information. What do the record librarians do with them. To get some estimate of the possible variations we should expect from this particular source, we took two or three troublesome diagnoses and sent them to record librarians in a majority of hospitals and asked them to "code" them for us.

This slide shows the three diagnoses and the specific International Classification Codes given to them by 29 of our regional record librarians. The first one is one that

all speak  
fifty-five  
hemorrhage  
in in th  
arians d  
be less co  
hemie at  
this really  
particularly  
average age  
chronic br  
obviously t  
ther than  
the difficult  
prevented u  
that disease

Another  
collective dia

I mentio  
vascular acc  
The Interna  
rovascular a  
This next sl  
patients in t  
the severity  
diagnoses of  
three per ce  
this diagnos  
mitted and 9  
major degrees

If you th  
rage, you  
were uncon  
per cent were

Epidemiol  
diagnoses as  
together in  
scribed by co

We looked  
to see what  
as having die  
hemorrhage  
cerebrovascul

One-third  
hemorrhage  
cerebrovascul  
third) of addi  
diagnoses we  
noses.

In the pati  
of those giv

I will speak about again later—cerebrovascular accident. Twenty-five out of the 29 librarians coded it as cerebral hemorrhage which is correct procedure by instructions given in the index of the classification manual. Four librarians did not use this code number. The librarians were less consistent for the second diagnosis of transient ischemic attacks. The third diagnosis of chronic bronchitis really gave inconsistent coding results. We were particularly worried about this disease because the average age of admissions with a code number for chronic bronchitis was between 30 and 40 years. Obviously the group includes patients with more than (other than) chronic bronchitis. In this case we felt that the difficulties inherent in coding chronic lung diseases prevented us from learning about true distributions of that disease from our samples of morbidity data.

Another classification problem arises in relation to collective diagnoses and group outcomes.

I mentioned above that the diagnosis of cerebrovascular accident had concerned us in another context. The International Statistical Manual suggests that cerebrovascular accident be coded with cerebral hemorrhage. This next slide shows specific diagnoses given to stroke patients in hospital records. They are cross-classified by the severity of stroke on admission. If you look at the diagnoses of cerebrovascular accidents, you will see that three per cent of patients admitted to our hospitals with this diagnosis were unconscious when they were admitted and 92 per cent were free from any of our three major degrees of severity.

If you then look at the diagnosis of cerebral hemorrhage, you will see that 50 per cent of these patients were unconscious when they were admitted and only 18 per cent were without all three severe degrees of disease.

Epidemiologically, these are two very different diagnoses as physicians give them but they are lumped together in the descriptions of groups of patients described by code numbers in hospital statistics.

We looked at our total morbidity and mortality data to see what proportion of the people we were counting as having died or been admitted as a result of a cerebral hemorrhage had, in truth, been given the diagnosis of cerebrovascular accident.

One-third of discharge diagnoses coded as cerebral hemorrhage had, in fact, been a primary diagnosis of cerebrovascular accidents. The same proportion (one-third) of additional non-primary "cerebral hemorrhage" diagnoses were actually cerebrovascular accident diagnoses.

In the patients' past medical histories, more than half of those given a cerebrovascular hemorrhage code

number had actually had a diagnosis of cerebrovascular accident. In a sample of death certificates, again more than a third had cerebrovascular accidents. Once more the decision of the "coder" to put cerebrovascular accidents with cerebral hemorrhages and the proportion of each in the total group of patients can make a lot of difference to end-point measurements. In our stroke registry, we code all diagnoses separately so our end-results for cerebral hemorrhage will probably differ from a majority of others.

I would now like to talk about a different kind of bias; one I mentioned earlier and one that I did not fully appreciate before we started this survey. Our usual morbidity data come from the records of our best hospitals. By "best" I mean the biggest hospitals with adequate record keeping facilities and the most accessible diagnostic indices. These are the only hospitals from which investigators and planners can easily get the kind of listings of record numbers and diagnoses needed to collect morbidity data. The following slides illustrate why this is so. This slide shows the status of our record rooms in Maryland at the beginning of 1969. Twenty-one hospitals (half) could produce a computerized list of their admissions and used the International Classification. Thirteen had a card file and used the I.S.C.D. We went through these hospitals card files by hand and made lists from which we could prepare samples. Seven other hospitals had a card file and used standard nomenclature. For these hospitals we had to develop a code compatible with our selected I.S.C.D. categories and we had to go through the card file by hand to identify all compatible diagnoses in the given time period. At the time we did the survey two hospitals were without a filing system. We sampled from all of their records for one year and read large numbers of records to get our balanced sample of patients with stroke, heart disease, cancer, diabetes mellitus, and chronic bronchitis. The crux of the matter is that the likelihood of getting a list of patient discharge diagnoses varies enormously from the larger to the smaller hospitals. An even harder problem to deal with, and one that limits available data more than the actual mechanization of the index system, are hospitals that fail to identify which listed or coded diagnosis was the reason for admission.

They simply write every listed diagnosis into their card file with no indication which one the physician listed first.

The next slide shows that the proportion of hospitals that can identify primary discharge diagnoses increases steadily from the smallest to the largest group. However,

not all of the largest hospitals identify primary diagnoses. This failure is a major barrier to collection of evaluation information. You may want to know about patients with heart disease. If you go through all the index cards and count all people admitted in a certain period of time with heart disease, you end up with a count of everybody who had heart disease listed in any ranking order among their discharge diagnoses. This specific problem almost doubled the staff work needed to abstract information for our survey. To make sure that my complaints are about systems and not medical records staff let me first show you evidence of the magnificent effort and cooperation of our regional medical records departments. We asked for about 21,000 medical records. The percentages at the bottom of the next table describe the few records the record librarians could not produce for our review. It was a total of less than 2 per cent of 21,000 records.

The next slide shows the extra work we undertook to identify the diagnosis for which each patient was admitted.

In this slide the "rejected" records were those pulled and reviewed but unused. The main reason for non-use was that the disease of interest was not listed first among discharge diagnoses. You can see from our "control" sample of all admissions other than heart disease, cancer, stroke, diabetes mellitus, and chronic bronchitis that 12 per cent of the records were not *included in the sample*. There were excluded for the following reasons:

1. the disease was not coded;
2. the record pulled did not match with any record number in our sample;
3. the admission was either before or after the defined study period.

With the major RMP disease diagnoses we had to reject as many records as we accepted. The difference between the 12 and 50 per cent was due to the non-primary nature of the diagnosis.

To summarize, available morbidity information is biased towards large hospitals. These hospitals differ from smaller hospitals in their patient populations, their availability of diagnostic techniques, quality of the information in medical records and its method of storage and retrieval. We should recognize this bias when we make generalizations about changing medical care and service programs on the basis of local and national morbidity information.

One further problem in using morbidity data from medical records that I will mention today is that of missing information and the bias it may have on your final interpretation of those data.

We have tried to look at the patterns of care and flow patterns throughout the region. One of our chosen measurements was the interval between onset of symptoms and admission to hospital. The next slide shows these intervals. I would like you to notice the "not recorded" column. About 20 per cent of all records in the sample were without information that would help us decide the delay between onset of symptoms and hospitalization. These incomplete records were concentrated in the smaller hospitals.

Any assumptions from these data about patterns of medical care have to be made with the knowledge that one in five pieces of information is missing. It is even harder to find information about the places patients were discharged to from the hospital.

We wanted to know where patients go when they leave acute hospitals. From the next slide we see that in some hospitals, 50 per cent of the medical records had no useful information on this point. Are we going to generalize our findings with the Maryland region—we cannot. The data we have apply to only a very small number of hospitals and patients.

We have tried to use other types of morbidity data in our region to get some baseline measurements for expected changes over time. They are summarized on the next slide. We have used death certificates. Some problems in the use of death certificates are mentioned on the table. We only use deaths to follow up the outcome of individuals who fell into our sample. We have been trying to trace deaths in all of the people that have appeared in all of our samples. This is a large scale operation. All names in our samples have to be matched with names that appear in subsequent mortality data. Once we get the death certificates, the diagnosis is always in question and steps should be taken to get validation.

We have tried to get information about out-patient visits. Those of you who use out-patient records know their two major obstacles: There is no way of getting a list of diagnostic problems unless they are listed by a secretary in a log book or clinic file as patients are seen; the out-patient records themselves have no "interval" diagnoses.

We spent all last summer in out-patient clinics getting information about visiting patients. We found that patients attend diabetic clinics for years, and their record contains no definite statement that the patient has diabetes mellitus. The diagnosis has usually to be assumed.

Other problems met in our surveys are: definitive diagnoses are rarely entered. Further, the information on which we could make a survey diagnosis is limited.

I am not going to talk about functional end-point measurements because I know Dr. Kelman is going to talk about them.

In terms of disease measurements, out-patient records and physician records have very limited value.

Finally, I would like to show you some of the ways we are using these different kinds of disease measurements to get estimates of regional needs.

One of the questions we have asked ourselves about our region in general is: Should every patient seen (somewhere) with a stroke diagnosis be admitted to a hospital immediately. We are not talking about patients who never appear in any kind of medical care facility, only those who appear somewhere in the health care system. If so, how many bed days would be needed. The next slide shows an example of the type of construction we are making to get this information. From our in-patient survey we have estimates of all the patients admitted to our hospitals in one year. From a surveillance of the emergency rooms of certain city hospitals we know how many individuals with stroke diagnoses visit the emergency rooms of those hospitals and are sent home. From these two sets of numbers we can get a total number of people with stroke diagnoses seen somewhere in the hospitals in a stated period of time.

We have not yet added into our sum of patients the out-patients with new stroke diagnoses we identified during our out-patient survey.

Now, what else have we done. We have completed a follow-up study of all patients seen in emergency rooms and not admitted in a defined period of time. We visited all living patients two years later to find out what happened to them since the initial emergency room visit. We also visited all patients admitted to the same city hospitals two years after they were discharged and asked them the same question. We also know whether and when any patients in both groups died, and whether and when they were admitted to other hospitals. We know what they say about their experience since the time they went to the hospital when some were and some were not admitted. By putting together these various pieces of information we can look at all "recognized patients with stroke" and see if there is any evidence of a difference in outcome for similar admitted and not admitted patients.

Our outcome measurements for this study are death and hospital admissions.

This is obviously a time-consuming and slow study but I hope it will give us some basis for estimating our

short-term general hospital bed needs for stroke patients?

The next question, as far as stroke patients are concerned, is: Do we need acute care beds for admitted stroke patients; or for how many patients do we need acute care beds? We are on the planning road towards getting the answer to that question. Four neurology centers have funds for acute stroke units. They have all agreed and have already started to set up standard criteria for all centers. These standard criteria will allow us to describe the patients in the same language so they and their outcomes can be compared. The standard criteria will also ensure that all patients have at least a minimum number of standard diagnostic tests. Each center will add its own special tests to its protocol but each has agreed to use a standard basic protocol.

Above and beyond this agreement to develop standard information in the four centers, we are working on the design of a randomly allocated therapeutic trial to allocate patients with different degrees of severity into our limited number of acute stroke beds and into other neurological beds. This study will identify the kinds of patients for whom acute care makes a difference in outcome.

This is one of the very tight end-points that Sam Shapiro was talking about and one that we believe has tremendous implications for the country as a whole. We want to be able to say how many (expensive) acute stroke care beds we need.

Finally, I would like to discuss one figure I borrowed from Dr. Matthew Tayback who is a member of our department. It is a beautiful illustration of a point Sam Shapiro mentioned about the need for comparison groups even when you are looking at changes over time.

Dr. Tayback has been looking at improvements in the outcome of pregnancy in relation to maternal and infant care programs. These outcome measurements show a beautiful downward trend in phase with program development. (Slide)

My colleague is wise enough to look at trends in cities who chose not to develop maternal and infant care programs during the same years. Curves are shown for prematurity rates and for neonatal mortality rates in nonwhites. The hard lines representing cities with maternal and infant care projects are mirrored exactly by trends in the cities without programs. The initial assumption that these programs are easy to measure because they have dramatic changes over time is proven wrong. It is very hard to measure the value of these programs because the other cities seem to be doing just as well.



This one pair of graphs illustrates Sam Shapiro's point that I want to emphasize—the need for comparisons even when looking at changes over time.

I would be glad to answer any questions about other aspects of our studies later in the program.

### Discussion

DR. CARPENTER: That's fine. You noted the complexities of analyzing data from existing medical care records, and Mr. Shapiro said it was a complex job to devise new records and get decent information from those. This is one indication of the difficulty of end-result analysis.

Are there any questions for any of the panel members from the floor?

QUESTION: With regard to the stroke patients, when you listed other diagnoses like heart disease, presence or absence, what criteria were used in deciding whether a stroke patient had heart disease?

DR. HENDERSON: The data I showed you are from hospital records that were already in existence. The diagnoses we used were abstracted from the discharge diagnosis. In other words, we copied every discharge diagnosis listed in the medical record onto our data form.

Our data say that whoever wrote out the discharge summary in the medical record listed this disease as being present.

QUESTION: The diagnosis of heart disease in these patients may have been based on EKG findings or not?

DR. HENDERSON: May have been based on anything the physician used to make up his mind that the patient should be given the diagnosis.

QUESTION: Are these face-sheet diagnoses or extracted from the discharge summary?

DR. HENDERSON: Discharge summary. Not face sheets. I had my abstracters copy the discharge summaries at length.

QUESTION: What was your hypothesis in getting involved with reviewing these thousands of records prior to the operation of the regional medical program?

DR. HENDERSON: We had four reasons for this survey: 1. To get information for planning. We felt it was really unrealistic to plan to set up new programs or extend programs unless we knew what was already in existence.

2. The second reason was to get baseline measurements for evaluation. If there were improvements over time, for example, in doing more EKGs when people

were admitted—then we wanted to be able to say there has been an improvement. So we wanted baselines from which we could measure improvements both in the process and in the outcomes.

3. We wanted to be able to describe our region in terms of patient movements through the medical care facilities, consultation, delays etc. We felt this was the quickest way to get the picture.

The best alternative was to take a group of people with each disease and follow them through the system for a number of years.

We decided to use a cross-sectional approach.

4. The last purpose was to identify comparison groups. We now have a pretty good picture of the people seen in all of our health care facilities in the region. If, as has happened, one area sets up a program for chronic respiratory disease, then from our data we can pick out an area that has similar patients that doesn't have a program and maybe we can make comparisons.

That was our rationalization.

QUESTION: Point of information. I was wondering how you could define morbidity.

DR. HENDERSON: How I define morbidity?

QUESTION: Yes. It came up in your discussion a number of times.

DR. HENDERSON: I suppose I was just using it loosely. In general terms it is a measurement of illness, as opposed to mortality which is a measurement of death.

There are obviously different kinds of measurements of morbidity. You can describe the disease itself. You can describe the use of services by people who have disease or by people who do not have disease as a measurement of morbidity. You can use length of hospital stay. You can use measurements of function.

I used the term in a generic sense meaning measurements of everything related to disease separate from mortality.

QUESTION: How long did it take you to gather this data? And in the meantime did you wait to start a program?

DR. HENDERSON: No. Let me explain our situation in Maryland. We have an epidemiology center which has been busy collecting these kind of baseline data and which is now working with the directors of individual projects to set up their evaluation schemes.

The center started after the Regional Medical program began and projects were funded before any survey were set in motion.

It is supported by RMP funds, but is administratively in the department of epidemiology in the Johns Hopkins School of Hygiene and Public Health.

It is an advisory and a scientific arm of the program administered by the University and it was very unfortunate that the program began before this particular activity was funded. However, in spite of the delays, the result of our surveys (now being cleaned up) seem to be coming at a good time.

In my opinion, (with its limitations) the region has seen a lot of activity as a result of initial funding of programs. People became interested and began to work.

Now I believe we know the active and interested members of our professional society and we are at the stage where we really need some overall direction. I believe the Center's results are going to be available at the time when some overall direction needs to be developed.

How long did it take us to do the record survey? It took us a year to collect the in-patient data and another three months to collect it from the out-patient clinics. Over this same period of time we have collected data from three sets of admissions to a sample of nursing homes so we know about their population and its turn over.

I think it has been a fantastically rapid job in terms of the amount of information collected.

We are currently having a lot of problems with analysis because the sampling frame was really confounded by all of these problems we met in getting samples of records and finding which ones were usable. We are presently working hard at sample estimates. To change our estimates we had to look at all of the rejected records (five thousand of them) and tabulate the reasons why they were rejected. We have just finished this exercise. Overall it has taken two and a half years to collect, process, and begin to churn out data.

DR. CARPENTER: I think at the rate things move in Western Pennsylvania data collected within three years is bound to approximate baseline data.

QUESTION: Were you able to differentiate between the care received and the disposition of the patient that actually occurred and what the medical person in charge would have wanted for them?

DR. HENDERSON: Not from past records. We are doing that kind of thing in evaluation of separate projects. These are retrospective data so they are hard to validate.

QUESTION: Not even on discharge placement, where their first choice of placement would have been?

DR. HENDERSON: You mean you go to the patient and find out whether they actually went there? We haven't done that. It could be done.

QUESTION: Or whether a facility existed that they could be moved to that would have been a physician's first or second choice?

DR. HENDERSON: We have not done that. We have collected a lot of subsidiary data. For example, we did a survey to identify all of the relationships, both formal and informal, between and within all our hospitals and between our hospitals and all other institutions. So we do know with which nursing homes and which other hospitals each individual hospital has relationships.

QUESTION: Dr. Henderson, would you care to give an opinion about the necessity to have a program which would significantly improve the hospital systems for data collection and data management in view of the fact that it's terribly expensive and very difficult to set up a modern type of information system?

Do you feel that the data that's needed is so essential that this is one of our major problems?

DR. HENDERSON: Well, you've got to separate this into the data needed for patient care and the data needed for overall planning. The speed at which these systems have to run is different for the two purposes.

The fast systems are the most expensive. Dr. Williamson knows much more about this than I do.

I think we need to have our systems improved, there is no question about it. The major data problem is quality.

Most people improving data systems are really taking no notice of the quality of the data.

In my opinion, which is an epidemiologist's opinion, a great deal of effort in RMPs across the country has gone into the technical improvement of data systems without taking any notice of information that will come out of the system in the long run.

Perhaps we tend to go the other way and place too much emphasis on the exact meaning of the information and its accuracy.

There may have to be some approach between these two points of view before we reach the best data systems.

But we obviously need an improvement in medical records systems.

The biggest holdup in Maryland, if you want to look at the speed with which information becomes available, is in making the record summaries, getting them completed and getting them into the record room. No system is going to do that. You have to get substitutes for the physicians or give the physicians time to write their summaries.

DR. CARPENTER: Dr. Williamson, do you want to comment on this?

DR. WILLIAMSON: (Johns Hopkins University, Baltimore, Maryland): Yes, I agree fully with what Maureen has said, and my own bias is against trying to throw a lot of money into developing a fancy record system when nobody has an idea for what purpose it's going to be used or what kind of decisions are going to be influenced by the information you will get out.

I think it's much better to try to aim at developing an assessment function within the medical care group and then especially to get the physicians involved so that it becomes a part of the problem of trying to attack and identify what problems or priorities to aim at first and then what kind of measure they want to make and then start to work backwards to the system to say, "Now, if we are going to measure this particular disease, this particular problem, we are going to have to have a much better form, and let's standardize it so all the physicians will use it, and we can get that and start collecting standard data on that problem to see if we can arrive at some conclusions as to where we can improve the outcome of care for these kinds of patients."

So I would strongly stress going after the function of evaluation and setting up expertise and getting physicians and members of the group involved with that, rather than taking some part of the process and trying to bring this up in a not very sophisticated way without the balance of the other parts of the system that will eventually lead to decisionmaking and altering the system itself.

DR. CARPENTER: It's almost what you were saying, Sam. One needs a special data collection system to measure a specific end-result. Existing systems seldom work.

MR. SHAPIRO: Well, by themselves they are almost invariably not adequate to serve the purpose of the kind of end-result studies that I was describing.

But at the same time I think that there is a danger even within the framework of these large studies that I was describing to overlook the important role that existing record systems may play.

For example, the multiphasic health testing program that I very briefly described and that's just getting underway in HIP is very heavily dependent for some of the evaluation on what will be found in the physicians' records.

There is being developed a retrieval system to obtain diagnostic and physician and other medical personnel utilization information from the existing records.

One of the questions—it isn't an end-result type of question. It's a process type. It's related to process. But one of the important questions that we're raising is the

extent to which the poverty and nonpoverty group differ in what is already known to physicians about the health conditions and the health conditions that are found through this health testing program and what the medical care behavior of people in light of what is known and can be retrieved from the medical records.

In every one of the studies with which I have been associated in the past, the existing medical record system has played a very important role. In some cases it has been more subsidiary than in instances to the specialized effort to obtain information, overcoming some of the problems that existing records pose.

QUESTION: Does the HIP regularly monitor or have a quality control system with regard to how physicians enter their medical records? What do you do about keeping up on a certain minimum level of quality?

MR. SHAPIRO: There's no continuous monitoring system for quality of the information being reported in the medical records that cuts across all records. But some of you may recall the quality of care studies that were carried out during the 1950's and early 1960's in HIP in which the information in the medical records provided a critical source for evaluating physician performance.

This left its mark on the system and had a very profound effect on the way in which records have been organized and maintained.

Also, the payment system within HIP does contain a provision for annual review of a sample of records in each medical group, for new entrants into the system and for entrants during the previous two years, and the quality of records is judged on the basis of this review, and money flows to the ones that meet the criteria.

So while we don't cut across all categories of patients in the system, there is a considerable amount of attention given.

I might also point out one of the strengths of the record system in our plan is the ability to retrieve information without breaking your back going out into the general community.

There are 30 medical group centers, and a centralized record system provides a very powerful means for efficiently obtaining access to the total medical care received in the system.

QUESTION: Many dollars have been spent on studies in the name of health planning, and I was interested in the suggestion that as a result of the work you have done in Maryland you anticipate some of the course of the program might be modified from some of the results of your study. I wonder if you could just enlarge on this a little bit.

DR. HENDERSON: Well, I'm speaking here before we have looked at our total data even for one disease. However, I do believe that it does point out differences in services, diagnostic services, differences between the severity of disease at the time people get into hospital, and also follow-up differences.

I think we have already made a beginning with stroke, for example. We have met with neurologists and pointed out that there are groups of patients who were not getting follow-up care. And I think just looking at our series of descriptions of patients, investigations and modes of therapy, the neurologists are going to come up with ideas about what is needed to improve care across the board.

That's not a very specific answer I know, but I have looked at enough of the data to believe we are seeing tremendous variations.

If the neurologists agree that certain standards of diagnostic investigation are necessary, and we show variations in frequencies of diagnostic investigations, it is the responsibility of the region's neurologists to begin to set up a program to see that necessary investigations are available to all patients.

We are trying to provide the clinical specialists with data they can use to make decisions about gaps in the care. I think we see enough variation to predict there will be enough gaps to keep everyone busy.

QUESTION: One of the morbidity figures that you showed in the first slide, Maureen, was prevalence. And I wondered if you agree with the viewpoint that prevalence should be one of the last measures one would ever use to assess the effectiveness of regional medical programs, in that prevalence, which is the frequency of disease at any one moment of time, is likely to go up if regional medical programs are truly effective.

And this is a somewhat embarrassing finding that we would probably not want to show, although we may want to know it ourselves.

DR. HENDERSON: That is very true. We are aware of it. We are trying to get some estimates of care needs in terms of prevalence, not look at the outcome in terms of prevalence.

The MIC program is, again, a good illustration of your point. In areas without good facilities before the program began we are getting an increase in the stillbirth rate, and an increase in the mortality rate, only because we are finding babies never registered in the past. So the rates are going up as the care initially improves.

DR. CARPENTER: Let me ask the audience and the panelist whether anyone has now in their hand end-result measurements which have led to new decisions in

your own region or health care system. Does anybody want to claim credit for that? That doesn't mean you claim good data; it just means somebody did something because you showed data to them. Sam, you must have had that experience.

MR. SHAPIRO: Yes. I was waiting for Maureen or Chuck or somebody else.

Yes, the breast cancer screening program has had a very direct effect on what is being included in our multiphasic health testing program. We're going to have mammography there. There's a move within HIP to include as part of the general physical examination not only palpitation, which ordinarily is included, but also mammography.

Now, this may sound like a trivial affair, but mammography is a costly procedure.

Now, if the information we currently have hardens over the next couple of years, there's— Well, I'm going to be the optimist to say there's no question in my mind but that there will be major efforts in many parts of the country, including efforts among those groups that are concerned with the regional medical program's regionalization and expansion of services, to include breast cancer screening.

In fact, in your area, Abe Lilienfeld has a project that ties with RMP to have every woman admitted to hospital go through mammography.

This program I am sure he will acknowledge is a direct consequence of the tentative findings that we have made in the breast cancer program.

DR. CARPENTER: Incidentally, were the mammography cases usually curable? That is, the cases discovered only by mammography?

MR. SHAPIRO: Well, I'm not going to be able to give you a direct response to that because our numbers are still quite small. But the histologic type of breast cancer picked up through mammography is more heavily concentrated of the intraductal type where there is evidence outside of our program that survival rates are much more favorable.

At this point, both those cases picked up through mammography and those cases picked up independently on clinical examination have very favorable and very similar types of survival.

DR. LOGSDON: I would only add as far as end-result evaluation of a test that the dental examination that included oral cytology had such a very low yield in number of positive cases that were thereby treated that this was deleted from the process rather than adding to it.

So that end-result evaluation can delete as well as add.

DR. CARPENTER: There's another good example then of how end-results can change the system — end-result measurements.

Anyone else?

Do you want to give us an example of how some of the end-result measures you have made, John, have motivated either your own institution or your Regional Medical Program to undertake health care a little differently?

DR. WILLIAMSON: I guess the two most dramatic illustrations might be, first, our heart failure study at Baltimore City Hospital, where we took a look at a range of outcomes from case fatality rates to people who were still out of work a year after leaving the hospital that should not have been out of work.

And having found that the results did not meet some very stringent criteria we set up, the administration of the hospital was impressed but didn't do anything. But then they did okay some more studies. And then one of my graduate students took and replicated the same type of study in another area and found the same kind of rate. For example, the case fatality rate was almost double that which they predicted under the worst of circumstances.

We identified that the problem had been that the care given during the time they were at the hospital was great but it was that year after they left. Then this other study found the same thing — they then appropriated money and hired some new staff and set up what they call a follow-up clinic to follow the patients after they leave even through the hospital may not have responsibility. They still wanted to find out what could they do to see that these patients get to another physician, to see that they do fill their prescriptions, to see that they are going to be followed. With heart failure there are disastrous results if they don't take certain medications and have certain medical care.

And this has resulted in, I think, quite an innovative approach to this whole organization of the clinic system, and we are rather pleased with what seems to be happening.

Now, the payoff will be to—which we want to do—repeat the study and now see if we find any different results as far as outcomes go or see if we are just measuring something where there are other factors that might explain this.

But this is a definite change that occurred in the whole clinic system as a result of these systems.

DR. CARPENTER: Good.

DR. HENDERSON: We have one that was not really an outcome. That's why I was not speaking. We did follow-up study of patients a year after they had been discharged from three hospitals in Baltimore with diagnosis of stroke. We wanted to find out what medical care problems they had had in between times.

One of the reasons for doing this study was that the Maryland Heart Association wanted to develop stroke programs and wanted to know the needs of stroke patients living in the community.

We found that many, many patients said that they could not get to their usual medical care facilities to get their blood pressure measurements, their pills, and the kinds of month-to-month care patients with this kind of chronic disease require.

They could not get there not because they were paralyzed and couldn't be gotten out of bed but because they could not speak well enough to feel confident to travel or because they were too insecure or unstable to go without an escort.

And as a result of this study the Heart Association was given a van, and it now has a transportation program—the van driven and staffed by volunteers. It has started to offer a free service to needy patients in the metropolitan area to take them to their medical care facility if they have no other means of getting there.

So we did have a particular effect. It wasn't RM sponsored, but it was a community organization.

The process of doing evaluations has, in fact, had numerous effects on programs. The simplest to describe is in our coronary care units. We have been looking at coronary care units throughout the region. All the units have beautiful patient information forms which include all kinds of measurements. Few of them actually measure and record weight.

We have been abstracting information from one of these units for some time, and they are now beginning to make much better attempts to get complete records. We are having a real effect on the recordkeeping systems of the unit.

Again we have been going to hospitals looking at the performance of nurses who have been through stroke education programs. We look for care plans and whether care plans changed after the nurses attend the course.

The process of evaluation inevitably affects ordinary programs. This is not decisionmaking; it is a sort of infiltration from the bottom.

DR. CARPENTER: Very good. Does anyone else have any examples?

Are there other matters then that came up this morning in the discussion of the problems of health

status measurement and the need to choose proper problems for the considerable effort required?

DR. MARGULIES: You know, I was impressed by a couple of things this morning. One of them is the sense of reassurance I got that in the years that I have left medical practice the problem of medical records has remained so stable that I don't have to relearn anything. It's about where it was.

But the other thing that I really wanted to raise on the basis of the experience of the people who have been describing their efforts in evaluation and measuring outcome is how important they feel this issue of medical records is. Obviously you feel that it is very important.

This then raises the basic question in my mind in any kind of evaluative procedure of having adequate information. And if you are going to measure what you are doing and measure the effects of any kind of activity, whether it's regionalization or clinical procedure, whatever it may be, as you very correctly pointed out, you have to have something that you can compare with something else.

For Regional Medical Programs that could very easily be a major goal—to look at the capacities which we have for influencing medical records, for introducing stability, consistency, and so forth.

Now, that's one aspect of it, but you also pointed to another which is of real concern, and that is the varying perceptions of medical record librarians of how they perform in a record system.

Have you pursued this particular issue further and do you have some advice for us?

DR. HENDERSON: Well, we have been pursuing it in several ways: One, through setting up meetings and instruction. As a matter of fact, it is not really instruction. Interestingly enough it is easier to use the International Code than a standard nomenclature. We have to "unlearn" the record librarians. We have tried to encourage them to change their use of codes.

We also have a pilot study setting up an educational program for a new kind of person that we are calling a medical records summarizer. I said earlier that medical record summary is one of the biggest hold-ups in the record system. We have had in our research programs for many years medical record abstractors who are very competent and can abstract a medical record perfectly well, getting all the detailed information we need. We are now trying to see whether we can train an assistant with this capacity who can be used in a service function to summarize the medical record to the physician's satisfaction so that he will sign it. This would really give us a much more rapid flow of records and we will get better

summaries of the essential information we want for both patient care and research. We currently have a girl who is over-educated for the position working with us to set up the content of a training program. We are hopefully going to add another couple of candidates in the spring.

We are comparing the girl's summaries against medical residents' in one special area after the other. A successful program would be a great step forward in speeding things up.

DR. MARGULIES: Of course, this still confines you to what you can do in improving medical records in hospitals. And on a continuing basis, as you pointed out, you have had to confine your observations to isolated incidents and, in fact, to patient response on the basis of their own experience.

DR. HENDERSON: Right.

DR. LEWIS: I'd like to jump in and comment.

I think besides Maureen's program with medical records librarians your comment raises the issue as to whether or not ambulatory care, traditional or radical, can ever be evaluated without a problem-oriented approach to recordkeeping.

DR. MARGULIES: That's really what I'm getting at. Can it be? I doubt very much it can.

DR. HENDERSON: No, I do not think it can.

DR. LEWIS: The second point—and it's even more subtle than that—is the problem of distinguishing between the actuarial content versus the contractual elements of the medical record.

Let me put it back. Sam and many of us would be interested in data that allows us to do a life table kind of following of what happens in time on patients from the actuarial kind of prognostic point of view.

In fact, however, when one looks at medical records in tracing backwards the history, one is a prisoner of the kind of medical information which that physician chose to write down which really was in part a fulfillment of his contract with the patient.

And one has a highly biased view of the world, much of which serves to remind everybody who will read that record that he was in fact doing a good job as he saw his job with that patient.

This sort of contractual, or legal, ethical reason, I think, is one of the more serious problems which has been cited by Garfinkel and others outside of the medical care system, and one which raises the question as to whether or not professionals can record actuarial information without the kind of super-structure that has been built in special long-term studies to get information that is other than almost a self-fulfillment prophecy.

DR. CARPENTER: Let me see if I can take a more positive point of view about the medical records. Our data was obtained from medical records, and we were curious as to what we could find out about the medical records, how bad or how good were they.

We tried to find certain expected correlations. One would think if a patient came in comatose that he ought to die more frequently than someone who came in alert. And this kind of correlation, in fact, we could find in the records. But maybe the correlation is so strong that even with a much muddleheaded recording in the charts it is evident.

If you get past the diagnosis sheet and look at what the doctor wrote in the record you can learn some interesting things.

For instance, in the county we studied, the significance of coma with no neurologic signs is really not adequately recognized. Often if the spinal fluid is examined and blood is found, the diagnosis of subarachnoid hemorrhage is not made. And if spinal fluid is not examined, the urine may not be examined either.

It may show unrecognized 4-plus sugar and 4-plus acetone.

So by getting past that face sheet into the details of care, somebody who is adequately trained can learn a fair amount.

We are now in the process of saying to the people in our study county, "Some of you lose more patients than others. The difference is not related to age, sex, or certain measures of severity." We also can say, "it looks as though you're not all doing an adequate neurologic exam. Generalists lose many patients without definite neurologic signs who are diagnosed as stroke. Similar patients (without clear signs) who are treated by the internists die less frequently."

These and other data lead us to conclude that though hospital records are imperfect, they do contain useful data.

QUESTION: Was there any evidence that hospitals that take part in the PAS program of the Committee on Professional Hospital Activities keep any better records than those who don't take part in that program?

DR. HENDERSON: No. No, we looked at that. The only difference was they could supply us with a printed index. That was very helpful.

DR. CARPENTER: Are any of your hospitals using any kind of automated history?

DR. HENDERSON: No.

DR. MARGULIES: Does the utilization of the screening program, the automated multiphasic screening,

have an influence on hospital records that you can perceive?

MR. SHAPIRO: We don't have the experience yet. We are going to become operational in November. So that's a very easy question to answer. We don't know. But this issue is one part, one phase of our evaluation.

I want to comment on the quality of records issue. What Chuck has referred to as the actuarial approach can be thought of in terms of prospective studies. There are enormous difficulties even under the best of circumstances when you try to use information collected during a previous period.

You have the problems of reconstituting a population. You have problems related to, again even under the best of circumstances, absence of information that did not appear to be terribly relevant initially.

This in no way detracts from the importance of major efforts to improve medical records, and we too are getting involved in new approaches to improve quality of records.

I want to emphasize that in every research project in which outcome measures have been used, we have depended on the HIP medical records in one way or another. The record has not supported completely the investigation, but without the record we would have been in terrible difficulties. Even in the "purest" types of outcome studies, existing good records systems can be of invaluable assistance.

DR. LEWIS: If I may make a comment, since the issue has arisen, I think there is a tendency to confound technology with validity, or neatness with validity.

For example, Maureen's comments about PAS hospitals who had a printout, but had no better records than those who didn't support this. I'm sure some of you are aware that lots of people are pushing automated history-taking and computerized forms so that the physician gets a very neat printout. The issue of validity seems to have been totally overlooked in a good number of these projects.

Whether or not it looks neat and comprehensive is one thing, but whether or not it means anything, whether or not anything has really been measured that is of any value, is another.

### How to Measure Health Status

HOWARD R. KELMAN, Ph.D.

What I would like to talk about are some ways in which — and I think for this audience this will not necessarily come as anything new or unique — others have

looked at health status and have tried to measure it focusing principally around the measurement or the determination of disability and related kinds of measures of discomfort or dissatisfaction.

I think it was probably Kerr White who first coined the five "D's" of measurements of health status — death, disease, disability, discomfort, and dissatisfaction.

And it seems to me to be as good a way as I can think of to define the different kinds of ways in which health can be thought about and determined.

Our speakers this morning have concentrated, or focused I should say, most of their discussion and attention on measures of, and utilization of measures of, mortality and disease or sickness, and I'd like to talk a little bit about the third D and maybe get a little bit into some of the other D's.

Of course, I couldn't help but get the feeling following the discussion early this morning that, why bother even to begin, when we have so much ground to cover in terms of defining really what the RMPs are supposed to do, to begin with, and to achieve and develop a kind of apparatus for assessing these largely undefined or global objectives.

But I suppose if we waited until objectives were clearly delineated and everybody was really sure about what they wanted to do, we might not even be meeting here.

Let me go on a little bit further and talk about disability measures and why and how it might be utilized in RMP programs — which I know very little about because my connections with RMP have been rather peripheral. That is, I've been approved but not funded.

I suppose it's worth starting out by asking: Why get concerned with disability or discomfort or dissatisfaction? What has that got to do with medical care?

I leave the obvious answer to that to you to think about for only two seconds, because, for a variety of reasons, we have become increasingly concerned with the social and economic and psychological consequences for living of individuals who survive medical care (or chronic illness) and what is done to them or for them or on their behalf.

The increasing concern with chronic or long-term illness and the consequences of that for individuals in terms of their ability to function physically, socially and psychologically has led to the desire to regard disability as a sequela of long-term illness and how this might or might not be affected by the care that people receive.

One of the major problems I think we face in trying to look at disability, to measure, to define it, and to then try to relate it to medical care, is that the further

you get away from biologic measures of how people function, the more the function of the individual is influenced by non-biologic factors such as their immediate social environment, their aspirations, their past histories and future desires.

So that what we might try to attribute to medical care maybe gets less and less influenced by what medical care can do and more so by what the patient's social situation is like.

I wanted to put that out to begin with because I think that sometimes we make the assumption — and I think I'm as guilty as anybody else — that whether a person can or cannot walk or will or will not go to work is due solely to whether he feels and actually is healthy or appears to be healthy.

For example, there are questions as you know in the national health survey which ask people whether their activities have been restricted due to illness. Well, this is a loaded question, it seems to me. It's something that I think we need to consider with regard to measures of disability.

The other thing I think we need to think about are the data sources for information of this sort which are different than those stressed by the previous speakers.

There is less dependency here on hospital records and those kinds of reporting systems with their degrees of unreliability and uncertain validity, and more reliance on a hard source of information — namely, the patient or somebody who cares for him.

Now, I know that it has been traditional to think of measures of social functioning as relatively soft and measures of morbidity and mortality as relatively hard. But I'm convinced by what the first two speakers told me this morning about how really soft the latter kind of information is — and I would say I'll put my bet down on the patient.

But, quite seriously, I think the whole question of the reliability of patient in terms of asking him how he feels and what he's able to do or not do for himself, or asking his relative or asking somebody who has observed the patient either in a treatment program or on a visit, whether it be a visiting nurse or an occupational therapist or an interviewer, does pose technical problems of reliability and validity which are related to but somewhat different than the kinds of problems that we have heard about this morning.

Now, in thinking about this subject and in some prior conversations with Dr. Carpenter, he inquired as to whether there were some kind of standard measures of disability, social functioning — you didn't use the word but I did "happiness" — those kinds of things.



There are "standard measures," and each person who does a study develops his own "standard" measure. There are good and sound reasons for this.

One of them is that it is exceedingly difficult to get any real consensus that goes beyond the confines of perhaps an advisory group about what you mean by disability and what you mean by social functioning and whether any of these things have anything to do with the program that was supposed to influence any of these states.

The other problem is that what might be regarded as disability in a person with physical impairments is not necessarily going to cover the same kind of ground for presumably well people out in the community.

So that if you are interested in a small increment of change, let's say, in whether a person can now dress with or without some kind of assistance because they have sustained some kind of motor or neurologic impairment, that would not necessarily be an appropriate measure or question to ask of somebody who is out there in the community and who is unemployed for one or another kind of reason.

If you wanted to develop a battery of measures of social, physical and psychological functioning to run the gamut from patients who may be nearly or completely bedbound to those who are both fully ambulatory and who work quite effectively as physicians or legislators or RMP coordinators, this is as yet a quite formidable task to get anything beyond the crudest kinds of information.

I think the other point that needs to be made is that those of us — and there are many of us in this room that I recognize and many who are quite expert in this field—don't view these measures really as replacing the more traditional and hard-to-get-at and harder kinds of information centering on, you know, mortality and morbidity, but really try to see these measures as perhaps other kinds of ways in which the benefits or lack of benefits of programs can be documented or tested.

What are some of the ways in which disability has been thought about and how have some people been going about it? Perhaps a word or two on that.

I have already referred, I think, to the National Health Survey, and I think it's important particularly for persons concerned with broad population groups and planning for their care and meeting their care, like RMP, to be aware of the kinds of information that are produced not only out of the National Health Survey, but also more recent studies conducted by the Social

Security Administration with regard to disabled persons and how they function in the community.

Essentially, the kinds of information that they collect are geared to basically well or "non-sick" populations, so that how relevant it is to populations of sick people is something you really have to decide for yourself.

But with regard to the question you raised earlier, Maureen, about denominators, I think this is where this sort of information may prove to be helpful if you can use the current information and if you find that the numbers are adequate for the population you are talking about.

Now, one of the problems, of course, is that these are usually national surveys, and depending upon the size of your local community, you may only have a sample of six or eight people in this national study.

In any event, you may be able, with the aid of very competent people, to relate the local population you are concerned with to adjusted rates based on these national sources. I think it's something that we don't ordinarily think too much about, at least in this area.

Now, some of the kinds of things that they try to collect information on in this survey— I'll just run through it very briefly. I'm sure that many of you are familiar with it.

They ask questions about days lost from work wholly or partially—the extent to which the individual has restricted activity days I guess is the actual term that they use—whether there are mobility limitations present or whether the person in a sense is either confined to the house or can get about without any kinds of difficulties.

And they also inquire, very interestingly it seems to me, about the person's social role activities — that is, the occupational information, if the person is a housewife or if a child, whether there have been any activity restrictions with regard to those roles.

I'm not sure what they do about people who are 65 and over, because we have no real social role defined for those individuals unless one can call retirement a role. So I think part of the problem in talking about people age 65 years and over is that they would probably score pretty low on these scales. They don't work, perhaps. They may never get out of the house. (I shouldn't say "never.")

Now, ranging from those types of very global measures— and I again want to emphasize that the individual is asked and his response categorizes him — to say whether limitations or restrictions in activity are due to illness limitations. Nobody examines him. Nobody decides *a priori*. The individual categorizes himself in terms of his response.

So that you have disability measures of that type on one end of the continuum, if I can put it that way, and at the other end of the continuum you have a variety of different kinds of measures of function which center around a core of what have come to be defined as ADL activities — toileting, dressing, feeding, ambulation, transfer activities. You name it and you can find “scales” for it.

And these kinds of measures have been developed essentially to look at rather severely disabled people, or those with potentiality of becoming quite severely disabled, who require a great deal of care and who have rather profound limitations in the ordinary activities of daily living.

And you will find on this end of the continuum a variety of different kinds of scales, all of which, or many of which, have proved to be quite useful in terms of evaluating change in patient status over a period of time or over a period of exposure or lack of exposure to one or another kind of treatment programs.

The scales vary in terms of the actual dimension that they are cutting. But what they really are trying to get at, it seems to me, is the extent to which not only the individual can perform at one or another level but the extent to which this performance is based on some increment of care or assistance, whether this assistance may be given by someone in the home, a relative, or by somebody in the treatment institution — the extent to which the individual can perform this particular function, dressing, toileting, transfer, etc., independently or dependently.

And the scale endpoints usually range from some level which is either “unable to” or “completely bed-bound” to, “can do by self” or “requires no kind of assistance.”

Now, investigators have usually used a battery of these scales, and in some of these scales different weights are assigned to different functions — ADL functions. Other scales do not assign different weights but rather give equal weight to performance on each of these functions. Still other kinds of scales are concerned with whether the patient needs help or doesn't need help.

Some years ago in a study we had underway, we found several different ways in which disabled people could be scaled or the scores manipulated, and we found that where there was change each of these scales revealed pretty well who was going to be changed.

Where there wasn't any change, it didn't really matter which of these scales were used.

Now, one of the other things I should mention with regard to these scales is that they have for the most part

been based on information obtained from professional people who know the patient. Sometimes it's a team making a judgment based on their experience with the patient, coming up with a group judgment about the patient or individual. In other scales it's a single individual who knows the patient, who may have worked with the patient, or who maybe sees the patient or former patient out in the community — a nurse, perhaps or an occupational therapist. Sometimes the patient himself or a relative is the source of information.

It may be almost like splitting hairs, but we sometimes seem to take these three rather disparate sources of information on particular individuals and throw them all together as though they possessed similar qualities of reliability and validity — and of course they don't really.

However, with all of these problems, as I said earlier, judicious selection and use of these scales has proven quite valuable in terms of determining whether a given program is having some appreciable effect on raising levels of function of disabled individuals or on whether it has reduced their need for assistance.

Particularly with regard to individuals in nursing homes or who require great amounts of nursing care, a small increment of gain from dependence to independence, let's say, in an activity like toileting can mean a great deal over a period of time in an institution where many in the population may require a great deal of care and assistance in terms of toileting.

Certainly I don't need to remind this group that a small increment of gain in toileting in a patient who has to be taken care of at home, while it may only reflect a jump from 3 to 2 on the scale position, may reflect a great deal more in a home situation if someone has the responsibility for the care of that individual.

I think then that one of the other problems with these scales is the fact that a stepwise jump from position 4 to position 3, while it looks mathematically neat, may not have the same kind of social meaning as a jump from 4 to 3 on another scale.

But these are generally problems of scale, and I don't think they are specific really to this kind of problem.

When we move from this more or less traditional area of definition of disability or disability determination and its application either to broad populations or to more narrowly defined clinical or patient groups, into the area of discomfort, into the area of dissatisfaction, into the area of social functioning with regard to let's say the family or the community, we get into terrain that is not nearly as well worked over.

I guess in large part we don't really think about or try to affect family relationships, if I can put it that way, when we think about stroke patients.

I suppose the connection between whether the stroke patient will now get along better or worse with his spouse, and the application on the other hand of medical measures to first see if you can keep the person alive and then to make living a little more livable for the person in biological terms is distant. Life saving does and should take precedence. But we pay a lot of attention, at least on paper, to social well being, and maybe we ought to begin to think of broadening some of our concern into some of these areas.

I shouldn't want to leave you with the impression that there aren't studies of social well being of well or sick people and that there aren't studies of family well being or compatibility, community participation and a variety of other kinds of social measurements; e.g. social isolation, work satisfaction, work performance.

But what I'm suggesting is that in terms of at least some of the kinds of programs that we are talking about, it may be well to think not only of scales which more directly seem to be related to biological efforts centering around disability, but also scales which seem — only seem — less related, a little more remote from our interests — for several reasons.

One is our own bias. That is, it may very well be that while we may be increasing the person's ability to function independently in one or another area of activity, this may have quite deleterious effects, when this person gets home, on his family. We don't know that unless we look at it or think about it.

The reverse may also be true. We may have very little success, for one or another reasons, in terms of basically affecting the physical level of functioning of an individual, but perhaps the application of other aspects of the program has had beneficial consequences in terms of how the family may now function or how the person may function in other kinds of areas.

I think part of the problem in moving into these areas is twofold. One is to make, as we all do, some kinds of decisions out of the plethora of dimensions of psychosocial functioning, those which have some kind of more plausible relationship to medical care programs than others.

And I think here that we do have a wide selection of — "scales" is hardly the word I think to use in this regard — but dimensions out of which scales that have been developed or can be developed can be applied.

Certainly it seems to me that with regard to sick people, and particularly with regard to some of the

comments Dr. Henderson made about followup studies that we ought to be interested in things like whether the patient is now better or less able to communicate, to use the medical care system, to manipulate it to their own benefit. Maybe this ought to be, if it isn't, one of the kinds of things we ought to be aiming at with sick people.

Their whole knowledge of what is wrong with them and what they might do about it, I think, represents another area that might be thought about with regard to looking at some of the kinds of programs that have been, or that ought to be, developed.

What I'm suggesting is that for a variety of reasons we may not be able to affect very basically the biologic functioning or biologic status of many disabled individuals. We may be better able to affect some aspects of the individual social situation, his social or psychological functioning, or the function of those around him.

I don't know why, for example, the National Health Survey doesn't ask at least for information from family members, and what their input is in terms of care of the sick person.

That is, if the individual replies that he is not able to work because of illness, oughtn't we to get information on whether the social role of some other individual has been altered as a result of that? Is that not really part of the disability picture that we all see pretty often? Does this now mean that somebody else in the family is now working? For somebody that is disabled and cannot work, does this now mean somebody else in the family situation's work role has been affected?

What I'm suggesting then is a kind of broader view that we might think about with regard to the plethora of effects that programs that have been developed ought to be looked at in terms of status or benefit.

Now, the obvious retort to that is that you can extend the concept and idea of health status to a point where it begins to be so diluted as to lose its meaning. But I don't think that some of these kinds of questions that I have raised or some of the areas of inquiry that we ought to be undertaking are that far afield for us at least to think about.

There are also areas in which there is enough methodologic experience and technique and enough familiarity in the health field in terms of sample survey that a ready transplantation — and I use that word advisedly in this particular context — of these kinds of efforts would appear to be appropriate, with some cautions.

And I want to end up with noting some of these cautions and then see what you have to say about these kinds of things.

These kinds of studies — you have heard how much it costs from Dr. Henderson. Well, going out and interviewing people or sending out people into the community to ask and get detailed information about activities of daily living is more expensive. It's more expensive to generate, to reduce, process and make available information from this source than it is using available hospital records with all of their limitations. And I want to mention that because it's a consideration we all do think about even today.

The other thing I think one has to think about is the practicality of obtaining many of these measures. I recall in one study that I was involved in we were concerned, in addition to getting physical and social measures of functioning, with getting psychological measures of functioning. And this involved first obtaining and then sending a highly qualified, highly trained clinical psychologist into nursing homes with a suitcase which he opened up and then did his testing in front of the patients.

This whole apparatus — and I won't even get involved in terms of the development of this procedure, ran anywhere from one to three hours.

Well, I'm not suggesting that these very detailed kinds of measures on memory, on judgment and recall are the kind of thing that ought to be done routinely, but there may be some programs where this kind of measure would be entirely appropriate and may be the most relevant criteria of benefit for some kinds of patients and therefore shouldn't be excluded. But it is expensive.

Again one or two comments. The data, of course, considering again the source, is highly — well, I was going to say highly reliable. Relative to other forms of information on mortality or morbidity it is no more or no less subject to problems of reliability and validity than these data are, although the problems are different.

And, finally, I would end up with just a reminder that, as I said earlier, the further one moves away from physical and biologic measures of function, the more the actual functioning of the individual and the patient is going to be influenced by things other than what was done to him medically or what his biologic status is. And this presents a problem in terms of evaluating programmatic impact.

I think I'll stop there and ask if there are any questions that I could try to answer or any points that you would like to have me try to elaborate on.

### Discussion

DR. CARPENTER: Thank you very much, Howard.

I think you said that it's often worth measuring disability and I was fascinated that you talked about measuring family disability, not just patient disability. By the time we get that on the front sheet of the medical record, we'll be quite far down the line.

DR. KELMAN: Not in your or my lifetime. Well, maybe yours.

DR. CARPENTER: It's interesting you pointed out that sometimes it's hard to understand the validity of a measure of death unless you know a concomitant measure of disability. Dr. Stoneman pointed out that probably those of our patients whom the internists appear to have saved went home comatose and wet the family bed for ten years before they finally died. And so it is necessary to measure both death and disability to understand the value of their treatment. By the way, disability on discharge was the same for both physician groups.

QUESTION: As you were speaking of different ways of measuring health status, one thing struck me. I think some of these measurements have to be reproducible if we are going to use them in evaluation.

In evaluation evidently we are going to pick them up at one point in time and then later on pick them up in order to evaluate programs. How can we pick up reproducible measurements?

DR. KELMAN: I think many, if not all, are reproducible. I think the question is whether it's a measure you want to get and whether it's relevant to your program.

For example, it is not difficult to ask one or more times or of one or more points in time, not difficult to get from the patient the answer to a question, "How do you feel?" I think the prior question is: Do you want that piece of information?

It is not, I think, difficult. Now, some of the considerations you would have to take into consideration are: How stable is that feeling state? Is this going to be something that he is telling me right now and is it going to be based on what has happened to him in the past five minutes, or is this a more or less enduring state of being that I am concerned about?

One of the things in the program may have been directed towards altering favorably the feeling states, the moods, the emotional status, whatever you want to call it, of certain kinds of patients. So that may not be the most efficacious way of getting information.

But you can get reproducible information by asking those kinds of questions.

Now, whether they are the kinds of questions that relate to the kinds of information you are seeking is the prior question.

This is true also of measures of social function and measures of activities of daily living which are a little more enduring than the example I cited but I really used that just to make a point.

DR. WILLIAMSON: Howard, if you were to recommend to us one key reference in the literature on the validity and reliability of disability measures, what reference would you recommend?

DR. KELMAN: On the validity and reliability?

DR. WILLIAMSON: This question of looking at the reliability and validity or usefulness and general applicability of these measures. What literature could be pulled out that would get us going in studying this more thoroughly?

DR. KELMAN: I think one of the places I would start is with— I forget the author—but it was a monograph put out by the National Center for Health Statistics.

DR. WILLIAMSON: Sullivan?

DR. KELMAN: Sullivan I think the name is. I think that's a good reference to start out with not only because of the kinds of questions he raises and how he tries to relate disability to the broader questions of health status, but also because I think he has an excellent bibliography.

I think the article by Ellinson in the *Handbook Of Medical Sociology* on sociomedical measures or measurement problems is an excellent discussion of methodologic problems.

DR. WILLIAMSON: Levine's?

DR. KELMAN: Right. In that book. I think you would do well, if you haven't already, to write to Murray Wylie and get some reprints from him.

And I think with that a person would be well armed and well acquainted with not only the problems of the application of these kinds of measures but their potential and actual utility.

DR. CARPENTER: You can also look up Kelman in the literature. That will get you a long way down the road.

DR. RIKLI: A small observation. Those five D's that you attributed to Kerr White — I have heard them on many occasions — are most useful in taking a project or a program and running down those five.

And as you talked about disability, you talked about the independence. It seems to me that probably a sixth "D" might be dependency — financial dependency, emotional dependency and physical dependency. And I think that's probably the greatest concern of people — when a parent or uncle or aunt becomes dependent on them in some manner. And dependency is measurable,

and I think that is one of the parameters you have to watch pretty carefully.

DR. KELMAN: I agree completely. And it's my impression that most of the scales you get into with the disability measures, whether it's vocational or occupational or activities of daily living, are really geared towards estimating how dependent or how independent a person is either occupationally, vocationally, socially or physically.

DR. RIKLI: I'd just like to add one point there, and that is about disability. A man may be missing a leg or missing an eye or have many other disabilities which are really compensated for and are not really of serious concern to society when a person makes the adjustment. But if they are unable to adjust and have a dependency, then they become a serious concern.

DR. KELMAN: Right. And your comment reminds me of something, namely that we have to distinguish I think between an impairment such as this and a disability. They are separate things. There are many of us who function with a whole variety of impairments quite well.

That is, if I were an engraver, with my level of impaired vision, I might be quite disabled occupationally. But in terms of the kind of vocational situation I am in now, I'm not at all.

These are other problems, and this is part of what I was trying to get across in terms of the point I made about when you begin to move away from the biologic functioning of individuals to estimating how they function in social terms and in social situations. The biologic becomes less influential. Not uninfluential, but less influential.

I'm glad you raised the point of distinguishing measurement of disability from measurement of impairment. They are both important, but they are different kinds of things. We sometimes tend to think that when we measure impairment we are measuring disability, and vice versa, but we are not.

DR. CARPENTER: It's hard to get the diagnosis adequately on the front sheet of the chart and a little easier I think to get survival indicated on the front sheet of the chart. Are there any obvious measures of disability on discharge that could be coded on the front sheet of a chart?

DR. KELMAN: I think there are a number of things that would be very useful to try to get in some standardized fashion. I think it would be extremely useful to know a few pieces of information. (I say that as though it's so easy and so simple.) The extent to which an individual is able to perform certain limited activities of

daily living—and God knows why on any hospital record you can't find out how much education the person has. I wish we could get that in. And where the person is going to, more explicitly than "discharged, improved." If we could routinely know whether the person is returning home or to some alternate living situation, I think would be a way of starting out.

Again, if you can do that in some kind of standardized fashion, fine. But to have the ward clerk or somebody else, you know, just scribble down some things is not worth the bother.

DR. LEWIS: I was looking for some of the participants in a four-center contract from the National Center now, and I guess there isn't anyone here.

The Harvard center, Western Reserve, Syracuse and Johns Hopkins, have a contract with the National Center for Health Services Research to develop a classification system for patients that deals with three levels:

- The problem of the patient, the actual management of the patient.
- Second, the problem of institutional management, in terms of length of stay—the usual issues that an organization or institution is concerned about.
- And a third level of coding which has to do with interorganizational needs — in reality, what the community has to furnish patients in the way of extended care facilities, etc.

The classification is in the early stages, first developing a common language, and is designed to work at several levels like any taxonomy which progresses to deeper levels in which information is going to be obviously less and less available.

This is an attempt at the kind of classification which would lead to a series of codes some place on a record that would describe a functional disability, to an institutional problem and interinstitutional problems.

And, as I say, I don't see anybody from one of the groups here.

DR. CARPENTER: And this language will describe activities of daily living?

DR. LEWIS: Part of it will. In essence it looks at a lot of the kinds of things Howard has been talking about. It tries to take into consideration all of this. They have been reviewing the literature trying to develop between four institutions a common language so that when somebody says they are impaired, for example, in mental status, they will now know it is coded, in terms of disoriented by place, by time, by person, etc.

It poses a real problem in numerical taxonomy.

MR. SHAPIRO: Well, Chuck, does this effort contemplate major changes in the contents of hospital records?

DR. CHARLES LEWIS: But looking at institutions to examine the feasibility of recording information, I think the real question is the one you raised: If you have a marvelous language which is somehow or another codable in a series of digits, so what? How will it be accepted? How will it be involved in medical records? To what extent will it actually influence patient care, organizational behavior, interorganizational behavior?

But I think that's maybe a remote question because the real issue is that there is no way of communicating this between institutions, between patients.

This is an attempt to try to standardize — to deal with Howard's original point that everybody starts out by inventing a new wheel.

DR. KELMAN: There was an attempt — some of you may be familiar with it. I haven't heard what happened to it, but it was called "rehabilitation codes." That effort involved a number of advisory committees who for years tried to develop a common way of coding relevant information for patients in rehabilitation and related kinds of programs, institutions and facilities. And they developed reams and reams of material. I don't think it was ever used much by anybody.

I don't really know why it wasn't, because there were many, many places and many, many people and advisory committees that worked on development of it and worked very hard, and a lot of it was good. Maybe all of it was good.

But I guess there's a different set of problems involved in developing these beautiful codes and then trying to take that and translate it operationally in terms of some ongoing system like a medical care system. And I really don't know what happened to it.

### The Relation of Process and End-Result Evaluation

CHARLES LEWIS, M.D.

I want to approach this from the stance of an operator, somebody who has to make decisions about evaluation data as well as someone who is supposed to be providing it.

And I will assume at the beginning that we evaluate things in order to change things, not as some form of self-amusement, (which it does turn out to be sometimes), but in order to provide some guidance for those who would like to really change the way things operate, *if* they need changing.

Now, I'd like to restate very, very simply what was said more eloquently this morning. Something — and I have decided to call it a condition, not a problem, not an

event, and not an in-put — just a condition at a time zero, whatever time you care to choose that to be — is usually measured in some kind of units.

And the units are hopefully relevant and possible of being measured, assessable, and hopefully available. And I think we would like these to be valid, replicable, practical, and sensitive.

For operators, the thing we are currently concerned with is that the condition needs altering, or else there is a question of it being altered. After looking at this condition we do something, not just anything but something specific, and that the thing we do is also measurable or assessable.

Having done something, a whole bunch of messy things happen that are called processes. And I would just say that one man's process is another man's end result, that somewhere in here people may choose to stop and say, "That's all I'm interested in."

And this is particularly true I think in looking at continuing education, in which maybe all we want to do is show they were sitting in the room.

The next thing we may decide we'd like to know is that they sat in the room and learned something.

Then we'd like to know if they took it home they did something with it.

Now, as I have just indicated here, most of the times when we are concerned with process we are concerned about the number of things that are done, the number of things used, the nature of things done or not done in terms of quality. Basically, process evaluators count heads, or something, or the use of things. People who look at disability, deaths, and so forth, as in the morning's discussions, are concerned with end results.

The major point I'm going to make — I hope — is that it is difficult to affect change without doing both, that end-result and process evaluation need to be carried out conjointly if one is going to be an applied evaluator and attempt to use results to redirect efforts.

Let me just point out some of the other things that by some of these terms I think relevant.

The use of evaluation data depends upon two sets of factors:

One, organizational factors. Organizations need to maintain themselves. They need to perpetuate the status quo, their prestige and individual's vested interest. Evaluation basically questions the reason for being in a certain business, and doing certain things. Fear of the consequences of change, change in rank, or change in the structure of an organization are certainly sufficient causes to reject evaluation data.

The second factor is the state of the art of evaluation in general.

If we present those who are coordinators of programs with end results which say "it worked" or "your program is good" — that's all they want.

But if one is going to present someone with information which is other than socially acceptable, it's useful to be able to tell them what processes went wrong, because this provides alternatives for strategies in terms of restructuring programs.

Donald Campbell and others have talked about the problems of reforms as experiments, and the social legislation that has been enacted to create social change and why evaluation of these programs has been so difficult.

If you had a million dollars riding on a program in which it was announced *a priori* there were no alternatives to success except through this approach, you have some idea of why individuals resist evaluation (at the risk of going out of business).

The failure to specify strategies, alternate strategies, for experimental programs creates a problem.

Perhaps one of the few ways we can deal with this type of program is by looking at the nature of the processes that went on while reaching an end result and presenting these data to those who have to make policy decisions.

This is particularly important, I think, if one is going to institutionalize experimental programs — that is, change the way people do things. The transfer of a program which seems to produce results into a different setting is difficult. Unless one has some idea of what went on.

Maybe this is related to some of the problems in disability evaluation.

I didn't stop and spend as much time as I should have here talking about the measurement of "do something." I think there are probably more "good" programs that have succeeded because the "do something" was, in fact, a phantom treatment that never got done than other kinds, in which something rather dramatically happened.

It's very useful to know what it was you did that made a difference, and I would just suggest as you look at the literature (as one moves from clinical trials of drugs where we are sure we injected "something", to a program in which we install a new kind of health manpower) that we really don't take the same consideration to standardize the dosage, the blood levels, and other things that we are concerned about. The process of evaluation begins with knowing what the experimental treatment was.

I have selected out of the literature eleven papers that are concerned with evaluation, and I'd like to comment on them and then talk about the kind of evaluation that was done in each of them and what it would mean (to me) in terms of trying to implement these results.

Let me begin, though, by over-simplifying certain classes of end, and process, evaluations that these papers represent.

The first one and the simplest type is the reporting of end results — end results in a group of patients or institutions that describe the fact that different things happened.

The second class or type of paper looks at variations in end results between groups or among groups or within groups, as a function of patient or doctor characteristics.

Type three is similar to type one but related to process evaluation. These describe what happened — the processes that were carried out and how they varied. This ranges all the way from results of chart audits and a whole bunch of things that are done to people or things that are used.

A little more complex, and fourth, is the study of processes of care as a function of certain provider characteristics. This is an attempt to describe the differences in the way processes were carried out as a function of the professional's background, training, and so on.

The fifth type, is a look at both process, or treatment done to somebody, and the end results of that treatment, without any comparison to other similar events.

In the sixth class, there are two processes — one I have listed as "C" for control, in which there was no treatment, and an examination of the end results among two populations or groups with different kinds of treatments.

I'm staying away, in this discussion, from the kind of complex experimental designs that many of us would like to carry out and are very comfortable with in the laboratory, i.e. cross-over, factorial designs. Because they don't come along very often in the business we are involved in.

There are some other kinds of quasi-experimental designs that are possible such as a time series observation that was pointed out this morning, regression discontinuity designs, etc. I refer you to the paper by Donald Campbell in the *American Psychologist*, for discussion of these.

With this very crude and perhaps debatable classification I'd like to go over eleven papers. I really didn't choose these with any bias, except that they illustrate these types of evaluation.

1. The first one was a sub-study that came out of the national halothane study of the incidence of hepatic necrosis with halothane. This was a report of institutional differences in post-op death rates. Among 34 hospitals, the end results (death rates) in surgery varied by a factor of 27. They were subsequently adjusted for a few things like age and sex and other things, and that difference is resolved to 10-fold. There were subsequently readjusted for severity of procedure, and the difference collapsed to 3-fold. This is the kind of study which says the death rates in hospitals are different — nothing else — and if we age-adjust and do some other things that we know how to do, they are still different, but we don't really know why.

2. The second paper is by Leon (Gordis) on the evaluation of a program for preventing adolescent pregnancy. This is a paper that looked at a program in which teenage girls who were sexually active were treated in a special clinic by social workers, by physicians, gynecologists, and placed on oral contraceptives.

The design then was to follow these girls to determine how many of them stayed under treatment month after month. About 50 percent dropped out of the program within the year. The characteristics of those young ladies who did not stay in the program versus those who did were compared.

3. The next paper, an evaluation of community nursing services in the care of the mentally ill, was done by Tayback. It looked at what happened when a bunch of patients discharged from mental institutions were provided services by visiting nurses in the home, in terms of a criterion called rehospitalization. The result was that there wasn't any difference among control and experimental patients.

The paper raises some interesting questions as to why there wasn't any difference. I think from the description, I might point out there wasn't any standardization of treatment. One really didn't know quite what was being done and how this might have varied or how certain subgroups of women might have had a better prognosis than others. In terms of looking at the probability of rehospitalization as a function of the patient, this is another kind that fits in second category also.

4. The fourth paper comments on genetic counseling. And if any of you know any other studies of the efficacy of genetic counseling, I'd appreciate knowing them. This is about the only one I have come by.

Families who had had one or more defective children for whom the genetic inheritance patterns were known, were provided counseling services (not further described)



and then followed forward for a period of time. Approximately 60 percent of the patients went ahead and had another child. It would suggest about 40 percent of this counseling, however it was done, had some effect on further child-bearing.

Here again there was no discussion of the effects, no discussion of the characteristics of patients. It represents a straight-forward statement that so many children were born who had major congenital anomalies or minor congenital anomalies to families who had been counseled.

5. The fifth paper presented is from San Francisco data on the neighborhood clinics for a more effective outpatient treatment of tuberculosis. This was prompted by some observations that (in San Francisco) about 80 percent of alcoholics, (50 percent of blacks and 20 percent of Chinese) broke their appointments to the TB outpatient clinic.

The public health department went into each of these neighborhoods, organized clinics with the help of the local citizenry. The compliance rate with broken appointments, sometimes used as a measure of satisfaction, dropped to about 5 to 10 percent.

The interesting thing about the paper is that nobody reported whether or not there were any readmissions or active cases of TB.

This is a discussion essentially of processes and change in processes related to the structure of a program, which, oddly enough, did not look at the payoff — which is whether or not any of these tuberculous patients complied with their medications, or were readmitted to hospitals.

6. The next paper is a study of variations in the incidence of surgery. This was a study which looked at all Blue Cross subscribers in the state of Kansas and looked at the incidence of certain common operations, T&A, appendectomy, etc., in various economic subregions of the state, defined so they'd be fairly homogeneous in nature.

The "Glover" effect or variation in rates for tonsillectomy was reconfirmed, as was a 3-4 fold variation in rates for appendectomy, cholecystectomy, and a variety of other procedures. The rates for surgery were studied as a function of the availability of surgeons, beds, and general physicians in the area. The percent of the variance of these rates that could be explained was rather phenomenal. For appendectomy, 70 percent of the variation could be explained by beds and surgeons.

It has some interesting implications, but it doesn't say anything about the consequences of these surgical pro-

cedures. It looks at processes as a function of certain variables in the structure of medical care.

7. The next paper is by Thompson and his group at Yale on end result measurements of the quality of obstetrical care in two U.S. Air Force hospitals.

Thompson looked at two Air Force hospitals and perinatal mortality by race, and found out that in one hospital, the black perinatal rate was higher, but in the next hospital the white prematurity rate was higher.

He went back and looked at utilization of care by trimester of pregnancy and found out that all of these ladies were using prenatal care rather early. It's a very fascinating paper because the more you read it, the more you have trouble reconciling some of the results.

8. The next paper measured the quality of medical care through vital statistics. This is a comparative study of appendectomy rates in the hospital regions around Rochester, New York. There were large variations in rates at which appendectomies were performed. And no relationship was found between rates of appendectomy and deaths due to appendicitis — an example of looking at a process, and the variations in process as they relate to an end result.

9. The next study of comprehensive outpatient care in rheumatoid arthritis is one of the ones that deserves reading if you're going to read any of these. In this one Dr. Katz does several things. He defines the condition that he's trying to deal with. He measured disability with all of the problems that Howard Kelman mentioned earlier this morning. He describes the processes of care for a group that got physical therapy, nursing, public health nursing, comprehensive team approach, and describes it very well. He measures outcome, significant changes in disability, as a result of applying comprehensive care for ambulatory patients with rheumatoid arthritis.

10. The next one is a study that we did in Kansas on continuing medical education. This is a study which basically looked at the tremendously aggressive program in continuing education that had been mounted at the University of Kansas for over 30 years with circuit riders, with regional courses and with conferences and seminars held at the medical school.

It was an attempt to look at the participation of all physicians in the state for each year at risk over a ten-year period.

We took a look at the predictors of use, as a function of physician characteristics, and found among other things that it's related to being near a regional center (having it available); being a specialist, and being a recent

graduate, but not at all related to place in class on graduation.

11. The last is from the nurse clinic study at the University of Kansas by Barbara Resnick and myself which looked at activities, events and the outcome of ambulatory care in which a population of patients with defined chronic illness were previously examined, described in fair detail, were randomized into two groups. One went back to medicine clinic, the other received care by nurse practitioners.

The critical incident technique was used to try to measure some of the activities of the nurse clinic, some of the things that John Williamson and Paul Sanazaro piloted. We looked at outcomes; death rates, in which there was no difference; the level of disease, no difference. There were significant differences in disability rates at the end of one year of care under these two systems; the nurses' patients were far less disabled. There were significant differences in discomfort and satisfaction levels. This paper attempts to look at processes and outcomes.

If I were presented with the data on institutional post-op death rates I would say, "I don't understand why our hospital is either so good or so bad." But I don't have any answers, and if we were good I'd be happy.

I think that regarding the second paper, evaluation of the program for preventing adolescent pregnancies, I would say, "This looks good, but I really can't tell what you're doing to these young ladies, and I really can't tell if anything is happening. Therefore, I think you'd better try to measure what you're doing to them a little better if you want me to pick up the tab for this kind of a program after the grant support wears off."

For the third paper, an evaluation of community nursing services, I think the comments would be as for the previous study.

This comment on genetic counseling. I don't know what you can say when you're confronted with information that says patients don't do what doctors tell them to do except begin to deal with their patients in a little more sophisticated way.

For the TB clinic study, this looks good on the statistical sheets, but did anybody get TB? Again the lack of outcome data creates major problems.

This morning when Bob said, "Does anybody here have end results that influenced decisionmaking?", Sam Shapiro talked about mammography, and someone else mentioned dental cytology. And John talked about the heart failure study and the creation of a follow-up clinic

that was discovered when it was found out that the deaths occurred after discharge.

Let me tell you about one that I'm willing to talk about, and it's a negative one, about how process information, and perhaps some outcome data influenced program planning in the Kansas Regional Medical Program. Perhaps we can get a postscript from Bob Brown who is now in charge of the program.

In 1967, the very start of the program, we like everybody else were trying to get people involved and trying to convince everybody it was their program. No one believed this.

We were always saying, "If you just bring us projects, we'll help you get them funded." And they brought us one from an area in Kansas that has some problems with economic growth, where the population was relatively aged, the physicians likewise, and no younger physicians were going, and there were lots of rehabilitation problems.

Some of the people in that area said, "We want funds to train assistants in occupational therapy and PT assistants, because we have a junior college, and we can train these people, and then they'll provide our rehabilitation."

We said, "Fine. We need some data to support it."

We had done a survey and were quite aware this was a very disabled population.

We also took a look at the occupational and physical therapy facilities in hospitals in this nine county area and found without exception all of them were operating at less than 50 percent capacity.

We interviewed a sample of about 50 percent of all practicing physicians in this nine county area and we sent our young ladies to them, and they asked:

"Have you seen anybody who needed occupational or physical therapy?"

And then there was a little probe to explain what occupational therapy was.

The next question was, "Did they get care?"

The final one was, "Do you think we need more?" — to which the answer was always yes.

When we took this data back, we were able to say to the people, "Look, you have lots of problems, end results that need to be changed; but you have facilities that are being underused. There are occupational and physical therapists who are going to leave their jobs because they don't have any work to do."

If we look at who creates demand for rehabilitation services (doctors) and talked to them, we found that they (the doctors) were not aware of the need for this

service and had identified patients for whom these services should be prescribed.

We didn't try to make any interpretations. We presented this to influential citizens whose comment was, "It looks like we have a job to do with our own doctors."

I don't know, Bob, whether there is still pressure for this. But I think that in one case we were able to show that by looking at the processes, that is, why patients who need care do not get it, we were able to avoid spending some money at least at that time.

I have asked some of the experts around the room to give me some feedback on some questions that I have raised. I think I'll start by asking Sam Shapiro. It seems to me that one of the reasons you have been so effective, Sam, in influencing programs is that you really have been looking at end results, but also describing to your own group the processes that they were pursuing and carrying them right along with you.

### Discussion

MR. SHAPIRO: Yes. Well, Chuck, I almost have to say "of course."

The influence of an end result observation is going to be very heavily affected by the ability to understand the process by which you achieve the end result, and as much attention has to be paid to the issue of process as the end result.

The only reservation that I would have is that there are occasions when it becomes incredibly difficult to tease out of the situation anything but very, very global descriptive information about process. But yet the end result in itself can be a very firm one. And I have a very specific situation in mind.

Some time ago we looked at the question of perinatal mortality and prematurity in HIP in contrast to the rates among patients of private physicians in the community and did all the necessary standardizing. We came up with a finding of lower mortality and prematurity in HIP.

And the next question we raised was: What is there in HIP that produces this type of result; in other words can we identify the process of care responsible, as well as other factors?

But, it was just not possible for us to examine the process by which people received their prenatal care and the other circumstances in the process of medical care that might have influenced this result. I think the whole cause of reducing infant mortality would have been advanced if we had been able to get at the process, but

certainly the end result standing by itself in conjunction with the particular kind of setting in which it was carried out has been of an enormous importance in assessing the impact by prepaid group practice's impact on health.

So while I want to repeat that, of course, process is terribly important, there are on occasion very important practical considerations that make it extraordinarily difficult if not impossible to get at process.

The reverse is true too. An advance in understanding process with some implied benefits from process with no ability to get at the end result is also worthwhile.

DR. LEWIS: I think that's an excellent example. And the question has always occurred in my mind: If this sort of care system is related to these kinds of outcomes, then why have the, let's say, perinatal and infant mortality social gradients in the United Kingdom not been totally eradicated by the emergence of the national health system?

MR. SHAPIRO: Do you want to get into a discussion of that?

DR. LEWIS: No, sir.

MR. SHAPIRO: Look, in a system like HIP, we know that there are very important gradients by social class. I don't want to get into that issue because I think it opens up a new, highly complicated issue.

DR. KELMAN: Well, I would like to go a little bit further and reject if I can, just for the sake of a controversy, your emphasis on process evaluation. I'm not against it. Let me say that like everybody else, I'm for motherhood and all of that. No, these days you're not supposed to be for motherhood. I'm not opposed to process evaluation. However, I think, Chuck, that at least as I look over much of the evaluation literature, I'm struck by the fact that we have many more overall descriptions of program and process and visits than we have end-result evaluations. As I look over the process kinds of things — and this may be strictly personal, but I don't think it is — they raise no questions in my mind about program. However, when I look at outcome evaluations with or without process, they at least raise a question and would give me some pause about programs.

Now, I don't agree with the kind of response that you made to the first study — that if it's good, fine, and if it's no good, let's forget about it. I don't think that would be an appropriate response to outcome result where you may not know the process or channel.

I can give you an example of a study we're involved in where the outcome was negative. We had excellent descriptive material on the process. Nobody paid any attention to it because it was a negative finding. So that's one point.

The second point I would make has to do with the utility of evaluation. I think that obviously there are factors aside from the presence or absence of process information that would make the acceptance or rejection of an outcome result affect its acceptance. If there is no question to begin with about the program, if everybody is sure that is the only thing that can affect family planning, if this is the only alternative to patient neglect, then I submit this is not a question to be studied or to be evaluated and an evaluation is strictly eyewash.

It would seem to me that what is really wanted is documentation of the efficacy of what people's faith is in something, and I submit this is not an appropriate condition for evaluation of either process or outcome.

DR. LEWIS: Let me respond and say that when I was making comments about these papers, I hope you didn't lose the fact that I have been in and out of character in this discussion, one of which is a political animal concerned with getting things done and trying to keep peace and run an organization.

And maybe that's what all this is about – interorganizational conflict and the ways one deals with it using evaluation information.

It seems to me that the majority of people who want to evaluate something, Howard, come at it the way you just said: "We have a good thing. Wouldn't it be nice to show it?"

DR. KELMAN: "For you to prove it."

DR. LEWIS: "For you to prove it." I think sometimes the most fascinating opportunity for evaluation comes serendipitously that way. And you can say "We don't do that kind of evaluation," wherever you are locally enshrined, or you can say, "Okay, buddy, we'll have a go at it but let's be prepared to take the worst answer you are prepared to hear."

It seems to me that evaluation almost could not be separated – just a personal opinion – from the political and the ethical context in which it is performed and without the consequences to those who are involved in it. That may be a little more philosophical than I'm supposed to be.

DR. FOX: Two comments. I agree with what you say. However, I think that one must separate two very important issues. One is the bureaucratic and political pressures to prevent good evaluation.

Now, that is a very important product. In fact, I tend to believe the primary reason why good evaluation doesn't take place is more for that reason than the reason that technology doesn't exist.

The second aspect though, the relation between process and outcome studies, is itself a terribly impor-

tant separate question, and I wouldn't treat them as necessarily intertwined.

The other thing is that my own hard evaluation experience – I mean in terms of doing long-term studies – has been in mental health, which is a little different from a lot of other studies.

But we did a study where we were looking at rehabilitation of chronic VA patients with control in an experimental ward and reached a conclusion on most of our variables that the experimental ward was a little better and on one variable it was worse. And in a sense that was hard, you know. I mean the data was as good as you ever get in psychiatry, which is a little weak.

But then I think the creative part of this in some sense came in a bunch of us sitting around the table – by a "bunch" this included some patients too, incidentally – and trying to figure out, "Well, gee whiz, we thought we were going to get big differences." And yet we were only getting very small differences.

What was the process? And, furthermore, what were processes that didn't exist in either ward that might have been instituted that one might want to carry forward in further experimentation?

That's a very soft set of procedures. I think it's very important that this be done.

I agree with one of the comments that was made that there's a great tendency to get so embroiled in process because outcomes tend to be more difficult to measure, that you end up patting yourself on the back as the process looks pretty good.

DR. LEWIS: Let me restate. I have tried to say that I think both have to be done whenever possible – but there are circumstances in which only one or the other can be done and appropriate circumstances when maybe only one or the other should be done.

But I don't think there is such a thing as process or end results. And this gets to be an ideology, and it really breaks down between the denominator and numerator people in the world, those who are concerned about groups and don't give a damn about cases, and those who are only interested in what happens with the case. And these two subcultures have always existed.

MR. SHAPIRO: Present company excepted.

DR. LEWIS: I don't want to – I'll run up a flag in minute. But I think, quite honestly, this is one of the problems in trying to diffuse this issue of what are you going to do, because it really is related to personal orientations about how you see care.

DR. CARPENTER: Dr. Brown, there is a lull here. Do you want to give us that followup? Are they still trying to train occupational therapists in way-out Kansas?

DR. BROWN (Coordinator, Kansas Regional Medical Program): Well, it's a very complicated thing, and there has been a great deal of study of the situation. It's essentially where it was at that time.

Another similar thing, however, Chuck, having to do with changing conditions. It's the phenomenon we see with the home health care service. If the nurse makes rounds in a hospital with the physicians, she builds her clientele for the visiting nurse association very rapidly. If she is at headquarters and doesn't go into the hospital and make her own, she doesn't get referrals, which is the same — which has to do with awareness of physicians, you know, of whether everything is really lovely or where it isn't.

The same with the PT. Since they don't know and have personal experience, they really think everything must be all right and they really don't need it.

It's a complicated problem hooked up with our whole educational process in the state. So they haven't really made any progress.

DR. CARPENTER: It was effective evaluation I gather.

DR. LEWIS: We didn't spend some of Dr. Brown's money anyway.

DR. BROWN: They still want it.

DR. CARPENTER: Well, could we get some discussion around the question, "should end-result analysis be undertaken by every region funding a coronary care program?"

MRS. BLAXALL (Budget Examiner, Office of Management and Budget): I don't know if we want to specifically limit it to that. But a year ago we had a session with Pete Peterson and Karl Yordy and a couple of people — the Assistant Secretary for Planning and Evaluation — and Don Schon was there and a couple others from his firm.

And the whole point of the meeting was to try and get a handle on the kinds of evaluation criteria, indicators, whatever you wanted to call it, that the Bureau of the Budget might use not so much in evaluating in a hard sense but perhaps even describing the process of the activities of Regional Medical Programs in the budget appendix, for example.

We were using such things as the process indicators — how many participants in the training program, how many regions were operational, just, you know, just indicators, nothing that really explained anything related to Schon's systems transformation model, nothing that gave any flavor of Regional Medical Programs in the description.

It was the most elementary kind of analysis which we are all used to.

And the conclusion of the meeting, which was — When you think about it, a year ago we didn't really know as much then as we do now. The conclusion of the meeting was that we had to get a handle on ways to describe Regional Medical Programs, from my point of view, that would be able to focus in on what kinds of transformations were taking place in the health care system through Regional Medical Programs.

You know — big deal — that's the conclusion.

Well, we haven't really got any further than that, and yet I feel when I go looking at the budget submission when it comes in to me and I have to make some recommendations that I can't really justify Regional Medical Programs budget just on terms of additional trainees this year or whatever. That's not really what Regional Medical Programs is about any more.

And I don't know what kind of indicators to use. This is a tough question.

DR. CHARLES LEWIS: To drop back and say something here since I'm out of the RMP business, isn't this the whole problem since 1966, that the RMP was based on a promissory note which could never be delivered, which was really the elimination of heart disease, cancer, and stroke, and some of us had a strong feeling that besides providing "improving the care of the patients" it was really about regionalization, and the establishment of relationships, and the introduction of change within the system which occurs only under certain conditions.

It sure helps to have a little money. It helps to have some doctors who are hurting.

I think it's fascinating that we have focused most of our attention on university medical centers, which are about the last things in the world that are going to change because of the density of prestige and population.

I think if one really wants to see innovation in the medical care system today you go to the small towns any place in the country and you find nurse practitioners and physicians' assistants and mergers of hospitals and all sorts of interesting things that aren't making the *New York Times*.

But I suspect if one were going to invest a little RMP cash, one could very easily facilitate regionalization outside of those sorts of procrustean things that have probably already died but the message just hasn't got to the brain yet.

MRS. BLAXALL: That's right. I agree with your statement.

DR. LEWIS: A lot of people don't.

MRS. BLAXALL: But it doesn't help me in the question I have. This is a tough question we're still working on. For example, does that mean instead of using the old indicators that we should focus in on anecdotal elements?

DR. LEWIS: No, there are end results that can be measured I would assume.

If you would like to talk about the availability of care for populations and the provision of care to populations that don't have any care, as a byproduct of RMP, I think that can be measured — providing that's what your objectives were.

But there have never been any objectives except to "improve the quality of care for patients with heart disease, cancer, and stroke" — starting at where the care was probably the best.

MRS. BLAXALL: Does this get back to the question then that I hear when I go around and talk to some of the regions, "Who's making the objectives for RMP?" Washington or our local RAG? Is that the kind of question that you're getting towards?

Because if there aren't any concrete objectives at the national level, which is what I suppose I have to worry about, then —

DR. LEWIS: I think RMP when it emerged in 1966, for some of us that really got seduced into the planning process without knowing what is going on and found ourselves operational before we really knew what was going on, we had been at that time fascinated by the fact that this was a program in search of objectives, that there was an enormous amount of money to be spent for doing something, but no one ever defined from hierarchical quarters up there what was expected of regions, and regions grew depending upon, essentially, the philosophy of the coordinator or the parent institution.

And at that time I think many of us felt the taxonomy of RMP. There were hardware-oriented regions and software-oriented. There were disease-oriented and there were people-oriented. They were centralized and there were decentralized. They were clearly determinable by the nature of the people involved in the original programs.

I do not know whether it has changed or not. This was the equivalent of the identity crisis which overwhelms the teachers of preventive medicine annually.

DR. CARPENTER: You know, it's interesting that now we are stuck with really so many objectives that there are people who say we don't have any. Each individual region has a large number of objectives, some of which are immeasurable, some of which, though, are

measurable. The diversity — the major strength of the law's permissiveness toward local innovation — makes for such difficulty of expression that it now becomes the bane of the evaluator's existence. Having no national decision that a priority, for example, for coronary care is acceptable, he has less clear evidence as to whether his Region has placed significant priority on such care.

DR. LEWIS: I think if your programs had written real objectives and not statements of vague goals, they might have been evaluable. And it's like teaching, you know. If you just tell them what you want, which we usually do, it's a mess. Writing educational behavioral change objectives is a very difficult job.

DR. HASTINGS: It occurs to me maybe we have got a new definition of what RMP is really about. If we make the assumption that RMP's real business is social change, if we are supposed to be changing things, then perhaps we should shift our statement of what our objectives are from disease-related, medically-related criteria as listed in each of these articles, as enumerated in each of these articles that you just discussed, and frankly say that we're in the business — that we're in a political business, an organizational business instead of being in a task-related business, that we're in the business of changing a system.

And if we define ourselves that way, then it's possible to state objectives that one can measure, different kinds of objectives that people have tried to measure.

But if that's what we are about, maybe that's what we should be doing.

DR. LEWIS: It would have been nice if the original law hadn't said in it as long as it doesn't interfere with current patterns of practice.

DR. CARPENTER: But interference and change aren't the same.

DR. HENDERSON: That's right.

DR. STONEMAN: I think there is a real gap that has developed in this conference. I think it's been there all the time. I think Dr. Lewis alluded to it. It concerns me. I'm sure it concerns many other program coordinators.

I think a lot of us were seduced into RMP by the bright hope of local initiative and local decision making and system building within the context of the law as it was written, with perhaps a few liberties with the interference clause.

But we did develop regional advisory groups. We did develop systems. We did spend a couple of years teaching them what the law says and what it's all about. And we did do this on the thesis that unless we put a system together that could work together we were never going to be able to move the system in any effective way. We

have begun to make some progress toward doing that, but we aren't there yet. I don't know all about all the other regions. I know we're not there yet.

The law is being renewed. It's written by Congress. It's still virtually the same language except for some kidney wording and a few other minor changes. And yet the Bureau of the Budget and others in Washington are coming through with a clarion call that we're going to be judged on whether we're agents of social change and whether we can materially, with the dollars we have, affect the health status of the nation very soon.

Now, we spent all day finding out nobody can tell us how to measure health status to begin with. So we can't evaluate that pursuit except in individual program activities, and that's out. We're not supposed to measure activities as much as we do broad program. The people back home still think we're working under Public Law 89-239 and renewals.

Now, it seems to me that there is an obvious question here that I hope will be addressed before the meeting is over. I don't think we can do what we have been asked to do until we do what we set out to do — put a system together. And I don't think we can do it by fiat within the next four months or within the next 12 months, probably not in less than several years.

And this comes back to the question the young lady asked about — what do we put down to justify your existence? I don't think we're going to with \$94 million this year produce enough product in additional health care delivered to amount to a minuscule fragment of the total systems production.

Maybe we're going to produce a process that can put us in a position to do something about that, but I can't give you much more justification than that.

DR. HENDERSON: I want to just try to remove one misconception I think I heard.

I would not say we cannot get measurements of health status. I say we can. I tried to say that it is a difficult task and it takes experts in many fields to apply their knowledge and do it efficiently.

I think you have seen that. There are experts in several kinds of measurements here today. We have all tried to say that it takes a lot of effort, a lot of skill, and a lot of skilled personnel focusing on doing the specific kinds of evaluation. I do not think the RMPs have had people with the right kinds of expertise in their programs to start off with—for good reason. The majority have been planners and people who had to get programs implemented and were well versed and became well versed in these aspects.

This may be just a time lapse. But I do not think that you should say or anybody should say that we cannot do it. Given enough money and the proper input it can be done. But it cannot be done except by collaboration between many kinds of experts with background and training in the sciences needed for the purpose.

DR. STONEMAN: I know, but given the fact that each region is doing its own thing, if you will, even give the kinds of that you describe—and I listened very closely this morning, very interestedly—at \$200,000 for the first year how long with that kind of a data base would the regional medical program have to go with operational activities directed toward the soft spots and gaps that you identify and develop before you can come back with a continual status evaluation that will answer the question that she asked—for one region?

DR. HENDERSON: I can in part answer your question. I cannot give you a time limit. But I can tell you a problem about the whole program that I think extends this time. Because of the insecurity of funding, from year to year, our unit has no full-time professional person. No one with enough epidemiological and statistical experience to organize this kind of center can at that stage in their career afford to go full-time on a program without surety of continuity and funding. So if the program had a more stable base, it could be done in much shorter time because you would get people working at the job full time. The very nature of the program is extending the length of time it takes to do evaluation.

DR. LEWIS: I think just to reintroduce Buck Rikli's question as we have come full circle, it's whether or not the kind of data that we are talking about will influence planning and operation.

MR. SHAPIRO: I don't see how you can answer that question—in a kind of global way any more than I could possibly grapple with the global way of stating the issue of changing medical care systems. You can think in terms of a change of medical care systems involving a total approach. This is a \$65 billion-a-year industry. And anybody who thinks that RMP is going to change medical care systems in a very fundamental and decisive way just doesn't know what's going on. It's unthinkable.

But you could define medical care systems in clusters, in smaller units, in a dimension which you can begin to grapple with.

I hate to come back to our own experience and our own aspirations, but the program that I was describing this morning in coronary care is directing itself at a categorical disease, but to be effective, the way we view effectiveness, it means a change in a system. Hopefully, through a demonstration of the kind we are projecting

there can be an influence on a much broader segment of the community in developing approaches to a specific disease.

So I think that there is a danger of stating issues in such broad terms that it becomes absolutely impossible to cope with them.

DR. KELMAN: Well, again I think if you came away from the discussion all day with the idea that we can't measure health status, then really we failed.

I don't think we here could allow you to slide out from taking a hard look at RMP easily by saying, "We can't measure health status so therefore RMP can't be evaluated in those terms." It's not appropriate.

The discussion we have heard thus far initiated by the young lady in the back is very similar to many discussions I have been in after a program has been launched and they say, "Well, we'd better get an evaluator in here to tell us what we're doing because we don't really know." And I think that's pretty sad after all this time. I cannot for the life of me understand how we could get into the sorry state of spending all of these millions of dollars setting up all of these regional offices and then come around and say, "Well, I really don't know how to judge whether one or another region or one or another unit should get more or less funds for what it wants to do."

This is an extremely dangerous kind of situation, tying it back into some of Chuck Lewis' comments, for an evaluator to operate in, because he or she can't possibly win in such a situation. In other words, you're putting the evaluator in the position of defining the objectives of the program. Do you really want that? I don't think you do.

DR. CARPENTER: Bill, you stimulated a lot of this.

DR. STONEMAN: Yes, I'd like to respond.

I didn't mean to sound like an evaluation nihilist. The thing that bothers me is that we have had for some time now some very broad and general aims for RMP outlined which are extremely vague, and if I overstated their vagueness and the unlikelihood of their immediate accomplishment, I apologize. But I apparently made the point.

If we are going to go to program evaluation at the regional level instead of concerning ourselves with individual project activities, then I would submit that most of the evaluative techniques that were described this morning are more appropriate to project evaluation, if I can use that term, than they are to program evaluation.

Then it's necessary for us to hold our own feet to the fire in terms of setting some precise program objectives

before we can begin to decide how we are going to evaluate them.

And I must confess it's still not clear to me what evaluative methods we are going to use for that. I have got some strategic concepts of why I'm doing many of the things I'm doing, but they are steps along the way to what has been discussed in terms of more profound changes in the system than the reorganization of a given subsystem within our coronary care process. I hope that clarifies what I said to some extent.

DR. BROWN: It gets back though to this business about process and end results. If you're going to try to define how many people's lives you saved or so on, that's going to take a very long time and may not be possible and probably isn't even important. But the process is important, the process by which subregionalization or regionalization occurs.

Now, that may be hard for people in the Bureau of the Budget to measure, but that's their problem as well as ours, because that is where maybe the \$96 million can have some influence on what is happening in terms of the whole.

Now, that's about as global as I could make it, and within that there are 55 sub-sets and probably 25 approaches within those sub-sets of 55 regions, and then within that there are a lot of other smaller things that Dr. Shapiro refers to which are terribly important, but I don't know how you measure those in terms of lives you save.

DR. KELMAN: Could I be antagonistic and ask why it's important to have all these subregional clusters and paraphernalia?

DR. BROWN: It's a mechanism because someone feels that there might be a better way or a more economic way or something to deliver health services.

DR. KELMAN: I'm asking an outcome question.

DR. BROWN: I was struck with this business here of the neighborhood health clinic where the analysis of the report says that 95 percent of the patients get followup contrasted with only — what? — 10 percent or 20 percent. Therefore the neighborhood health center is a good thing?

DR. KELMAN: I don't know if it's good.

DR. BROWN: Well, nobody knows, but that's one of the objectives it seems to me we're hearing, one of the goals of the regional medical program. Access. Isn't that access? It doesn't make any difference whether the outcome was better for the patient. Nobody measured it. But if we could guarantee access, that's politically important right now.



Now, I'm not saying that's good or bad. I'm just saying if you take stability of data you could say, well, here are X number of people who did not get followup. Now they get followup. Therefore, you've improved the system.

Maybe all you have done is added a component to it that costs you money.

DR. CARPENTER: I suppose the fear is, Bob, that although that is politically important today, it doesn't sound as though it's going to stay politically important, whereas whether or not there is increased access to improved health care may have a little longer staying power as an argument.

MR. SHAPIRO: Let me give one example briefly. Then I've got to leave. And I'm going to oversimplify the situation.

During World War II, there was an EMIC program — emergency maternal and infant care. Nobody thought in terms of an evaluation of that program. There were millions of women who were delivered through this program. After the war that program was abandoned. There was no supporting evidence that could be used to sustain a program that roughly corresponded to the EMIC program. There are a lot of people who are convinced that some form of EMIC program would have been maintained after the war and hopefully would have resulted in further reductions in infant mortality in this country instead of the long sustained period of small decreases if those responsible had taken the trouble to think through the importance of evaluation.

There is currently a program in maternal and infant care, and there's a huge amount of money being poured into that program. I don't believe that that program will continue in the long run unless it can prove itself in one way or another.

I think in the RMP there are very similar types of situations. I don't care how carefully you regionalize an ambulance service to respond to coronary care emergency situations. You may have a beautifully operating program. But unless somebody can establish whether or not that program is really accomplishing something in terms of outcome, that program is going to be chopped. That's the rationale behind outcome.

DR. LOGSDON: Could I just comment about another program that had a similar type of outcome in the migrant health bill that was passed and which operated on a budget much less than this, about one-fifth of the amount, and was passed primarily because of Steinbeck, his writing, and some special interest groups that were able to get enough support in the Congress. And this program provided health and environmental

services to migrants all over the country. But because of the lack of solid evaluation information and because of the lack of grass-roots support, this program is in real jeopardy right now of being lost in the shuffle of another bill that was passed. And if I was any kind of prophet I would say that the same thing could well happen to the RMP.

DR. CARPENTER: Dr. Fox?

DR. FOX: I think Martha and I would like to respond to some of the comments.

For those who don't already know, this is Martha Blaxall who is a budget examiner in the Health Branch in the Office of Management and Budget. She also helps me write speeches for places like here and ropes me into interesting meetings.

I think a couple of points have been raised. The problem of insecurity of funds, for example, has some validity. The issue of lack of goals may or may not have validity. I think that can be carried too far.

I wonder, for example, whether you were at lunch and listened to Dr. Margulies' speech. He enumerated certain things that were as clear as they are going to be enumerated, and if you people don't understand what they are, then I don't know what else can be done.

You also heard in the morning that the concept of themes versus specific objectives was talked about by Don Schon, and I haven't heard anybody dispute that as a concept. You know. The messages that you're going to get will consist of themes. You'd scream if you were given specific objectives in terms of numbers of this and that type of unit that you must engage in.

We have heard that you can't measure health status. Well, you know, I made a big point of this yesterday in my talk, and presumably you heard that. Not that you can't measure health status, but that you won't get a single measurement of the impact of RMP tied up in one cost-benefit measure. We're aware of that.

On the other hand, there are things that can be done. I sure learned a heck of a lot today. It (the panel) has some of the best information of what the state of the art in measuring health status really is.

Let me tell you some of the things that I think one can expect. I think one can expect movement in directions. What those directions, the precise directions, ought to be, that's up to you people again. You know the themes. What are some of the system changes? Is duplication in facilities being eliminated or new duplication being prevented?

We're on the verge of entering into the kidney field for big. Are we going to have the same fiascos there we

have had in open heart surgery? If we do, then maybe the program should be questioned.

We have also heard examples here, and I have heard them, anecdotally, of important situations where duplication of facilities has been prevented.

We know manpower is important. You know. Is RMP doing something to rationalize the introduction of new manpower in the project areas?

These are meant to be seed money projects. Are they engaging in projects that are real projects that are absorbed into the regular system after, say, a two- to five-year period?

One can look at core staff and ask whether they are developing a regional strategy that intuitively makes sense or is it a case of just responding to individual requests and interest groups that come in?

We have heard statements that the evaluator can't set objectives. Well, this is true in a purist sense. But if the evaluator can't help the decisionmaker set objectives, can't start to ask questions that assist the decisionmaker in setting objectives, then the evaluator ought to be fired — and I really mean that — because that may be the most important thing that he can fulfill.

And I know to some extent the regions have to come forth and say, "Look, within these themes these are the

good things that we think we can do, and these are our objectives. This is what we think is reasonable to measure us by. Here are some measures that might be tempting from your point of view but we think they are unrealistic because —."

And I think the regions have to come forth with honest information, not with snow jobs.

Now, in a sense, things are bad. There's uncertainty. But the uncertainty isn't, I contend, anywhere near as bad as what your statements make us believe.

DR. JESSE B. ARONSON: I'd like to ask the question as to why in all of these discussions of measurements we haven't brought in or I have heard really nothing about the measurement of the cost factor.

We know that we are far from getting cost-benefit studies. We certainly can get cost-effectiveness of process. And if we are going to start measuring process without measuring costs, I don't think we're measuring process in any realistic sense, in any case that will in any political sense certainly be realistic.

And I think we ought to put more of our thinking, and we ought to have examples of studies, where the cost-effectiveness of process becomes an essential element in our whole measurement system.

## WORKSHOP ON PROGRAM EVALUATION

### Participants

Harold W. Keairnes, M.D. – Moderator  
Coordinator for Evaluation  
Tri-State Regional Medical Program

Avery M. Colt  
Field Associate for Eastern Massachusetts  
Tri-State Regional Medical Program

Claire G. Farrissey  
Special Projects Coordinator  
Tri-State Regional Medical Program

Osler L. Peterson, M.D.  
Associate Director for Data Collection  
Tri-State Regional Medical Program and  
Acting Chairman  
Department of Preventive Medicine  
Harvard Medical School

Dean J. Siebert, M.D.  
Associate Coordinator for Dartmouth  
Medical School  
Tri-State Regional Medical Program

Ruth B. Mott  
Research Associate for Data Collection  
Tri-State Regional Medical Program

James J. Dunlop, Ph.D.  
Arthur D. Little, Inc.

Robert K. Ausman, M.D.  
Deputy Director  
Florida Regional Medical Program

Jack E. Thomson, Ed.D.  
Coordinator of Evaluation  
California Regional Medical Program

Robert Beckman, Ph.D.  
Director of Research  
Nassau-Suffolk Regional Medical Program

William R. Thompson  
Deputy Director  
Washington/Alaska Regional Medical  
Program

Paul E. White, Ph.D.  
Division of Behavioral Sciences  
Johns Hopkins School of Hygiene and  
Public Health

Jeannette Forsyth, Ph.D.  
Project Administrator, Information Support  
System  
Tri-State Regional Medical Program

---

### Approaches to Program Evaluation

H. W. KEAIRNES, M.D.

Evaluation is assuming a larger role in the planning and management of Regional Medical Programs. The new procedures for anniversary review program applications and in-depth site visits indicate that increased local autonomy in management of activities and funds is contingent upon a clear understanding by Washington of yesterday's achievements by the program. Under these conditions, past performance is equally as important as future plans. Evaluation, whether done formally or informally — if done at all — helps build the bridge from the past to the future.

Recently I tape-recorded a brief interview with Mr. Robert Lawton, Deputy Director of Tri-State Regional

Medical Programs. After talking about the impact of the anniversary review guideline on local programs, I asked, "How do you expect the evaluation activities to contribute to the development of these program applications?" This is the dialogue that followed:

MR. LAWTON: The program application and program itself has to demonstrate that it can manage the process in its own region of good health service problem solving. The (evaluation) technique for doing that must not only exist in the region but must be visible in the application. The region has to know how to apply and use the technique and how to use the results of the evaluation technique. I think it's a good circle involvement. You have to develop and put down a technique that helps you do your job — better.

DR. KEAIRNES: You talk as if evaluation has something to do with planning.

MR. LAWTON: I find them hard to separate. I think that the credibility factor is extremely important here. I think if

you are going to do good things for patients and good things for patient care through rationalization, then you have to demonstrate that what you did yesterday had some merit and improved patient care – so that evaluation is an on-going thing. Today's planning and tomorrow's results are pretty dependent on yesterday's evaluation.

In the game of improving patient care, that's another way of saying that evaluation is part of the process of winning. Planning and action are, or should be, based on experience. Evaluation involves the systematic description of these experiences and the associated achievements. If done well, evaluation can supplement the gut-level feelings that play such a prominent role in most decisionmaking about the future. Unfortunately decisionmakers have functioned so long without systematic evaluation that many feel that they can win without it, or, at least, by paying no more homage to it.

The model for winning through the use of evaluation has been established by that multi-million dollar industry – professional football. Each week each team records the process of their winning or losing in the game movie. The coaches and, to a lesser extent, all members of the team spend many hours reviewing the game movie. They evaluate every plan and the performance of every member of the team. Those plays that worked well will be used again. For any play that didn't work, decisions will be made about the performance of each player and the appropriateness of the play. On this basis plans are made for practice and for the next game. And then they practice. There is little mercy for teams that continue to make the same mistakes in decision making and performance that were obvious in the game movies. Of course they have to take into account the limitations of their personnel and their system and the new challenges presented by the next opponent. In next week's game, if they have successfully evaluated, corrected and planned, they will win. And they may even win over a team that has superior personnel and resources.

The task of a broad-base social change organization such as Regional Medical Programs appears more complex than that of a professional football team, but only superficially. They both have the same over-all objective – winning. RMP's goal line, however, is less well defined. There are many more ways of scoring points. The process of moving down the field involves many more players. The opportunity for fumbling is much greater. The rules and the officials are much more difficult to identify. The fans are often not interested in paying to see the team win. And there are no time outs during the game or between games.

But none of these differences negate the value of the game movie and the process of planning for tomorrow on the basis of what happened yesterday. What follows is a description of the concepts and methodology for taking an RMP game movie that will allow a clear assessment of the performance of the teams involved in winning or losing the game of rationalizing and improving the process of medical care delivery.

### Concepts of Information Support

Evaluators in Regional Medical Programs play the role of cameramen, not coaches or players. In their role, they must keep the camera focused on the crucial activities on the playing field if the coaches are to have useful game movies. Evaluators have not been hired as judges. Only those persons whose decisions influence the fate of an organization can really be considered as judges. Evaluators are hired to provide information to decisionmakers so that their judgments are not made on incomplete, inaccurate or biased information. In this sense, they are concerned much more with INFORMATION SUPPORT than with judgmental evaluation.

This concept of information support makes sense only when the decisionmakers utilize the information. If no one but the cameraman sees the game movie, then the plans for next week's game will be based on the rather undetailed and unsystematic recollections of the coaches and players. Similarly, taking two weeks to develop the film destroys its usefulness. If the film is available and utilized, then it must be of such quality and content that the coaches and players find it useful. If they feel that it is useful, they will utilize the information in their planning and decisionmaking and they will request that the service be continued. In Regional Medical Programs information support services can be justified only when there is utilization of the information and requests for additional information by the decisionmakers.

### Decisionmakers in RMP

Who are these decisionmakers in Regional Medical Programs that correspond to the coaches of the professional football teams? One of the important differences between the two games is the larger number of players and decisionmakers involved in RMP activities. RMP decisionmakers fall into several important groups:

1. Coordinators or directors – the senior executives who are responsible for the implementation of the planning and operational activities of the program.

2. Planners — the committee members and core staff personnel who determine the direction — objectives — of the program and the activities that will move the program and the region in that direction.

3. Project directors and officers — the core staff and project personnel who manage the process of project development and operation.

4. Grantors — the members of local and national advisory groups and the staffs of granting agencies whose decisions determine which activities and programs become funded.

5. Consumers — both professional and lay persons whose support determines the success or failure of most broad-based social change programs.

If the decisions of all these people about how the game should be played are correct, Regional Medical Programs stands a good chance of winning the battle for rationalizing the medical care system through voluntary mechanisms. The thesis of this paper is that meaningful information based on past experiences and provided in a useful manner will improve all crucial decisionmaking. Of course, evaluative information becomes only part of the decisionmaking process and, by itself, cannot overcome problems in communication, resources or constraints that also influence the decisionmaking process and its results.

#### Work of the Evaluator

Meaningful information forms the context for the work of the evaluator. He must understand how the game of Regional Medical Programs is played, who the players are, what direction the team is heading, and what the coaches want to see before he knows where to focus the camera. For example, focusing on the wide flanker while he sits on the bench during a defensive play may be the same as focusing on the evaluation of a project when the decisionmakers really need to understand how all the components of the program are working together to further the task of winning the game. Narrowly focused observations have limited value to understanding the total game process. Indeed, focusing on the wrong area may prevent the coaches from observing the process of scoring. Meaningful information that is useful for planning the next game depends on a description of the entire field including the play of all members of both teams and the success of both teams in crossing the goal line. It includes all the projects — both operational and planning — as well as all non-project activities of staff and committees. It also includes everything that the opposing team — the forces for the status quo — is doing

to resist the activities directed towards rationalizing the health care system.

The evaluator in focusing his information support services must first know the location of the goal line and the rules of the game. Then, if he understands who the key players are and how they participate in the game, he stands a reasonable chance of providing a meaningful service; that is, he will make the appropriate observations on the appropriate players during the entire game. Being guided by the decisionmakers in this process of focusing his observations improves his chances of making a game movie that the decisionmakers will find useful. If the decisionmakers will not provide the assistance or if their assistance is not sought, making the game movie becomes an irrelevant exercise. Fortunately for both decisionmakers and evaluators there are some general guidelines to follow.

#### *Location of the Goal Line— Problems and Objectives*

Each problem in the medical care system defines a different goal line. Setting objectives is the process of specifying which goal lines should be crossed. Planning specifies the activities which if carried out should lead to crossing the goal lines.

Analysis of published studies, surveys, reports, and applications gives the first level view of the problems of a health care system in a geographic area. Interviews with all classes of decisionmakers and other key persons are required to understand the relation between described and perceived problems. The degree of consensus or agreement on high priority problems gives some indication of the potential cohesiveness of the medical care system for problem solving.

Obviously the Regional Medical Programs cannot cross all possible goal lines or solve all the problems of the medical care system simultaneously. Objectives and priorities help direct the team towards those problems that most need to be solved or are most amenable to solution. Published objectives may or may not be the true operational objectives. Discrepancies arise when operational objectives are perceived as being not socially acceptable or when there is lack of consensus among decisionmakers about desired objectives. Such discrepancies make it more difficult to mobilize resources to accomplish the objectives.

Public objectives can be determined from documents. Operational objectives can be determined by direct interviews with, and by secondary interviews about, key

decisionmakers. Following these processes allows description of the nature of the objectives and of discrepancies between published and operational objectives.

Data Source: documents  
direct interviews  
secondary interviews.

Analysis: nature of problems and objectives  
consensus on problems  
discrepancies between published and operational objectives

*Rules and Playing Conditions of the Game -  
Resources and Constraints in the  
Medical Care system*

The Regional Medical Program's task lies in a setting created by existing institutions and their services, key persons both lay and professional, existing legislation and regulations, and financial resources both fixed and flexible. General socio-economic conditions, population distribution, transportation patterns, communication systems and educational resources are also part of the milieu. Describing these facts makes apparent the playing conditions of the game.

The constraints in the system are created by legal forces, institutional relationships and history. Legislation, regulations and guidelines may be found in published documents, but their impact and their ability to respond to new problems can be learned only from administrators who have had to work within and around them. Institutional relationships can be characterized by patterns of 1) institutional exchange of board members, staff, clients, and communications, 2) institutional domain for clients and resources, 3) domain conflict both actual and perceived, and 4) participation in joint planning activities. Historically the fate of previous change efforts and the general responsiveness of the system to new problems and new resources suggests the rules which influence the success of all future change efforts. These are the rules of the game.

This information, although crucial to evaluation, is the keystone of planning. It makes clear the condition of the playing field and the rules of the game. The evaluator should watch for ignorance or misperceptions of the conditions and rules by persons playing for the local Regional Medical Program.

Data Source: documents  
interview

Analysis: identification of key persons, institutions, and resources

distribution of key persons in relation to problem  
institutional relationships as characterized by  
exchange, domain, domain conflict, and joint  
planning activities  
history of previous change efforts

*Record of Team Performance -  
Results of Previous Resource Allocations*

Local Regional Medical Programs have up to thirty years of experience as operational programs. Unless a game movie exists, this description of team performance will have to be primarily performance statistics that are generally available, such as the number and types of plays, number of yards gained, and the number of fumbles, downs, penalties, and scores. Recollections of team players give some clues to the process, but they are subject to bias. Nevertheless this information is part of the planning for tomorrow.

The players in the Regional Medical Program game can be considered to be staff, committee and advisory group members, and all other persons in the medical care system. It is important to identify through interviews the members of the team, their skills and attitudes, their assignments in the change process, and their performance record. Their skills relate to their training, their position in their institutions, their concern, and their commitment. Their assignment as well as their performance vary with the activities.

Identifying all the activities or plays that are carried out is perhaps the most difficult task facing the evaluator. There are so many simultaneous activities with vague starting points, a paucity of progress reports, confusion as to who is participating, and a lack of agreement on when the play is completed and, therefore, when it is appropriate to measure progress. The easy way out is to restrict one's concern to funded operation projects. That is appropriate if operational projects account for 90 percent of core staff and project staff time and budget. Unfortunately that is rarely the case. The whole spectrum of activities that must be identified include operational projects, planning projects, committee activities, central administration services including communication, research, data collection, and evaluation, conferences, developmental negotiations.

Once an activity has been identified the players, the assignments, their performance, and the effectiveness of the activity should, if possible, be identified. The performance of individuals relates to how well they carried out their assignments. The effectiveness of the activity asks not only how many yards were gained — a short term estimate of progress usually based on a

achievement of project objectives — but also whether the play or activity resulted in a first down or the crossing of goal line — a long term description of the resolution of any of the specific problems on which the program objectives focused.

Resource allocation is akin to selection of plays and the assignment of players. Effectiveness in achieving program objectives is obviously related to having the right players and the right play. Resources include personnel time and funds which are directly accessible to the local regional medical program *plus* all available personnel time and dollar resources in the region that could potentially be mobilized towards achievement of program objectives.

Retrospectively many details of resource allocations and player performance are lost. Effectiveness both in moving down the field and in scoring, however, is apparent because *significant* gains are usually obvious. Details become more important when progress has not been obvious. In this circumstance winning in the future obviously depends on developing a more effective allocation of resources because new players and new dollars are not usually available.

Data Source: documents  
interviews

Analysis: descriptions of persons involved in  
Regional Medical Program activities  
identification of activities  
identification of effectiveness  
description of resource allocation

Once the location of the goal line, the rules and playing conditions of the game, and the record of team performance have been developed by the evaluator, he has two obvious tasks: first, to report this information to his organization, and secondly, to set up an ongoing mechanism for recording and reporting evaluative information. Both of these processes depend upon the specific conditions and needs of his program. He must remember that the information should be considered confidential and that the coordinator of the program should have complete control over the use of his analyses and reports. The evaluator should work closely enough with the coordinator so that the results are made available in a concise, meaningful, useful form, but with enough accompanying detail for use by other decisionmakers if desired by the coordinator. The evaluator in developing the information should recognize that the reports should be constructive and not destructive. The reports should allow an opportunity for development of winning

patterns and should not result in the players becoming so defensive that they will not participate.

The ongoing mechanism for recording and reporting evaluative information depends on the philosophy of the coordinator, the evaluator, and the core staff. But participation in the evaluative process will probably result in more effective utilization of the information. In his assigned role, the evaluator should be responsible for surveillance of documents, especially minutes of meetings, application for planning and operational projects and reports of projects and studies in order to maintain some general structure for all evaluative observations. He may supplement his observations by interviews with persons involved in the various activities, by participant observations in committee meetings, planning activities and consultations, and systematic reports from core staff and project directors. Involvement of many of the staff in reporting participant observations and their analysis provides an opportunity to train them in evaluation concepts and the use of evaluation information. Although discrete segments of the program may seem to require specialized research or project evaluation such activities are not a substitute for ongoing program evaluation. Program evaluation requires the identification of all activities, all the players on the teams or some other major category. The performance of the players in each of the activities, the success of the activities in making progress down the field and drawing first downs and the effectiveness of the whole mix of plays and players and short-term achievements in moving the program across the goal line in scoring gains against the problems that exist in the region.

In this context evaluation itself is one of the major activities of a program. Effectiveness of evaluation activities can be judged from its influence on the decisionmaking and the planning processes. Indeed, if evaluation cannot be demonstrated to contribute to winning the game it cannot be justified as an important activity of the Regional Medical Program. Effectiveness and relevance must guide the entire process of observation, analysis and reporting of evaluative information — that is, effectiveness and relevance to the decisionmaking of all classes of decisionmakers from consumers to Congressmen.

The importance of involvement was summarized quite well by Mr. Lawton in our interview when I asked him if he had an opinion about what proportion of RMP effort should be put into evaluation. Let me close with his response:

“No, I don't think I have.

I see it working in this way, an evaluation component, such as you and your associates, but in addition I think our

geographic and medical school coordinators have to apply evaluation in their work and, if they need it, get educated about evaluation technique. Our program implementors, whether they are in the RMP organization or in outside health institutions have to have an appreciation of evaluation also, because I find it impossible to separate evaluation from the objectives of programs or projects.

I would hope that there is a little bit of evaluation in everybody – including me.”

### Program Evaluation Workshop – A Case Study

HAROLD W. KEAIRNES, M.D.

This paper describes the educational content that was developed by the faculty, the educational methods that were used in a workshop session sponsored by the Regional Medical Program services, observations on the educational process, and implications of the workshop experience for training and evaluation and development of program evaluation methods.

#### EDUCATIONAL CONTENT

The evaluation of Regional Medical Programs at the local level is an open-ended story. Formal guidelines have not been established because they are evolving. Their evolution is based on a process that takes into account the multiple uses of evaluation, extreme variability in judgmental criteria and significant differences in the philosophy and approach of the 55 regions. The open-ended nature of the legislation - PL 89-239 - requires this evolutionary approach.

There are three basic uses for evaluation: 1) justification 2) control and 3) learning or planning.

When used for justification, evaluation must deal with judgmental criteria that have been established and are regularly utilized by various groups, e.g. National Advisory Council, RMPS, local and national legislators and political figures, organized providers, and consumer groups. Usually these criteria are unwritten and often unarticulated, but they do exist. In general these groups make judgments based on their private interpretation of the intent of the legislation, and these interpretations form the criteria for their judgments. Their interpretations may vary considerably from that of core staff and RAGs of local programs and often may not allow for problems of feasibility and practicality in carrying out the legislative mandate.

When used for control purposes, evaluation helps administrators and executive committees such as RAGs to be aware of progress in implementation of *all* the various

activities of the program and the effects, both observed and anticipated, of these activities. Program evaluation emphasizes a comprehensive approach that accounts for all activities supported by the resources of the program, not just discrete operational projects. Ideally program evaluation provides a mechanism for monitoring the progress and the effectiveness of all activities in furthering the purposes of the RMP. These purposes include the philosophy, goals, and objectives of the program as a whole, as well as the more limited objectives of each discrete project or activity. Program evaluation information can then lead to better understanding of the most effective way to allocate financial and personnel resources within the program and to maintain quality control of all supported activities.

When used for learning or planning, program evaluation attempts to assist the staff and advisory committees through a continuous feedback process to improve the quality of their decisions and actions. This is the traditional planning – action – evaluation cycle. Maximum benefit occurs when all the actors participate at least partially in the entire process.

#### *Role of Evaluator in Program Evaluation*

The evaluator concerned with program evaluation is the keystone for the following processes:

1. Identification of the uses for program evaluation.
2. Identification of those individuals and groups who will use program evaluation information.
3. Identification of appropriate evaluation criteria as described by each user for each use.
4. Identifying his own role in obtaining information that allows judgment on the established criteria for each use.
5. Carrying out his role.

In the process he must avoid substituting his criteria for those of the users of program evaluation information and must be sure that the information is accurate, reliable, and relevant to the criteria, the uses and the users.

The evaluator's role in program evaluation is quite different from that in project evaluation. Project evaluation is concerned with discrete activities based on objectives that are or can be well defined and agreed to by the project staff and the funding agency. Here the evaluator's role is to apply systems' concepts tempered by economic, educational, epidemiologic medical care or other technical considerations to the development of an economically feasible evaluation methodology – and to the supervision or performance of the evaluation plan.



Program evaluation is concerned with the conglomerate activities that have poorly defined objectives, that often cannot be clarified to the satisfaction of all the users of evaluation information. Here the evaluator must be extremely flexible and understanding in order to deal with the complexities of the task. Rigid application of traditional evaluation approaches, such as may be appropriate for discrete projects, becomes increasingly irrelevant as programs become larger or broader in their scope. Precise evaluation of one component of the program usually gives little insight into the total program and usually provides little assistance to those who must administer or justify the financing of such programs.

### *Process of Program Evaluation*

For the purposes of this workshop, program evaluation was defined as a process. This process definition took into account the very primitive state of the art of program evaluation. Although projects are underway to develop the methodology of evaluation of broad range social change organizations such as Regional Medical Programs, there are no generally agreed upon and tested methodologies at present.

This process of program evaluation follows the following steps:

1. The evaluator shall develop a thorough understanding of the philosophy, history, strategy, and activities of the program. In this step, he may infer from his observations what the objectives of the program are and how these observed objectives relate to published or reported goals and objectives. Such inferences should be verified when possible.
2. The evaluator shall determine who wants or should want program evaluation information. From each of these individuals or groups, he shall obtain the criteria by which they make judgments and their intended uses of the information: justification, control, or learning.
3. Based on these objectives, criteria and uses, the evaluator shall develop a program evaluation methodology. This methodology should be comprehensive, practical, and efficient. Unless he has outside financial support for evaluation research, the costs of carrying out the evaluation should probably be less than 10 percent of total program funds. The scientific disciplines incorporated into the methodology should reflect the needs of the users of the information rather than the particular scientific discipline of the evaluator. The evaluation should take into account the

temporal flow and sequence of activities — that is, effects on process and organization can be observed in 1-3 years, but significant effects of transforming the medical care system on the process or end results of patient care may take 3-10 years.

Stated in a different manner this process calls for:

1. Identification of all activities of the program — past, present and anticipated.
2. Identification of all possible effects of these activities.
3. Developing methods for describing the process and the effect of all activities, not limiting the scope to funded operational projects.
4. Conducting the evaluation in a rational time frame.
5. Reporting the information to decisionmakers in a way that helps them make more rational decisions.

Understanding program evaluation as a process rather than as a procedure is fundamental to evaluators being successful in their activities. In this context, success in program evaluation is defined as a development in a body of information which is perceived as being useful by individual and group decisionmakers concerned with the operation of the program and that played some part in decisions that were made.

### EDUCATIONAL PROCESS

The workshop attempted to reproduce this evaluation process. One particular regional medical program was selected so that the process and its associated problems could be illustrated.

Following an introductory lecture on program evaluation, a group of consultants met in a panel discussion with several members of the staff of the illustrative program. This panel had two major objectives:

1. To identify the philosophy, history, strategy, and activities of the program.
2. To identify the questions that the staff members felt needed to be answered by the evaluation process.

The staff members described their regional medical program as being directly concerned with transforming the medical care system through influence and a variety of activities into a system that filled the gaps in care, made better use of manpower, improved quality of care, and controlled the costs of care. They used the term "opportunistic intervention" to describe the fact that their activities were guided more by requests for assistance than by comprehensive, objective-oriented planning. "Tilling the soil" was the term they used to

describe activities designed to generate such requests in geographic areas, and among groups not aware of the availability of assistance from the Regional Medical Program. They told multiple stories of how this created change, e.g. a \$1500 contract was used to assist seven community hospitals in developing home care programs in less than twelve months; a neighborhood health center was opened by the city health department in association with the negotiations for a project on the screening and treatment of selected chronic diseases that was being submitted from the same geographic area.

The staff members developed a series of evaluative questions that they felt were important. Three major questions evolved:

1. How good are my cooperative arrangements?
2. How balanced are the program activities?
3. What changes in the health care system have the program activities influenced?

The workshop divided into three small groups to discuss the priorities of these questions for the evaluation effort and the methodologies that might be used to answer them. Each group contained staff members of the program and consultants. It was intended that each group, in addition to developing priorities and methods, deal with the practical problems of implementing the methodologies and reporting the results.

The closing session of the workshop, a general session, was designed to demonstrate the range of priorities, methods, and solutions that were available to solve common problems in program evaluation. This depended on the developments in each small group discussion and was intended to reflect the learning that had occurred in this open-ended educational format.

#### OBSERVATIONS ON THE EDUCATIONAL PROCESS

The faculty anticipated many problems in this educational endeavor. Approximately 150 man hours of planning plus two trial workshops with smaller groups made the faculty aware that most evaluators would require 10-15 hours of training before they would begin to understand the concepts of program evaluation as a process. In spite of that, an attempt was made to compress this learning experience into five hours. The planning, however, could not compensate for the short time allowed for the workshop. As a result the objectives of the workshop were only partially attained.

Participants in the workshop immediately sorted themselves into two categories: evaluation or program

had no previous experience in evaluation, participated more actively in the discussions. Evaluators, on the other hand, appeared very uncomfortable with this "process" approach to evaluation. They tended to react as if their previous highly technical, project-oriented approach to evaluation was being threatened. As a result their participation attempted to force the workshop from a process orientation towards a structured approach in which they, the evaluators, defined the questions, criteria, and methodology.

One example merits presentation: A staff member who was concerned with the quality of the "linkages" that she helped establish was badgered by an evaluator to provide a precise definition of a "linkage." Although she gave several examples of what she meant, she could not in ten minutes of interrogation give a precise definition. A program person pointed out that such an approach was not getting anywhere. This led two other program persons to present methods to describing new contacts and working relationships between individuals that, although not quantitative, did provide a basis for understanding the quality of these "linkages."

The discussion groups tended to focus on their own questions about program effectiveness and to discard the questions posed by the staff members. As the discussion groups followed this path, they became unable in the allotted time to probe into the areas of methodology. This reflected the difficulty experienced by many of the participants in understanding the concepts of program evaluation as a process and their inability to play the educational game that had been established for the workshop.

#### IMPLICATIONS

Evaluation of programs that have as broad a mandate as the Regional Medical Programs is very difficult. Proven methodologies do not exist. The effect of a social-change program is often much greater than the sum of the effects of the discrete operational project that they fund. Application of simple concepts developed in project-oriented evaluation activities is often inappropriate. Evaluators who hold responsibility for program evaluation are often the victims of their previous training and experience in project evaluation.

In order for meaningful program evaluation methodologies to be developed, traditional evaluation methodologies must become subservient to the broad range demands of these broad-range programs. The first step in this process is developing evaluation concepts that are similarly broad-range. It probably requires that evaluators no longer sit on the side-lines of the program

as judges and that they become actively involved in the entire change process with program responsibilities in addition to their evaluation responsibilities. Only when evaluators have a profound understanding of their program will they know which consultants and which methodologies are truly appropriate to the task of program evaluation.

Training and program evaluation begins with an understanding of the program to be evaluated. It proceeds to the development of program evaluation concepts. Having passed these stages, it can focus on the application of proven methodologies or the development of new methodologies. Training in program evaluation is, therefore, just as much of a process as is program evaluation itself.

## SUMMARY

The educational content and methods of a workshop session on evaluation of Regional Medical Programs has been described. The objectives for the workshop were only partially attained. Observations on the complexity of the subject, the time limitations of the workshop, and the previous experiences of the participants were related to the partial success of this particular training method. Further developments in the field of program evaluation depend upon evaluators actively participating in their own program activities and in a continuing educational process.

## WORKSHOP ON RESOURCE ALLOCATION/ECONOMICS

### Participants

John Glasgow, Ph.D. - Moderator  
Associate Coordinator, Research  
and Evaluation  
Connecticut Regional Medical Program

J.A. Florin, M.D.  
Coordinator, New Jersey Regional  
Medical Program

Michael Zubcuff, Ph.D.  
Lead, Health Economics Section  
Tennessee Mid-South Regional  
Medical Program

James R. Jeffers, Ph.D.  
Director, Medical Economics Research Center  
University of Iowa

John E. Wennberg, M.D.  
Coordinator, Northern New England  
Regional Medical Program

Conrad Seipp, Ph.D.  
Associate Professor, Health Services  
Research Center  
University of North Carolina

Robert L. Berg, M.D.  
Professor and Chairman, Department of  
Preventive Medicine and Community Health  
University of Rochester

Wayne A. Kimmel  
Department of Economics  
Public Service Laboratory  
Georgetown University Hospital

Charles W. Caldwell, M.P.A.  
Associate Coordinator, Iowa Regional  
Medical Program

Charles L. Joiner, Ph.D.  
Director, Bureau of Research and  
Community Service  
Alabama Regional Medical Program

---

### Cost-Benefit and Cost-Effectiveness Analyses in the Health Field

JOHN GLASGOW

Rising levels of health care expenditures; the associated increases in medical prices and alleged shortages of manpower and facilities; the declaration that access to medical care is a right, not a privilege; and the growing role of the government in the health care field have led to concern with the effectiveness of alternative delivery systems or resource allocations. Concern with the effectiveness of delivery emphasizes the importance of using scarce resources (or dollars) in such a way as to maximize the return per dollar spent. This, in turn, has led to the search for planning and analytical techniques which might aid in the task of rationalizing the resource allocative process. Two such techniques are cost-benefit and cost-effectiveness analyses.

The Crystal-Brewster paper<sup>1</sup> provides an introduction to cost-effectiveness and cost-benefit analysis. The present essay attempts to build upon this introduction and to suggest certain conceptual and methodological

concerns that the user of these techniques needs to keep clearly in mind if he hopes to use them effectively and if he is to understand what information these techniques do and do not provide. The purpose is not to present a step-by-step "how to do it cost-benefit manual" although one might be desirable and desired. Examples of calculations of both a hypothetical and theoretical nature, in addition to that provided by Crystal and Brewster, abound.<sup>2-13</sup> Neither is the purpose here one of exploring new theoretical frontiers. Indeed, as Klarman has pointed out "so much has been written. . . about the application of cost-benefit analysis to the health field that almost every point that might be made has been made."<sup>14</sup> Although perhaps something of an overstatement reminiscent of Mill's premature claim that everything that was to be known of economics was known, the observation has sufficient validity to narrow the present concern. The attempt here will be to ensure that terms and concepts used in cost studies are clearly understood as to their definition, the underlying assumptions, and the result and implications for the analysis. It should be clear that the objective is not to be

critical of previous work. However, an understanding of the limitations involved in such studies both increases their value to the decisionmaker and provides a reminder of the need for constant improvement of the analytical techniques involved. A secondary goal is to consolidate into one paper a number of points which are fairly well-developed in the literature, but widely scattered and therefore less accessible to the less specialized reader. A final objective is to provide to the interested reader a bibliographic resource for further personal investigation.

### THE NATURE OF THE BEASTS

Cost-Benefit and Cost-Effectiveness are terms often used interchangeably. In actual fact, the two are not the same although both concepts do derive from the same theoretical fount—capital budgeting theory. In essence, capital budgeting theory is concerned with the present and future costs, and the associated benefits over time, of alternative investment strategies. The goal is to allocate scarce resources to their most productive (profitable) uses. Thus, the theory is concerned with determining the effects, as well as the costs, of specific alternatives available.

In cost-benefit analysis, the monetary cost of a program, or intervention activity, is compared to the monetary value of the expected benefits. This cost-benefit ratio (of total costs to total benefits) might then be used to compare alternative programs to determine which is the best potential investment. For a specific activity, the comparison of costs and benefits is for the purpose of answering the question: Do the benefits received justify the expenditure (i.e., is the ratio greater than 1 or some other arbitrarily set number)?

Cost-effectiveness analysis, in contrast, attempts to compare the cost of alternative approaches to the achievement of a specific set of results. The goal, therefore, is not to determine the feasibility of achieving a goal (theoretically that has already been decided), but rather to select from among alternative approaches the one approach which will result in a given output for the least cost or the maximum output for a given cost.

Although somewhat artificial in nature, the definition of the terms does allow us to specify in some detail the major characteristics of, and distinctions between, the two concepts.

1. Cost-benefit analysis is more comprehensive in its focus than cost-effectiveness analysis.
  - a. Cost-benefit includes a consideration of social or external effects as a part of the complete enumeration of the costs and benefits. In

principle, cost-effectiveness should do the same. In practice, however, cost-effectiveness analyses are often less complete in listing the total cost and benefits. For example, external effects are often ignored and certain desired results or benefits are specified with all others regarded as constants or relatively unimportant.

- b. Cost-benefit analysis normally values the costs and benefits in monetary terms. This provides the common denominator necessary for comparisons of alternative types of programs. In cost-effectiveness analyses the measure of output often is not in terms of dollars, but rather in some other unit such as man-years saved.
2. These differences in comprehensiveness and technique result in cost-effectiveness being used most often "when various benefits are difficult to measure or when the several benefits that are measured cannot be rendered commensurate."<sup>3d</sup>
3. Cost-benefit analysis allows comparisons among several programs which have different objectives. Cost effectiveness is used to compare differing ways of obtaining the same objective.
4. The objective of a cost-benefit study is to determine if an action or program is worth undertaking; the objective of a cost-effectiveness study is to determine the best way of achieving an already determined course of action.

### CONCEPTUAL AND METHODOLOGICAL ISSUES

In this section, differences between cost-benefit and cost-effectiveness studies will be ignored for the most part. Here the concern will be with the terms used, the concepts involved, and the implications of the measurement techniques used. In general, the comments will be applicable to both types of studies.

#### *The Measurement of Costs and Benefits*

The essence of the cost-benefit approach is the assignment of dollar values to all resources so that the benefits of a specific activity might be compared to the cost of the intervention and to the projected benefits from alternative investment opportunities. Obviously, it is vital to include in the dollar valuation all the relevant effects associated with a given action.

*Economic Costs of Disease Defined.* The economic cost of disease or injury, as contrasted to expenditures for medical care, reflects both direct and indirect cost components. *Direct costs include the actual medical care expenditures necessary for the treatment of the disease or injury.* These expenditures would include both

personal (i.e., the cost of hospital care, nursing home care, physicians' service, drugs, nursing services, and similar type expenses) and non-personal expenditures (i.e., the cost of research, training, facilities, equipment, and a pro-rated share of the annual cost of health insurance). *Indirect costs are those costs to the individual or to society in the form of lost productivity attributable to the disease or injury.* In essence, this amounts to imputing a dollar value to the productivity lost through premature death or disability. Obviously, the imputation must take into account varying life expectancy, labor force participation and earning rates by different sex and age groups; the "value" of individuals outside the market pricing mechanism (i.e., housewives, elderly, children, unemployed); and the appropriate discount rate.

It is important to emphasize that the economic costs of the disease as defined above are really the projected benefits in any cost-benefit, cost-effectiveness analysis. That is, the benefits to be derived from an action are the eliminated losses in production output, personal well-being, and resource utilization which result from a successful program. The cost denominator is simply the projected budget of the program.

Enumeration of the types of factors included in the usual cost study makes it clear that a number of costs and benefits are typically excluded from the calculation. In addition, a number of assumptions, both explicit and implicit, underlie the definition of direct and indirect costs, the valuation of specific components of each, and the use of the technique of discounting which have major implications for the validity of any cost study. Again, the emphasis on the presence of biases in the technique and approach is not designed to be overly critical. Rather, the purpose is to explicitly recognize what conclusions these studies do and do not allow to be drawn from the data presented.

Despite the effort to define the economic cost of disease broadly to include both direct and indirect costs, it is obvious that not all costs and benefits are included in even the most rigorous analysis. For example, it is common to ignore the so-called "spill-over" effects. These are the desirable (or undesirable) secondary impacts of a given action. Illustrative of such a secondary impact would be the effect on prices and availability of medical care for the general population which resulted from the attempt to provide for the health care needs of the aged through Medicare and Medicaid. Another cost often not incorporated into the calculation is the cost of "locking" oneself into a given

technology when making a long-term capital investment<sup>17</sup> Other examples of significant omissions could be provided, but the point has been made. Many costs and benefits are excluded because (1) there is no known way of measuring the factor, (2) because it is assumed any undesirable side-effects could be corrected if desired through fiscal tax and transfer measures, or (3) because the analyst considers them of minor import for his purposes. Valid as the reason for exclusion may be, the fact remains that the end result is for most studies to concentrate on what is easily measurable. Unfortunately, in many cases, the easily measurable are not the most important effects which should be considered. *As a result, particularly in the health field, it is vital to avoid undue stress on the importance of economic measurements. In general, this means it is necessary to complement economic values with other non-economic values in determining the proper resource allocation.*

*The Quantification Assumption.* The most basic assumption in any cost study is that it is possible to quantify in monetary terms the benefits and costs associated with a specific activity. In actual fact, even assuming that all benefits and costs will be included, it is still not possible to quantify with precision even the most relevant factors despite major advances in measurement techniques.<sup>4,11</sup> The reasons are easily explained. The implications are somewhat more subtle.

It was noted that the benefits associated with the success of an activity tend to result from (1) increases in economic productivity due to decreased mortality and morbidity levels; (2) reductions in the need for facility and manpower resources given the eradicated or reduced health problem and (3) the existence of certain intangibles (consumer benefits) associated with good health such as reduced anxiety in the individual and society or an increased sense of well-being.

It should be clear that (1) and (2) above are more susceptible to precise measurement than is (3). As a result, most studies tend to ignore the latter effect. *The result is therefore to often significantly understate the potential benefit of any activity and to particularly underestimate the value of any activity in which consumer benefits constitute a major portion of the total benefit.* That is, since the consumer benefit component of total cost varies between types of diseases or illnesses,<sup>16</sup> the exclusion of such benefits, or even their inadequate valuation, will tend to result in a misleadingly low cost-benefit ratio for those diseases with a high consumer benefit element in comparison to

and a decision about the desirability of past expenditures. The second result might be illustrated best by a hypothetical illustration from an article by Warren Smith.<sup>8</sup>

“— assume — an investment of \$10 in a device which produces 50 units an hour and we learn of an invention of two improvements [to]increase — efficiency. Item A — costing \$5 — [increases productive capacity to]60 units per hour. Item B — costing \$7, can increase the output to 65 units per hour — which adapter would be the best choice from a cost-effectiveness standpoint?”

The total cost, if we buy item A, will be the original \$10 plus the added \$5, or \$15, and the total resultant output is 60 units per hour. Dividing output by cost gives us a ratio of 60 to 15 or 4. The total cost, using item B, will be the original \$10 plus \$7, or \$17 with a resultant output of 65, or a ratio of 65 to 17 or 3.8. The conclusion using this misleading analysis is that item A is preferred because it seems to give the largest ratio of effectiveness to cost.

The marginal or added cost for item A is \$5, and the added output is 10 units per hour for a ratio of 10 to 5 or 2.0. The marginal cost using item B is \$7 or 2.1. Our conclusion using this correct procedure is that item B is preferred because of its greater marginal ratio.”

*Additional Issues.* Although not technically cost and benefit measurement issues, four other comments need to be made concerning this general area. First, many studies distinguish between the effects of an activity on production (income) and the effect on the distribution of income resulting from the fact that beneficiaries are not necessarily those who pay for the program, that there can be an impact on relative prices and real incomes; that program investment implies foregone alternatives; and similar forces. Typically, these distributional effects are ignored in most cost studies and for good reasons. Nevertheless, it should be noted that the ignoring of these effects can lead to either an over- or understatement of total benefits derived. For example, if an activity not only treats a disease but leads to a more equitable tax policy, the ignoring of this latter fact seriously understates the value of the program.

Second, it was previously noted that one cost often excluded from most cost calculation was the effect of a program's initiation on the price and availability of resources which could have been used in alternative ways had they not been used in this activity. This implicitly assumed a state of full employment. However, where

there is significant unemployment among the resources in question, utilization in this activity not only entails little or no cost, it may provide an additional benefit. That is, the result may be a pure benefit composed of a net output gain plus reduced welfare costs.

Third, those who would make use of these techniques often desire the specification of a policy which would simultaneously provide the greatest benefit and the least cost. While theoretically possible, the attainment of this goal is limited by at least two factors: (a) limits on ability to spend and (b) requirements for expenditures of a given size. To illustrate, it is often possible to obtain a larger benefit from a larger expenditure and the increase in benefit size need not be proportional to the increase in expenditures. As a result, increased expenditures can often result in a much higher cost-benefit ratio than would be a lesser expenditure for the same activity. But if you do not have more funds to invest, the larger ratio is immaterial. In the same way it is possible that unlimited funds properly allocated among a variety of alternatives might provide a total benefit greater than the same amount invested in a single project. Yet if the required funds are limited, the use of funds in one area effectively precludes simultaneous investment in the alternative. That is, given the cost of doing A, you may not be able to do any part of B given its minimum cost requirements. This suggests two factors of import. (1) Cost analysis, in the usual case, will be able only to suggest policies which will provide the greatest benefit at a given cost or a given benefit for the least cost; and (2) in order to provide even this direction, there must exist a clear-cut statement of the objectives desired. *In short, cost studies are not a substitute for decision making, but rather a tool to help rationalize the decision making process.*

Fourth, it is also of some value to emphasize that the total dollar cost of a project does not always reflect accurately the allocation of resources which it theoretically summarizes. That is, the relevant market prices of resources do not necessarily reflect their true value (ie., actual costs) to the system within which they are being allocated. Some of the reasons why this is true have been previously alluded to (e.g., valuation of “non-market resources” or of human life itself and the use of previously unemployed resources). Other reasons include the fact that prevailing prices reflect a given income distribution. A different income distribution might result in a different demand and price structure. Finally, one might note that only if the structure of market prices is that which would occur under perfect competition

would the social opportunity cost\* equal the net cash payments for the project.<sup>18</sup>

Ideally, then, as Wennberg has noted,<sup>23</sup> the vigorous application of these techniques presupposes a detailed and accurate analysis of the system and the economic environment if the cost and benefit implications of the proposed project are to be fully understood.

### *The Discounting Procedure*

Previous mention was made of the desirability of expressing future benefits and cost in terms of their present equivalent value (i.e., to determine the present value of future dollars). The present value of future expenditures is the sum of money that would have to be set aside at present and cumulated at some rate of interest in order to equal the monetary cost of the expenditure at the time it will be incurred. Reversing the idea, one might discount a sum of future money by the interest rate chosen to get its present equivalent.

Obviously, the choice of the discount (interest) rate used in the calculation is of vital importance. Some argue that the proper interest rate to use is the prevailing market rate. Others argue that this is inappropriate for a number of reasons. No attempt will be made to examine the controversy surrounding the proper rate of discount to use since this entails a field in itself. It is of value, however, to briefly summarize some of the major issues involved in the controversy leaving to those interested the task of reading the references previously cited.

First, even a desire to use a market rate of interest is hampered by the fact that there is no single market rate. Rather the rate varies with the type of loan or obligation involved, the borrower, and time period, among other things. Second, in the choice of a proper discount for social benefits and costs associated with public investments, the choice is complicated by the existence of a close relationship between investment decisions and the social discount rate used in investment planning and between investment, the method of financing used, and fiscal policy. Third, a discount rate is intended to equate

---

\*Social Opportunity Cost is the reduction in consumption and investment which occurs due to the transfer of funds from the private to the public sector. It is the sum of (1) the amount of foregone direct consumption in the private sector and (2) the discounted value over time of the decrease in future consumption which would otherwise have resulted from the investment of the portion of after-tax income not presently consumed. For an excellent review of the concept and its development, the interested reader might consult the references to Feldstein in the bibliography. Further and more recent works are those by Baumol, Arrow, and Pauly also listed in the bibliography.

the productivity of an investment and society's reluctance to sacrifice current for future consumption. Attempts to utilize a private market rate of interest assume that the individual's time preference for money coincides with the collective preference as expressed in the market rate. This is not necessarily true.<sup>24</sup> Indeed it is argued that the individual's discount rate for the distant future will tend always to exceed society's.<sup>25</sup> Fourth, the time preference for money is not constant with age. That is, it tends to vary inversely with life expectancy. Finally, for any discount rate chosen, it is usually assumed that the general price level and productivity will remain constant over time. This is not a valid assumption, but an understandable one given the measurement problem involved. However, Klarman has suggested the desirability of developing an effective net discount rate by combining price and productivity changes that are simultaneously operative into a single rate.<sup>3b</sup> For example, one might divide the chosen discount rate by average price change (in percent). This ratio divided into the sum of the present value of output in dollars terms multiplied by the increase in productivity expected would give an effective net rate of discount.

It is clear from the above summary that the choice of the discount rate to be used, no matter how universally accepted, is an exercise in value judgment and quite arbitrary. Under these circumstances, one might wonder why the discounting exercise is performed. Blum, for example, suggests abandoning the practice.<sup>26</sup> However, it seems clear that there is no other effective way to reduce continuous and unequal dollar streams to comparable values. Consequently, accepting the need for and value of discounting, the concern is with the implications of the process for the results of the study.

*The most obvious implication is that relatively small variations in the discount rate chosen can produce relatively large differences in the cost-benefit ratio. And the greater the time span involved the greater the variance. A second implication is that the higher the discount rate chosen, the less likely programs with long delayed returns are to be given high benefit-cost ratios. A third implication, which flows from the second, is that service programs will be favored over research programs in the usual case.* As a result it often is suggested that studies should provide multiple rate analysis to demonstrate the range of priority ranking which results from different rates.

### *Miscellaneous Problems*

In addition to the biases and weaknesses imposed by measurement techniques or the discounting process, two



or problem areas might be mentioned. First, compared to the valuation of human life or some of the other indirect costs, the calculation of amounts expended for medical care (direct costs) is conceptually easy. Practically speaking, however, it may be as difficult to develop accurate estimates of these costs given the lack of available, accurate data, the difficulty of apportionment of total cost-benefits when multiple morbid conditions exist in concert, and, the existence of free services or payment in kind. To the degree this difficulty exists in a particular case, benefits may be either over- or understated by a significant amount.

Second, many cost studies subtract the cost of his maintenance from future earnings in calculating the economic value of a man.<sup>27</sup> If deducted it should be realized that the calculation can result in a value measure of output loss which might be negative. That is, it may appear that economically speaking the best course would be to "kill off" the population at risk. This is generally not considered a practical recommendation. In any case, the practice will tend to bias program selection toward those activities aimed at the "high income" or younger, more productive worker.

#### CONCLUSION

It should be clear that cost studies are not infallible guides to proper resource allocation. In fact, applied rigorously a comparison of cost benefit ratios would tend to result in a preponderance of programs serving the young adult, white, college male.

It is equally clear that given the present state of the art, no cost study can hope to include all the relevant costs and benefits or to measure even those included with any real degree of precision. Indeed, the whole process from conceptualization of objectives to measurement of benefits is a continuous exercise in value judgements compounded by a concern with events that are uncertain and often unmeasurable.

In that case, why bother with such studies at all? The reason is quite simple. If one keeps in mind that these techniques often give an unwarranted appearance of objectivity and that they are not a substitute for decision making, then these techniques can be of real value to the decision maker. They can be of value by forcing the decision maker to explicitly list the expected benefits and costs of a proposed activity and thereby allow critical examination of these claims. It highlights the presence of value judgements, assumptions and arbitrary valuations. It is, in short, a method for systematic information development, compilation, and utilization. Moreover, while it is not true that to be

useful these techniques must "yield unambiguous criteria on the project over another",<sup>23</sup> it is true that use of these techniques does force program objectives to be unambiguously specified.

Finally, one can argue that the difficulties involved in doing an adequate study has value in itself. Certainly, these problems should force the decision maker to question whether the technique should even be applied to certain problems or decisions. That is, in many cases the time required, and the sophistication of analysis involved, may be greater than required or affordable. After all, the study itself will involve the use of resource which might be more profitably employed elsewhere.

#### Bibliography

1. Crystal, Royal A. and A. W. Brewster, "Cost Benefit and Cost Effectiveness Analyses in the Health Field: An Introduction," *Inquiry* 3(4) December, 1966, P. 4
2. Department of Health, Education and Welfare, Program Analysis Group on Selected Disease Control Programs
  - a. Application of Benefit-Cost Analysis to Motor Vehicle Accidents, Office of Assistant Secretary for Program Coordination, August, 1966
  - b. Public Health Programs for Arthritis, Division of Chronic Diseases, U. S. Public Health Services, September, 1966
  - c. An Analysis of Planning, Programming, and Budgeting in Cancer Control, Cancer Control Branch, Division of Chronic Diseases, U. S. Public Health Services, 1966
3. Klarman, Herbert E.
  - a. *Economics of Health* (Columbia University Press, 1965) pp. 162-173
  - b. "Syphilis Control Programs," in *Measuring Benefits of Government Expenditures*, Robert Dorfman, ed. (Brookings Institution, 1965) pp. 367-410
  - c. "Socioeconomic Impact of Heart Disease," in *Second National Conference on Cardio-vascular Disease, The Heart and Circulation*, vol. 2 (Washington, D.C.: Federation of American Societies for Experimental Biology, 1965). pp. 693-707.
  - d. \_\_\_\_\_, J.O.'s Francis, and G. S. Rosenthal, "Cost Effectiveness Applied to the Treatment of Chronic Renal Disease," *Medical Care* 6(1) January-February, 1968, pp. 48-54.
4. Rice, Dorothy
  - a. "Economics Costs of Cardiovascular Diseases and Cancer, 1962." Health Economic Series No. 5 (U. S. Government Printing Office: Washington, D.C.)
  - b. "Estimating the Cost of Illness," Health Economic Series No. 6, 1966
  - c. "The Economic Value of Human Life," *American Journal of Public Health* (November, 1967) pp. 1954-1963.
5. Prest, A. R. and R. Turvey, "Cost-Benefit Analysis: A Survey," *Economic Journal* (December, 1965) pp. 683-735.
6. Report of the Committee on Chronic Kidney Diseases, (U. S. Government Printing Office: Washington, D.C.) 1967

7. Wildavsky, A., "The Political Economy of Efficiency: Cost-Benefit Analysis, Systems Analysis, and Program Budgeting." *Public Administration Review* (December, 1966) pp. 292-310
8. Smith, Warren F., "Cost-Effectiveness and Cost-Benefit Analyses for Public Health Programs," *Public Health Reports* (November, 1968) pp. 899-906
9. Hallan, J. B. and B. S. Harris, "The Economic Cost of End-Stage Uremia," *Inquiry* (December, 1968) pp. 20-25
10. Levin, A. L., "Cost Effectiveness in Maternal and Child Health." *New England Journal of Medicine*, 278(19) May 9, 1968, pp. 1011-1047
11. Muskin, Selma J.
  - a. "Health as an Investment," *Journal of Political Economy*, 70(2) Supplement, October, 1962, pp. 129-57
  - b. \_\_\_\_\_ and Francis d'A. Collings, "Economic Costs of Disease and Injury," *Public Health Reports*, vol. 74, September, 1959, pp. 795-809
12. Fein, Rashi, *Economics of Mental Illness*, (Basic Books, 1958)
13. Wiseman; Jack, "Cost-Benefit Analysis and Health Service Policy," *Scottish Journal of Political Economy*, vol. 10, February, 1963, pp. 128-405.
14. Klarman, H. E., "Present Status of Cost-Benefit Analysis in the Health Field," *American Journal of Public Health* 57(11) November, 1967, p. 1948.
15. Thedi, Jacques and Claude Abraham, "Economic Aspects of Road Accidents," *Traffic Engineering and Control*, vol. 2, February, 1961, pp. 589-95.
16. Weisbod, Burton A., *The Economics of Public Health* (University of Pennsylvania Press: Philadelphia, 1961) pp. 95-98.
17. Klarman, H. E., "Some Technical Problems in Areawide Planning for Hospital Care," *Journal of Chronic Disease* 17(9) September, 1964, pp. 735-747
18. See, Eckstein, O., "A Survey of the Theory of Public Expenditure Criteria," in James Buchanan (ed.), *Public Finances* (Princeton, 1961)
19. Feldstein, Martin S.
  - a. "Opportunity Cost Calculations in Cost-Benefit Analysis," *Public Finance*, 1964, No. 2 pp. 117-39.
  - b. "Net Social Benefit Calculation and the Public Investment Decision," *Oxford Economic Papers*, March, 1964
  - c. "The Social Time Preference Discount Rate in Cost Benefit Analysis," *Economic Journal*, June, 1964
20. Pauly, Mark V., "Risk and the Social Rate of Discount," *American Economic Review* 60 (1) March, 1970, pp. 195-198.
21. Arrow, Kenneth J. "Discounting and Public Investment Criteria," in A. V. Kreese and S. C. Smith (eds.) *Water Research* (Baltimore, 1966)
22. Baumol, W. J., "On the Social Rate of Discount," *American Economic Review*, 58 (3) September, 1968, pp. 788-802.
23. Wennberg, John E., "Cost-Benefit Analyses—Limitations and Uses," pp. 110-113 in *Proceedings: Conference Workshop on Regional Medical Programs*, January 17-19, 1968, Washington, D.C. Volume II, Public Health Services Publication No. 1774 (USGPO: Washington, D.C., 1968)
24. Huffschtmidt, Maynard M., "Standards and Criteria for Formulating and Evaluating Federal Water Resource Developments," (U. S. Bureau of the Budget, 1961 especially p. 11
25. Baumol, William J. *Welfare Economics and the Theory of the State* (Harvard University Press, 1952) pp. 91-92
26. Blum, Henrik L. and Associates, *Health Planning 1969* (American Public Health Association, Western Regional Office, 1969) p. 8-21
27. See, for example, D. J. Reynolds, "The Cost of Road Accidents" *Journal of the Royal Statistical Society*, 119(4) September, 1956, pp. 393-408

## Role of Social and Behavioral Scientists in RMP Evaluation\*

MICHAEL ZUBKOFF

First let me preface my remarks by stating that I believe the social and behavioral scientists' key contributions to RMPs are in the areas other than evaluation such as being an initiator of change in the region as well as aiding in the development of program strategy for achieving RMP specific goals of increased regionalization and more equitable distribution of health services.

Before turning to a "definition" of the role for social and behavioral scientists in RMP evaluation, it is necessary to spend a few moments reviewing: 1) the various levels of evaluation that exist and 2) possible strategies of evaluation within RMP.

### Levels of Evaluation

Basically there are three levels of evaluation:

1. Monitoring of specific projects.
2. Medical evaluation of specific projects in terms of quality of care.
3. Social, behavioral and economic evaluation of RMP specific goals of increased coordination and more equitable distribution of health services.

### Strategy For Evaluation

The following breakdown is suggested as a possible strategy:

#### *Role of RMPS*

1. The setting of priorities between categories and within categories, and the SUPPORT thereof, for complete end results medical evaluation of specific projects throughout the nation which RMPS feel

---

\*The author wishes to express appreciation to William Rushing, Dan Davis and Robert Metcalfe for aiding in the development of these comments.

may be worthy of possible replication (i.e., coronary care units, etc.).

### *Role of Local RMPs*

1. Monitoring of all projects.
2. Assessment of project and the program's ability to effect RMP specific goals of increased regionalization and more equitable distribution of health services throughout the nation.
3. Aiding in those "medical" evaluations that RMPS designates as needing such in-depth evaluation.

The reasoning behind this type of breakdown is basically that RMPs should "practice what we preach". In other words, we preach reduction of duplication of efforts within our region, while at the same time fostering continual duplication of efforts with respect to evaluation of projects. Without having access to RMPS records, it is impossible to tell the extent of this duplication; however, as one meets evaluators from around the nation, it is quite discouraging to discover that the same type of project is concurrently being evaluated, often without adequate support, in numerous regions. This is using up substantial portions of RMPs limited resources.

Thus, it would seem wise for RMPS to set priorities where in-depth medical evaluation should be undertaken to determine whether or not specific projects should be replicated throughout the country with RMPS supporting said evaluation in terms of dollars and manpower.

With respect to evaluation, a paradox seems to exist. RMPs are charged with trying to act as catalysts to initiate change with respect to increasing regionalization and increasing a more equitable distribution of health services, which as a process is definitely a long-term phenomenon, while at the same time the criteria being imposed by Washington for evaluation is short run.

It is important to understand that the effectiveness of RMPs must be measured as a long-term phenomenon and in fact I would suggest that if RMPs do their job as catalysts well, while documentation of change coincident to RMPs' entrance into a situation or setting will be possible, credit for their role will probably not ever be acknowledged. This can in part be explained by the difficulty and perhaps impossibility of sorting out the changes that have resulted from the program's activities and those changes which have come from other community activities.

### **Social Scientist's Role in Evaluation**

The role of social and behavioral scientists must primarily be related to those evaluations (the behavioral,

social and economic components) aimed toward assessing RMPs' specific goals of more equitable distribution of health services and better coordination of services. Here projects can often be evaluated on an individual basis although it is in terms of the TOTAL program's efforts (results of all projects) that this type evaluation is most relevant. Such program evaluation can only be done in the true sense in the long run.

The social scientist's tools for analysis of changes in the distribution of health services, regionalization and cooperation must be at the heart of ANY and ALL attempts to evaluate local RMP programs.

The methods of measuring RMPs ability to meet its goals will be many. One may study RMPs' role in bringing about:

1. Changes in functions of individual providers.
2. Changes in organization of providers.
3. Changes in the accessibility of care.
4. Changes in patterns of financing.
5. Changes in behavior following continuing education courses.

In addition to evaluation efforts aimed at judging the program's (and/or project's) achievement of its goals, there is in evaluation efforts another area in which social and behavioral analyses should pay off.

That is, trying to assess WHY a program (or project) fails or succeeds (i.e., what are the behavioral, social, cultural and economic forces that make for success or failure). There are a number of advantages to this focus. Foremost among them is the ability to anticipate the outcome of Project A (or Program Strategy #1), that is in many respects quite different from Project B (or Program Strategy #2), which has received evaluation - (e.g., if there are social and economic forces that are related to the failure/success of a physician's assistant project, the same forces may be related to success/failure of projects to recruit physicians, or even the success/failure of coronary care units).

## **The Application of Economic Analysis to Regional Medical Programs**

JAMES K. JEFFERS

### **INTRODUCTION**

Economics is the study of the allocation of scarce resources among competing needs for them. It is an economic fact of life that even our rich nation's resources are not sufficient to produce all the goods and services that we as consumers want. Therefore priorities

have to be established, and choices involving how much of our limited resources are to be devoted to producing particular goods and services must be made.

The real cost of producing a quantity of a particular good or service is the value in consumption of those goods and services not produced which could have been produced had resources been used to produce them instead of other things. Thus economics is the science of determining: (1) what needs exist, (2) how resources can be used most efficiently in the production of goods and services, and (3) how rational choices can be made among consumption and production alternatives.

#### METHODOLOGY OF ECONOMICS

The methodology of economics consists largely of abstraction, deduction, and induction. By abstraction I mean the formulation of models. Models are logical devices erected on a foundation of certain assumptions and empirical knowledge of behavior, custom, and institutions and are welded together by deductive logic resulting in one or more statements or hypotheses capable of empirical confirmation or refutation.

The trick in model building is to abstract sufficiently from reality in order to avoid the overwhelming complexity posed by the real world. At the same time, sufficient specificity with respect to key elements must be retained in order to provide reliable and relevant deductions as to how key variables are likely to be related and how they interact in real world processes. In a certain sense, abstraction plays the same role as "control" in the research methodology characteristic of the natural and biological sciences. Since social scientists in general, and economists in particular, seldom have an opportunity to "standardize" populations or otherwise manipulate social conditions with the exactness of environmental control provided by modern laboratories, theoretical abstraction permits, at least, clear thinking concerning a few highly important elements of a complex system or process.

The resultant of the construction of a theoretical model is the clear statement of behavior or of a relationship that logically exists given the assumptions and empirical knowledge on which the model is based. As such, these statements purport to say something about reality and may be useful in the sense that they provide a logical explanation of how certain things of interest work. Very often they are convenient ways of "looking at things" and are suggestive of new relationships and new "ways of looking at things" as well.

For the scientific researcher, however, things cannot terminate with accepting such propositions simply

because they are plausible. Many propositions are plausible, but not all are true. Such statements rightfully should be regarded as conjectures or hypotheses and should not be regarded as scientifically meaningful unless they relate specifically to a body of data that in principle could be examined by some means for the purpose of adding support or rejecting the existence of the relationships proposed by the theory.

This is the point at which inductive reasonings take over. The statements produced by theory are deductive generalizations set somewhat unfirmly on a foundation of assumption and on some not so certain "knowledge." The truth of the theoretical conjectures may be presumed to bear no more closeness to reality than that of the truth of the assumptions and "facts" on and from which they are drawn. Thus these theoretical conjectures must be tested against data purporting to describe reality.

In economics such tests are usually conducted statistically. While it is not usually possible to effect environmental control in sufficient measure to make data conform to the degree of abstraction required of the theory, advances in the theory of statistical inference and econometrics permit a degree of standardization of variables permitting the testing of many, but not all, theoretical conjectures. Multivariate analysis, as exemplified by analysis of variance and multiple regression techniques, permits the estimation of the relationship existing between economic variables of greatest interest while at the same time neutralizing the impact of other variables on these relationships.

Thus the final "proof of the pudding" in economic analysis lies in answering the question: Do the hypotheses advanced on the basis of theory square with the facts as exhibited by real world data? If the answer is no and if the assumption that statistical design used in testing is appropriate, the hypothesis must be rejected and the theory discarded as not being useful. If the answer to the question is yes, the theoretical conjectures should remain in the list of plausible explanations suggesting "how things work" in the real world until such a time as subsequent empirical investigation may refute the theory. Meanwhile the estimated values of the parameters of the relationships identified may be used for policy purposes.

Thus is the methodology of economics. It is summarized in Figure 1.

The emphasis in economics is on explaining the behavior of the economic aspects of a social system, and therefore a premium is paid for a theoretical *explanation* that is consistent with reality as opposed to a mere

*description* of reality. Pure induction involving statistical analysis, let us say correlation techniques, may provide a good description of what is "going on" in a social context. But statistics by themselves never provide answers concerning "why" things go on the way they do. Data do not interpret themselves, but rather they must be interpreted within a context of logic involving cause and effect relationships. Thus interpretation of economic phenomena is facilitated by a clear statement of the theoretical relationships that logically may be expected to exist. This logic is incorporated in what economists mean by a model which, as explained above, is merely an abstract prototype of how key variables may be expected to be related in the real world.

In my experiences many medical administrators underemphasize the importance of a clearly specified theoretical model prior to the collection of data. In many cases great haste is made to collect data without a clear conception as to how the data may be analyzed or interpreted to provide answers to questions essential to program planning and evaluation.

## APPLICATION OF ECONOMICS TO PROGRAM PLANNING AND EVALUATION

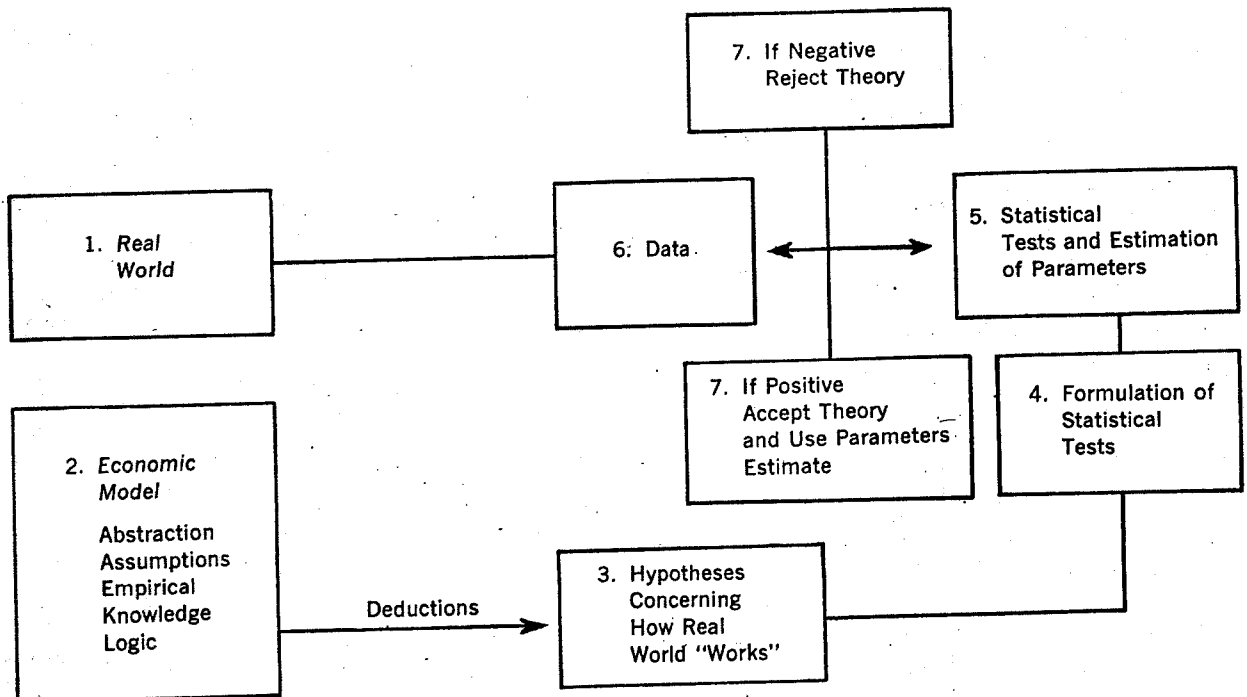
The process of planning involves a continuous conscious effort involving the following elements:

1. The specification of objectives of the course of action being considered.
2. The specification of alternatives by which objectives may be obtained.
3. The collection and interpretation of relevant data and information.
4. The specification of the potential costs and benefits of each alternative means of reaching each objective.
5. The development of a model that abstracts the relevant features of the situation being considered.
6. The specification of a decision-rule or criterion by which it is possible to rank alternative ways of attaining objectives in order of their desirability.

Effective evaluation is also a continuous process and differs from planning in the following respects:

1. Alternatives are not considered in the course of evaluation since a course of action for attaining a desired

FIGURE 1.—*Economics Methodology*



objective has already been selected. (However, after evaluation has been performed, it may be decided to terminate a particular program in favor of some alternative);

2. Costs and benefits are measured in actual rather than in potential terms;

3. The model abstracting the relevant features of the situation may be modified in light of experience, the accumulation of data, or refinement in its design; and

4. The decision rule or criterion adopted should apply consistently for all implemented programs for purposes of assessing their relative contributions to the overall objectives of the program.

One of the prime requisites of effective evaluation is the statement of the objectives of a given program. Statements of objectives should not be too broad and imprecise, should not be conflicting, and should be stated in quantitative terms whenever possible so as to facilitate both planning and evaluation.

A much too broad statement of an objective for a regional medical program would be: to reduce the pain, suffering, and mortality of heart patients living within the boundaries of the region of consideration. The statement is much too broad since any coronary care program, be it one of continuing education or one involving the use of a mobile coronary intensive care unit, would conform to the objective, and it would be impossible to judge the relative efficacy of these two programs.

An example of a conflicting statement of objectives might be: to reduce the morbidity and mortality of coronary disease in a given region. This statement of objectives is conflicting because the reduction of coronary mortality may well raise the average number of heart attacks experienced by many patients, thus raising morbidity in statistical terms. Clearly, reductions in morbidity and mortality are desirable, but it should be recognized that these objectives are conflicting. They should be stated separately, and decision makers must be prepared to compromise between the attainment of both objectives since they are in conflict.

Neither of the statements above are sufficiently quantitative in that they fail to clearly relate to a body of data that may be examined in the interest of planning and evaluation. A better statement would include a specification of the *extent* to which improvement in the condition of patients is expected. An example of a better statement would be: to reduce the morbidity of coronary heart disease by "X" percent over a specified time interval. Of course, the specification of the exact percent of reduction of morbidity or the exact time

interval must be reasonable and initially can only be determined or estimated on the basis of the experience of other programs conducted elsewhere or on the basis of expert opinion.

Economists can be of some assistance in developing statements of objectives. Economists can point out objectives that are conflicting and can assist in the development of quantitative statements. However, the ultimate responsibility for doing so lies with regional advisory groups.

Given an appropriate statement of objectives, economists can make a very significant contribution to the evaluation process in the areas of modeling, data collection, and the analysis of data.

#### COST-BENEFIT AND COST-EFFECTIVENESS MODELS

There are many specific models which economists have developed over the years that would be useful to program evaluation. They are too numerous to describe in the space allotted. Therefore at the risk of omitting many models that may be of interest for the purposes at hand, I will briefly describe the one that, in my opinion, is particularly useful. This is the cost-benefit or cost-effectiveness model. It is particularly useful for evaluation purposes since in principle it permits the simultaneous evaluation of the performance of several different operational programs. Economists generally regard cost-benefit analysis as an offspring of welfare economics and public finance, although the first practical applications of the technique were made by engineers in this country around the turn of the century. However, economists had initially developed the technique in the middle 1850's and had refined the principles of the methodology by the early 1950's.

In essence, cost-benefit analysis is a way of evaluating the desirability of a project or of a set of projects when it is important to view project activities over a long time span where there are likely to be many spill-over or side-effects on people, other programs, and other activities. In simplest terms the method consists of a careful enumeration of all direct and indirect elements of costs and benefits. It should be noted that when benefits and costs are not measured in comparable units (e.g., dollars), the technique is usually, but not always, labeled cost-effectiveness analysis—thus explaining the difference in the terms used to title this section of the paper. The phrase "cost-benefit analysis" will be used throughout the remainder of this paper.

Cost-benefit analysis involves a comparison of costs and benefits associated with a program or set of pro-

their relative significance, and draw conclusions concerning the nature of the patient's illness. Programs where the latter are viewed as alternatives or competitors for overall program funds. The cost side of the equation consists of estimated or realized program expenditures as itemized in program budgets with due allowance for the real costs of resources voluntarily contributed to the project effort. In general, benefits are viewed as future losses that will be avoided by the success of programs. The major purpose of health programs is to save lives and reduce illness. There are three general categories of benefits: (1) gains in economic output (usually measured in terms of income), (2) satisfactions from improved health, and (3) savings in the use of health resources.

Before going further it should be noted that some of the differences among authors as to how they measure benefits are due to differences in the availability of data and do not reflect philosophical differences as to the appropriate use of the methodology. However, some differences of philosophy do exist even if the same authors had access to identical data. Mention will be made of this later in the paper.

Once having enumerated all types of benefits and costs, usually some sort of discounting technique must be discounted by an appropriate interest rate to adjust comparable. This is because benefits are likely to be realized over an appreciable period of time and costs are usually incurred in the present.

For example, the benefits of a program designed to save lives may be measured by the earnings of individuals whose lives are saved over the period during which their lives have been extended. Since such earnings extend for a significant time in the future the income stream must be discounted by an appropriate interest rate to adjust future earnings downward rendering them comparable to costs that are incurred in the present. The choice of the appropriate interest rate is as yet an unresolved theoretical issue and thus in most applications several interest rates are used resulting in alternative estimates of the discounted benefits of each program.

Once having measured costs and benefits for several different projects we can make a comparison between them. If we are faced with selecting one project to the elimination of all others, the analysis is simply a matter of determining which project has the largest benefit to cost ratio and implementing that project. Note that no provision is made for the project with a benefit to cost ratio of less than one. Such a project would not be undertaken since the returns to such a project would be exceeded by the costs of the project.

Now we can consider the case for a set of projects which can be participated in at varying levels rather than in an absolute fashion. Attempt is made to achieve:

where subscripts 1-n represent different projects.

$$\frac{mb_1}{mc_1} = \frac{mb_2}{mc_2} = \dots = \frac{mb_n}{mc_n}$$

$mc_i$  is the marginal cost of the  $i$ th cure.

$mb_i$  is the marginal benefit of the  $i$ th cure.

It is profitable to participate in a program until  $\frac{mb_n}{mc_n} \geq 1 + i$ ; that is, as long as benefits achieved are  $i$  (where  $i$  represents the discount rate) times greater than the cost of producing the benefit. With this consideration in mind, the optimally sized regional medical program budget is one which allows that all projects a region wishes to undertake are participated in to the level that the return from each project is

$$\frac{mb_1}{mc_1} = \frac{mb_2}{mc_2} = \dots = \frac{mb_n}{mc_n} = 1 + i.$$

#### SOME CONCEPTUAL DIFFICULTIES

The first conceptual problem that one encounters is in developing appropriate measures of benefits. One is tempted to measure what appears to be objective and reproducible at the expense of other benefits not so easily measured. The economic gains of saving lives is usually measured by taking account of the increased income stream forthcoming to the individual whose life was saved. This is tantamount to saying that the value of a man is what he earns and neglects the affection accorded to the aged who have lived a productive life and who are retired and who are no longer employed. As yet a satisfactory measure of the loss of a "non-productive" member of society has not been devised. Similarly no indices of the welfare gains stemming from reduced pain and suffering exist.

Even if income or earnings are adopted as the appropriate measure of benefits, questions remain concerning whether income net of consumption should be the measure or whether gross income should be used.

#### CONCLUSIONS

Mention of these problems serves to underscore the fact that economic models in general and cost-benefit

analysis in particular cannot provide easy objective answers to all questions involved in program evaluation. However one of the major contributions of an economic model is to systematically categorize the key economic issues, variables, and relationships that are involved. Once these have been set out, analysis using objective data provides guides as to appropriate decisions. Even if complete answers cannot be provided on the basis of objective data and analysis, a systematic specification of the evaluation problem coupled with what objective evidence is available facilitate the consistent application of judgment and expert opinion so vital to correct decisions.

#### Summary of Remarks

JOHN E. WENNBERG, M.D.

A successful health planning and management capability requires the development of an adequate data base. This should be approached through the use of multiple disciplines in both the design and analytic phases. Relevant disciplines include biostatisticians, epidemiologists, economists, sociologists and systems analysts.

The NNE/RMP has developed a planning and evaluation base by assembling existing data sources into a compatible, computer-based system. The data base has been supplemented by ad hoc field studies involving retrospectively collected utilization data and facilities inventory. In addition, a complementary field social survey capability has been organized.

Details concerning the data system are reported at another conference session. Here I would like to report by way of example how socio-economic analysis, using information in the data base, can help clarify, if not answer, certain questions of concern to planners.

The questions chosen for example include those related to the cost of care and consumer preferences and opinions. The importance of these questions to the planning process will be emphasized.

#### Social Scientists and the Process of Evaluation

CONRAD SEIPP, Ph.D.

The field of evaluation is like the field of heart disease, cancer, and stroke. In both there is a serious gap between what we are able to do and what we are in fact today doing. We know a good deal more about the process of evaluation than current practice suggests, it is my contention. There is a substantial body of meth-

odology for evaluation, we command some potentially powerful techniques for this purpose, but we have harnessed very little of their promise in a systematic or organized way.

Like heart disease, cancer, and stroke we seem to lack the ability to relate the various pieces of the technical competence we command to pursue evaluation into meaningful total arrangements. The involvement of social scientists in the evaluation of regional medical programs is likely to prove productive only to the extent that there is widespread understanding and conceptual clarity on the part of program administrators about the evaluative process. Social scientists on the basis of the particular skills they possess are in a position to contribute to the evaluation of on-going programs. However their relevant role is restricted and confined to certain discrete levels of the process of evaluation. Further, their entry into the process most often presupposes the exercise of a great deal of prior normative judgment.

In order to use social scientists in appropriate ways in the evaluation of social programs, it is necessary to be clear about the different levels of the evaluative process and about the underlying values which assert themselves in any particular program under review. We must be able to specify the purposes to be served by evaluation and the criteria of judgment that are reflected in the formulation of those purposes.

Program evaluation is predicated on various essential assumptions, however obvious these may appear to be. It is necessary, for example, to accept the belief that a program embraces purposive activity, that socially valued resources are deployed with intent in order to accomplish something. Programs must have goals if they are to be evaluated. We are also sensitive to the fact that resources are limited. At a time when the availability of resources appears to be becoming progressively tighter, this is another premise which is easy to accept. Programs accordingly reflect the exercise of some form of rationing. The first essential task that we face in the process of evaluation is therefore to ascertain the extent to which our programs are accomplishing the goals which we have set for them. Programs consist of a bundle of more or less discrete projects. If the planning of a program has reached an acceptable degree of precision, each of its constituent projects possesses a clearly defined set of targets. A target is a statement of the end results which are sought through the activity that is called for in a project. It identifies the amount of accomplishments, if possible in quantitative terms, to be achieved within a specified period of time. A number of projects are collectively the means for achieving the objectives which



are set for the total program. Thus, the first level of evaluation is in principle at least relatively simple and clear. It is to measure the extent to which various projects are meeting their targets.

It is an integral part of the responsibility of the administration of a program to ascertain the extent of the progress that is made in fulfilling project targets. There is no point in becoming involved in other levels of evaluation unless this kind of intelligence is at hand. Perhaps the particular tasks and routines upon which reliance has been placed in the planning of some project came to be viewed as inadequate on their definition, their organization or their implementation, but it makes little sense to evaluate the adequacy of these unless there is firm knowledge of where a project stands in meeting the targets which have been set for it. Similarly, it is pointless to attempt to assess total program accomplishment until the extent to which the targets of the component projects are being met has been ascertained.

Buried in the targets of a project, however, are a host of value judgments which need to be made explicit if evaluation is to be pursued at a higher and more inclusive level of concern. Program people in the field of health and medical care still speak of securing the greatest possible return, however this may be measured, for the least expenditure of socially valued resources. One knows what they mean when they say that they want to get the most for the least. However, this kind of formulation of the economizing intent of a program is inimicable to evaluation. It must be challenged if evaluation is to proceed. For the most is in theory infinitely great and the least is zero, and this makes nonsense of their concern. A program administrator is motivated either to maximize output, the desired end results of a course of action, with a given input or he wants to secure some specified accomplishment with the minimum expenditure of socially valued resources.

Those responsible for a program are most often, in fact, motivated both to accomplish as much as they can with the resources at their disposal and at the same time to reduce what is required to achieve the objectives which they entertain. The evaluator, however, cannot simultaneously pursue both concerns, for they constitute separate and discrete analytical tracks. Each must be independently assessed as part of the process of evaluation. Further, the evaluator must ascertain the relative importance to be attached to each in a particular program. This rests upon a normative judgment which constitutes a given at this second level evaluation.

Correlative to this distinction is the differentiation between the effectiveness of a program or a project and

its efficiency. The quest for efficiency lies in reducing inputs per unit of output, of minimizing the resources which must be expended to obtain a target or a set of objectives. In the case of effectiveness, we want to know how much we are getting as return on the resources we are expending. These must be seen as separate problems to be dealt with in the process of evaluation. The evaluation of a program entails analysis along both lines. How the resulting intelligence is to be assembled into a comprehensive assessment of a project or a program depends upon the assumptions and the suppositions, the bias, if you like, which is incorporated in it. Program evaluation should conform to the norms and the criteria of judgment which are manifest, however covertly, in the planning and the design of a program, even though the evaluation that is done of a particular program by others may be predicated on different normative grounds.

The thrust of these comments is to underscore the importance of clarity about the values that inspire evaluative effort. Evaluation, involving the measurement or assessment of program accomplishment, proceeds on the basis of certain standards of comparison and particular, normative criteria of judgment which are current in a program and these must be understood and made explicit. The social scientist who is involved in the evaluation of programmatic endeavor has an important contribution to make in exposing and laying bare the construct of values which are reflected in a particular program. The need to insure a continuing explication of value premises is not only a requisite for meaningful evaluation; it must also be made an inherent attribute of program planning. This is the point at which planning, evaluation and research, meaning evaluative research, operations research, administrative research, call it what you will, emerge most explicitly as aspects of a single function.

The ways in which the social scientist is currently involved in this aspect of program evaluation is at best shadowy and uncertain. The relevance of his skill at this level of concern needs to be more fully appreciated and the role which he potentially can play requires more definitive delineation. The credentials which the social scientist commands to enhance the sensitivity of the staff of a program to the value implications of their actions are none too solid or convincing. His contribution in this regard is surrounded with difficulty. Further, the more penetrating and critical he is, and thereby the more useful, the less appreciated he is likely to be.

The task which I am suggesting for the social scientist at this level of concern is to ask those administering a

program why they are doing what they are doing, what evidence they possess to validate the assumed worth of those actions, and how they see the consummation of particular tasks and activities as related to the attainment of the broader objectives of their program. The social scientist is hopefully equipped somewhat more adequately than others to recognize the ways in which diverse values assert themselves in a program and to appreciate the various social roots of the normative judgments that are reflected in the activity he observes. His presence first of all may help to make this dimension of a program's endeavor more explicit. He is able to assist others in identifying and acknowledging the normative premises upon which action is based, in recognizing the existence of forces which militate for alternative standards of judgment, and in exposing inconsistencies between the value base of different parts of a program. In this respect the social scientist's role within a program is essentially one of education; it involves increasing the self-consciousness of the staff of a program about the social forces which impinge upon them and of which they are a part.

The social scientist can obviously make no exclusive claims to such a role. Yet he is in a position to deploy the special competence he is assumed to command in clarifying the normative bias of a program, particularly as it is expressed in the functional linkages and relationships which the program generates. In this he helps to expedite the process of evaluation at the same time that he contributes to the course of planning. His contribution, if he functions with effect, is to facilitate the formulation and appreciation of a clearer, more meaningful design of the interrelations between ends and means. Each project, I have suggested, should have an explicit target, an end result which has been operationalized as a measurable accomplishment to be achieved within a specified period of time. However, each project must also be seen as the means for attaining the objectives of a program. Further, the place of the program as a part of the endeavor to realize the aims of a more inclusive health plan must be adequately visualized. This is the essential conceptual matrix for the conduct of effective evaluation.

Given an adequate spell-out of this kind of a hierarchy of goals and of the interrelations between them, the problem of program evaluation is not especially complex, it seems to me. Assessment of the extent to which targets are achieved, even the measurement of the accomplishment of program objectives, can and should proceed without any particular need to enlist

the assistance of a social scientist. There do not appear to be any compelling reasons to suppose that the social scientist has a unique contribution to make in such tasks as ascertaining whether the development of a coronary care unit in a hospital is on schedule or determining where things stand in instituting a tumor registry. The same applies if the evaluative concern in regard to the tumor registry is less proximate and centers upon an assessment of its consequences or impact. If the planning of the program has been adequate, the problem is to determine the extent to which the project did in fact fit into the larger scheme of the regional endeavor as intended. Very possibly the talents of a social scientist might usefully be drawn upon if the issue that emerges in the course of the process of evaluation comes to center upon the efficacy or the validity of the technical prescriptions that a program has made to achieve a particular end result. Yet this type of concern, I would argue, should not be included as a primary function of program evaluation. Rather, it should be considered as an assignment for evaluative research which relies upon a different institutional base and set of resources. Regional medical programs will inevitably become involved in such activities but not, as I see it, as the agents who have a primary responsibility for undertaking such analysis. Rather, they should be a part of a larger consortium of concern that is involved in the pursuit of such questions. The social scientist does appropriately come back into the process of program evaluation in the appraisal of the broader and less specific objectives of a program. Here, for example, one encounters the need to evaluate success in promoting the legally mandated obligation to promote cooperative arrangements as an end in itself but also at the same time to see those arrangements as instrumental to improvements in the health care delivery system. However, at this level of concern the performance of those social scientists who have been involved in the evaluation of programs is far too often disappointing. There is a gross disparity between performance and promise. Social scientists tend to be mesmerized by a conviction in experimental design as the only road to salvation and they are reluctant to abandon the rigor and the apparent certainty that such procedures imply. Only slowly and with great pain are they learning of the tremendous practical difficulties of imposing experimental designs upon on-going social programs. Yet there are also theoretical grounds for suggesting that excessive emphasis has been placed upon the controlled experiment as methodologically essential in the evaluation of programs with broad and ambitious aims. These are as compelling to me as the many

obstacles which are encountered in practice in trying to execute an experimental design.

Let me with desperate brevity identify some of those problems. It is extremely difficult for those responsible for a social program to establish and maintain control groups, since this entails withholding the assumed benefits forthcoming from some intervention for some individuals or groups at the expense of others. Not only is such practice often difficult to justify, but problems are also frequently incurred in trying to enforce such discrimination. Many courses of action are simply not divisible so that the possibility of controls is precluded. Another issue is the growing resistance on the part of various segments of the consumer public to serving as subjects for research. Many are not impressed, to understate the point, with what they deem to be the academic games which the research worker engages in and they are unwilling voluntarily to offer themselves as a statistic in the promotion of a professional career. Yet even more important is the friction which the conduct of experimentally designed evaluation studies tends to create between those responsible for the administration of a program and the social scientist who is involved in such investigation.

From the beginning of a program, administrators are likely to seek advice and assistance from the evaluator in refining the design of the program and in implementing the plan as it is formulated. However, this inevitably commands second priority to the evaluator who is committed to a protocol which involves experimental design. Once he has established his controls, it is necessary for him to try to hold a program frozen in order to preserve the continuation of the experiment he has constructed. Good planning from the viewpoint of the administrator may involve maintaining the greatest possible flexibility in order to maximize opportunities to capitalize on unforeseen options. But good evaluation in the eyes of the research workers seeking scientific rigor is certain to militate against departures from a predetermined course of intervention. The exclusiveness of these concerns all too often eventuates in mutual recrimination, if not outright disaster.

Where the context and the environment of a program are highly fluid and dynamic, where objectives are far-reaching in their consequences and are hard to operationalize or specify with precision, where the unintended results of interventions are felt of major importance, the program administrator would do well to be critical of what the social scientist can accomplish within the framework of experimental design. The social scientist, in turn, should try more consciously to bring to the task

of evaluation some methodological alternatives for analysis at this level of the process.

If the social scientist can't rely as fully upon experimental design as he would like, what methods and procedures can he appropriately pursue in evaluating social programs? Admittedly, he does not have much that is firm to fall back upon, but there may be more than is currently acknowledged. I would identify a body of effort that has come to be identified as "process analysis." It is also possible to point to the successful use that has been made of the critical incident technique, an ad hoc kind of appraisal of developments as they actually happen to unfold. Much more can be made of investigations of the communicative network to encompass a systematic appraisal of the volume, the quality and the pathways of exchanges and transactions.

At this stage of things we must resign ourselves to the necessity of being selective and of concentrating our attention on certain indicator states and conditions, rather than the entire range of potential consequences of a program. One important area to look for help from the social scientist in developing such indices relates to the assessment of satisfactions and the crossmatch between measures of provider and consumer perceptions and aspirations. These are examples of some of the kinds of analytical concerns and techniques which the social scientist can bring to the evaluation of the broadest and most open-ended objectives of your programs.

The hope of enlisting social scientists in such endeavors depends to an important extent upon the effectiveness with which they are involved in the other levels of evaluative effort which I have tried to characterize. They must be given the opportunity to learn for themselves that the broader the aims of the programs they engage in evaluating, the more inadequate is their penchant for experimental design likely to be. The more rapidly they are able to recognize the necessity of relying upon less rigid evaluative methods and procedures, the sooner they will see their way to relevant endeavors which challenge them not as mere technicians but as seekers of new knowledge. Two things are necessary if this is to come about. The administrators of programs must become more certain about the importance of evaluation and they must acquire a greater confidence about their role in that process. The social scientist, in turn, must become more modest about the potential of his claims in regard to evaluative undertakings and seek to develop methods and procedures which can supplement, at some critical points, substitute for the design of experiments upon which he has too often and too unrealistically come to depend.

Cost Benefit—Cost Effectiveness  
Studies, and Their Application to  
Allocation of Resources

ROBERT L. BERG, M.D.

Making choices among alternate proposals in the RMP characterizes something rather new and uncommon in the health field: the acceptance that there are limits to what can be spent in money or in specific resources such as hospital days, or doctors' or nurses' time. In many respects, health care is dispensed as if there were no limits. Except in certain prepaid group health plans, there has been little budgeting in ambulatory or hospital care. Physicians behave as if they were guided by a fantasy that they always do everything possible for their patients. But none of us has ever done everything possible for his patients. We have always made choices: how much time we spend with a patient, how long we keep him in the hospital, how many tests we order.

The following proposal is an attempt to make explicit the grounds on which medical care choices are made. No new value judgment is suggested. Rather an effort is made to understand what values guide health care decisions, and in understanding them to make medical judgments more consistent with themselves and with these principles. Furthermore, if we can make the values sufficiently explicit, a whole new world of decision-making is opened through powerful econometric models which permit rational allocations of resources in circumstances where the complexity of data and interrelationships exceed the capacity of common sense decision-making.

Let us set a goal for any allocations problem: to get the most for our money. But the problem is to define "most." To most patients and doctors, the "most" is to save a patient's life. But one patient's life may not have the same social significance of another, an evaluation mostly related to the age of the respective patients. If there must be a choice, the 40 year old will usually be preferred to the 90 year old. This has led to the use of "life-years saved" rather than lives saved (cf. Michael, Spatafore, et al). Public policy adheres more nearly to this notion of life-years saved than the old favorite of economists, "life earnings saved."

What is the practical significance of life-years saved when resources are limited? On the basis only of life-years saved, would you give the next spot on an artificial kidney to a 90 year old or a 40 year old? Most of you would select the 40 year old. But what would be the choice if there were only time or resources to save either one 40 year old man from a burning building (or a sink-

ing ship, or on a desert expedition) or two 65 year old. Other things being equal, the "life-years saved" principle should lead you to opt for the alternative involving the most life-years saved. In this case (using 1967 data) the maximal number of life-years saved would result from saving the 40 year old (34.3 life-years versus  $2 \times 14$  years = 29.6 life-years for the 65 year olds). This leads to an action favoring one life saved over two lives saved,—troubling to all of us committed to doing all we can for our patients, but probably compatible with public policy in the health field.

If you have survived this painful moral dilemma, let us turn to another perplexing issue: that medical care may not save many lives but is mostly concerned with improving the condition of life. Instead of simple life-years saved, our more usual achievement may be something like years of improved function saved. Here we come against the major unresolved issue in the benefit field (for we are clearly dealing at this point solely with the benefit half of the cost-benefit problem): namely how do we weigh years of improved function as compared with life-years saved. We intuitively make such judgments in medical care situations. We work harder to save a life than to improve function, but how much harder and under what circumstances?

If we can explicitly articulate the basis for these judgments, we can be more consistent in future judgments. Take the example of the burning house. It contains three 40 year old persons. Two single quadriplegics (totally bed-ridden) are on the bottom floor, and a single working person on the top floor. Assume all three have the same life expectancy. In your best judgment, you can either save the two quadriplegics or the single working man—what would you choose?

If you save the working man, then you are assigning a value of something less than one half the value to the life-year of a bed-ridden patient as a working man. Or, if you save the two quadriplegics, you are assigning them something more than one half the value of the working man.

This comprises the basis for a proposed benefits scale as below. The assigned values are arbitrary but not out of keeping with public opinion.

- 1.0 good health, working
- .7 not working, at home
- .4 not working, institution
- 0 dead

If policy makers were able to decide on such a benefit scale, and with whatever data is available on the results of medical programs, rational explicit decisions could be made with such a scale as a basis.

For example, relative benefits of competing programs could be calculated in terms of life-years saved. But saving the life of a working man counts more than saving the life of a bed-ridden patient: more than twice as much.

Relative benefits could also be calculated for improvement of function. For example, returning two and a half, not-working institutionalized patients to work would be equivalent to saving the life of a working man.

These rough and ready calculations have not taken into account the number of years a patient would gain in each functional category, and a proper calculation includes these adjustments.

A theoretical example with hypothetical numbers illustrates how these calculations could assist in choosing among four proposed projects to be funded from an RMP.

The calculations (Appendix A and B) indicate how to get the most for the money: put it all into the program where you get the most for the money: the EDDU (Early Disease Detection Unit) at a cost of \$2351 per benefit unit.

But this solution does not take into account the number of stroke patients to be rehabilitated nor the number of doctors, nurses, or hospital beds available. It is this quandary that the multiple equation linear programming model helps to resolve. Given 20,000 doctors' hours (on an annual basis), 100,000 nurses' hours, 15,000 bed days and 100,000 office visits, what is the optimal mix of programs (not the single best program)?

The example in Appendix C indicates that the optimal mix would be:

750 stroke patients rehabilitated

6992 persons screened in EDDU

0 patients cared for in ICU

430 patients entered in a cancer registry

Some RMP's may operate as if there is only a money constraint and no limit to the number of available doctors, nurses, or hospital beds. But increasingly plans must determine the most efficient way to deploy our limited resources. Doctors and nurses will be available for a new RMP program only if they are lifted from other programs.

It must be pointed out that the data are imaginary and indeed identify a major need for any allocation technique: better data on the effectiveness of medical programs. The lack of such data is no comfort for the intuitive planner as compared to the explicit cost-benefit planner. Both are at a great disadvantage. But the technique forces any planner to specify what expectations he has for his proposed programs and what values he attaches to the results.

In summary, cost benefit analyses encourage the planner to specify his value system and behave consistently with it and to be explicit as to the benefits he expects from given programs. It then allows sophisticated solutions for problems too complex to be solved intuitively.

#### REFERENCE

Michael, Jerrold M., Spatafore, George, and Williams, Edward R., "A Basic Information System for Health Planning," PUBLIC HEALTH REPORTS, Vol. 33, No. 1, January, 1968.

APPENDIX A

Calculation of Benefits

1.0	$S_1 =$ Working	$\left[ \begin{array}{l} 1 \text{ Working Year Saved} \\ \text{(Patient Otherwise Would Die)} \\ = 1 \text{ Benefit Unit} \end{array} \right]$
0.7	$S_2 =$ Not working – at home	
0.4	$S_3 =$ Not working – institution	
0	$S_4 =$ Dead	

Example

Coronary Care Unit

Assume benefits of program equal any improvement in functional state due to program.

Assume 1 patient in 20 survives who would otherwise die, and assume his functional status for balance of his life is as follows:

$S_3$ (Acute Hospital)	$0.1 \text{ year} \times 0.4 \times 0.05 = .002$	$\frac{[S_4 - S_3]}{}$
$S_2$ (Home – Not Working)	$0.3 \text{ years} \times 0.7 \times 0.05 = .0105$	$\frac{[S_4 - S_2]}{}$
$S_1$ (Working)	$5.0 \text{ years} \times 1.0 \times 0.05 = .25$	$\frac{[S_4 - S_1]}{}$
$S_2$ (Retired–home)	$5.0 \text{ years} \times 0.7 \times 0.05 = .175$	$\frac{[S_4 - S_2]}{}$
	<u>.4375</u>	Benefit Units Per Patient

By Similar Calculations:

Stroke Rehabilitation	= .9400 Benefit Units Per Patient
Early Disease Detection Unit	= .0265 Benefit Units Per Patient
Cancer Registry	= .1107 Benefit Units Per Patient

**APPENDIX B**

*Comparative Costs Per Benefit Unit  
For Each Patient in Program*

Assume            MD Hours            \$ 15 (not counting overhead in institution or office)  
                      RN Hours            \$ 4  
                      Hospital Days        \$100  
                      Office Visits         \$ 5 (not counting MD income)

	<u>Stroke</u>	<u>EDDU</u>	<u>ICU coronary unit</u>	<u>Cancer registry</u>
MD hours	20 x \$ 15 = 300	0.1 x \$15 = 1.5	20 x \$ 15 = 300	10 x \$15 = 150
RN hours	80 x \$ 4 = 320	0.2 x \$ 4 = .80	150 x \$ 4 = 600	20 x \$ 4 = 80
Hospital days	20 x \$100 = 2000	0	10 x \$100 = 1000	0
Office visits	10 x \$ 5 = 50	12 x \$ 5 = 60	0	20 x \$ 5 = 100
Total cost per patient	\$2670	\$62.30	\$1900	\$300
Cost unit benefit	$\frac{2670}{.9400} = \$2840$	$\frac{62.30}{.0265} = \$2351$	$\frac{1900}{.4375} = \$4343$	$\frac{330}{.1107} = \$2981$

APPENDIX C

*The Most Efficient Program*

3

$$\text{Maximize} = 0.94X_1 + 0.0265X_2 + 0.4375X_3 + 0.1107X_4$$

	<u>Stroke</u>	<u>EDDU</u>	<u>ICU</u>	<u>Ca. Reg.</u>	
MD hours	20X <sub>1</sub>	0.1X <sub>2</sub>	20X <sub>3</sub>	10X <sub>4</sub>	20,000 (10 doctors)
RN hours	80X <sub>1</sub>	0.2X <sub>2</sub>	150X <sub>3</sub>	20X <sub>4</sub>	100,000 (50 nurses)
Inst. days	20X <sub>1</sub>	0 X <sub>2</sub>	10X <sub>3</sub>	0X <sub>4</sub>	15,000 (45 beds)
Office visits	10X <sub>1</sub>	12 X <sub>2</sub>	0X <sub>3</sub>	20X <sub>4</sub>	100,000
Solution:	750	6992	0	430	
Total Patients					

					Total	Left Over
MD hours	750 x20 <u>15,000</u>	6992 x 0.1 <u>699</u>	0 -	430 x10 <u>4,300</u>	20,000	0
RN hours	750 x80 <u>60,000</u>	6992 x 0.2 <u>1398</u>	0 -	430 x20 <u>8,600</u>	70,000	30,000
Inst. days	750 x20 <u>15,000</u>	6992 x0 <u>0</u>	0	430 x0 <u>0</u>	15,000	0
Office visits	750 x10 <u>7,500</u>	6992 x12 <u>83,904</u>	0	430 x20 <u>8,600</u>	100,000	0

**Accountability and Decision-Making  
in the Iowa Regional Medical Program**

CHARLES W. CALDWELL

My charge is to describe how Planning-Programming-Budgeting concepts are being implemented in an accounting/decision-making system in the Iowa Regional Medical Program. I will note some of the advantages of the system over the more traditional accounting systems and relate some of the problems which we face in our constant effort to remain true to the concepts we are incorporating.

The Iowa Regional Medical Program is a small program, funded at a level of slightly over \$700,000. Our core structure consists of ten professional staff members.

Our system cannot be compared with PPBS structures in large bureaucratic agencies, but it illustrates how certain PPBS concepts can be applied at any functional level.

I can offer no pat formula for evaluating a Program's overall impact on a Region, for establishing priorities, even for determining broad program direction. But I can tell you of a system that does permit the core structure to provide certain objective information to the decision-making process.

It should be emphasized that the system does not make decisions. It merely provides objective information which, in actuality, may be completely ignored by the decision-makers in favor of information that is purely subjective in nature.



Because of our organization's size and due to the numberless intangibles which confront all of us in the health field, we turned to the PPBS approach described by Samuel Greenhouse in an article that appeared in *Public Administration Review*, to guide us in devising our system.<sup>1</sup> His approach is simple and clear. He listed the major structural members of PPBS as: (Slide Number 1)

- Objectives
- Programs
- Program Alternatives
- Outputs
- Progress Measurements
- Input
- Alternative Ways To Do A Job
- Systems Analysis

So that we do not become confused by semantics, I would like to offer a precise definition for each of these terms:

#### SLIDE 1—DEFINITION OF TERMS

##### Objectives must

1. be directly related to overall mission;
2. describe an important end service;
3. be amendable to quantitative measurement;
4. be honest;
5. be broken down into immediate and long-range expectations.

##### Program

A package which encompasses each and every one of an RMP's efforts to achieve a particular objective or set of objectives.

##### Program Alternative

Other possible programs besides those already decided upon.

##### Output

Tangible outgrowth of a particular program.

##### Progress Measurement

Answers the question: How closely does the progress planned for match the progress actually realized?

##### Input

Total quantity of manpower, facilities, equipment and materials applied to a program.

##### Alternative Ways to do a Given Job

Rearrangement of input to an already-existing program in order to improve output.

##### Systems Analysis

Application of cost studies.

##### Objectives

The success of our system stems largely from accurately defining this term. Without doubt, it is the "apex term" in the PPBS idea-structure. These are criteria for judging the validity of an objective within our system:

1. It must be directly related to the overall mission of the IRMP.
2. It must contain a description of an important end-service.
3. It must—at least to the fullest extent possible—be amenable to quantitative measurement. What is not quantifiable has no valid usefulness within the PPBS context.
4. It must be honest. In other words, the stated objective must be identical to the true or real objective.
5. When appropriate, it must be broken down into immediate and long-range expectations.

##### Programs

A program is a package that encompasses each and every one of an RMP's efforts to achieve a particular objective or set of allied objectives. A program could consist of a single comprehensive project or of several projects which have allied objectives. It is confusing that in RMP jargon, the overall effort within a region is called a "program." But for the purposes of our system the term will be used as just defined.

The whole PPBS idea is to facilitate the coordination of all our efforts to meet a particular objective, so the validity of each program may be judged in terms of its overall strategy, dimension and costs. This permits it to be compared with other programs, potential or existing.

In our system no objectives are acceptable unless they suggest a program specifically designed to fulfill them; and no entity can be described as a program unless it is designed to accomplish explicit objectives.

##### Program Alternatives

Program alternatives are programs to the same general end other than those already decided upon. Program

alternatives suggest a choice between two or more programs designed to advance the same overall mission.

### *Output*

An output is a product or a service. As produced by the RMP, it is a tangible outgrowth of a particular program. It must be a kind of service that can be singled out as an indicator of program results. It must be an important end-service and must satisfy an important objective.

### *Progress Measurement*

If output means only those pragmatic end-services that satisfy explicit RMP objectives, then program fulfillment demands an output that was planned and has been produced. Therefore, progress measurement must satisfy one question: Does the progress achieved match the progress anticipated?

### *Input*

Input is the total quantity of manpower, facilities, equipment and materials applied to the program. Like most, we summarize this input in units of dollars.

### *Alternative Ways To Do A Given Job*

This concerns the rearrangement of input invested in an already-existing program to expedite production or upgrade services. In other words, one would rearrange the manpower, facilities, equipment and materials going into a program in order to improve the quality of service or arrive at the stated objective in a shorter period of time. Do not confuse "alternative ways to do a given job" with "program alternatives." Program alternatives are output oriented. Utilization of a program alternative changes the output, because it is a substitute for a whole program and has different specific objectives. Alternative ways to do a given job are input oriented and deal with the best way to achieve an already chosen output or objective.

### *Systems Analysis*

Systems analysis within the IRMP system is primarily the application of cost studies. These studies are of special usefulness in two areas of the system: (1) the determination and evaluation of alternatives and (2) the measurement of costs versus progress within a given program.

These might be called "pure" definitions. As I proceed, you will see how we bend and abuse these definitions within our system.

## THE IOWA SYSTEM

Six major steps have been identified by the IRMP as essential to significant progress toward its overall mission. In aspect, each step is continuous and open-ended, and its influence changes as new information is gathered and updated. (Slide Number 2)

The first of these six steps involves the gathering of morbidity and mortality data and related information that permit us to evaluate the effectiveness of the existing health care system in Iowa.

The second step is the assessment of all existing health resources within the region that fall within the parameters of RMP legislation.

The third step is the identification of needs. From the information provided by steps one and two, an Iowa Regional Health Profile is being developed. It should be emphasized again that this profile is open-ended and will continually change as new information becomes available. On the basis of the existing profile, we endeavor to identify where existing services need to be expanded, coordinated or reinforced to meet the needs identified. We determine where new services need to be initiated and supported.

The fourth step is the establishment of priorities. Conventionally, the criteria considered in the establishment of priorities should include evaluation of need, scientific feasibility, practicality, effectiveness, timing, amount of resources available and community acceptance.

The fifth step is the planning and implementation of programs to meet these priorities.

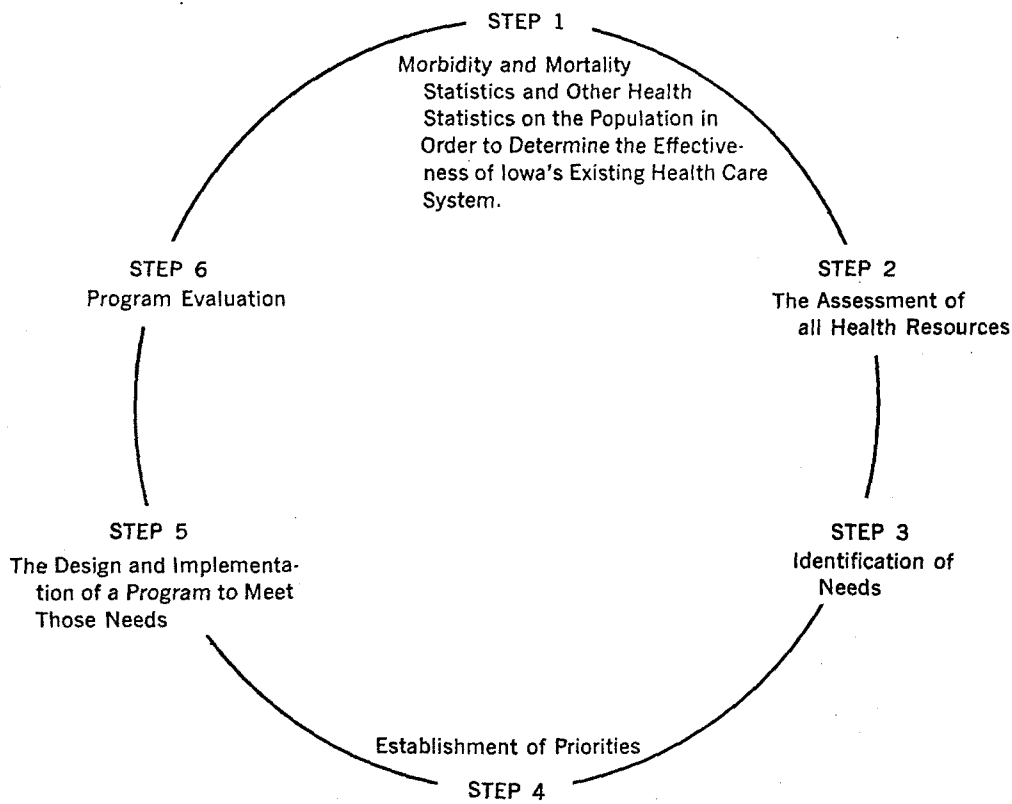
The sixth step is the continuous evaluation of those programs accompanied by modification based upon how well they meet their planned objectives and—insofar as it can be determined—the impact that meeting these objectives is having on an achievement of the overall mission of the organization.

Our accountability/decision-making system involves only steps five and six. Not until we reach step five can we measure precisely how well an objective is being met and consider—if indicated—an alternative program to meet those objectives.

Other than intuitively, we have no way of evaluating the overall impact of the IRMP on the health system in Iowa. A principal reason for this is that presently we have no way of obtaining accurate morbidity data. We hope to solve this problem soon.

Good data on morbidity will certainly aid us in selecting priorities. Such data will not, however, better enable us to evaluate the overall impact of the IRMP—because we will still have no basis for relating changes in

SLIDE 2



morbidity to the existence of our organization. Too many assumptions would have to be made, due to uncontrollable variables.

Now let us examine how the Iowa system can be used in the development and selection of programs to meet priority objectives by looking at the decision-making process from another perspective. (Slide Number 3) Visualize a hierarchy of objectives that relate to different levels of this process.

1. At the top is the organization's overall mission.
2. At the second level we put program objectives designed to meet priority needs.
3. At the third level are groups of project objectives that make-up a program package.
4. At the fourth level we find the objectives of specific activities within a project.
5. The final level is occupied by the day-to-day objectives that are to be met within a project activity.

Our system is applicable at the second level and downward, since it supplies data that grow in objectivity and preciseness as we travel toward the bottom of the

hierarchy. Actually, the system can be applied at any level—so long as we remember that, viewed from the top of the hierarchy, all these levels are means to an end and no ends in themselves.

In Iowa, all staff members contribute to our system—particularly in gathering information, making program evaluations and undertaking cost studies. They also disseminate the resulting information to the decision-makers.

The nuts and bolts of the system can be best recognized by breaking it down into four broad areas of activity:

1. Establishing the costs of program alternatives.
2. Establishing the costs of Alternative ways to do a given job.
3. Accounting and costing of existing programs on a monthly basis.
4. Accounting and evaluation of core activities on a monthly basis.

First, cost estimates are made on all program alternatives. Most of our program alternatives come to us in the form of new project proposals. The cost estimates

SLIDE 3.—Hierarchy of Objectives

OVERALL MISSION

1st Level \_\_\_\_\_ (To improve the availability and level of health care for all persons residing within the Iowa Region without regard to age, color, or economic status, but with special emphasis on heart disease, cancer, stroke and related diseases.) \_\_\_\_\_

2nd Level **A** | **B** | **C** | **D**

**PROGRAM OBJECTIVES**

Example:  
(To improve the availability and quality of care for Iowans threatened by or suffering from Stroke through Projects designed to provide continuing education, demonstrations of care and better availability of care.)

3rd Level **B1** **B2** **B3** **B4** **C1** **C2** **C3** **C4**

**PROJECT OBJECTIVES**

Example:  
(To develop subregional stroke programs that will demonstrate comprehensive patient care (including out-of-hospital care) and provide education to physicians, nurses and other allied professions and information to the public.)

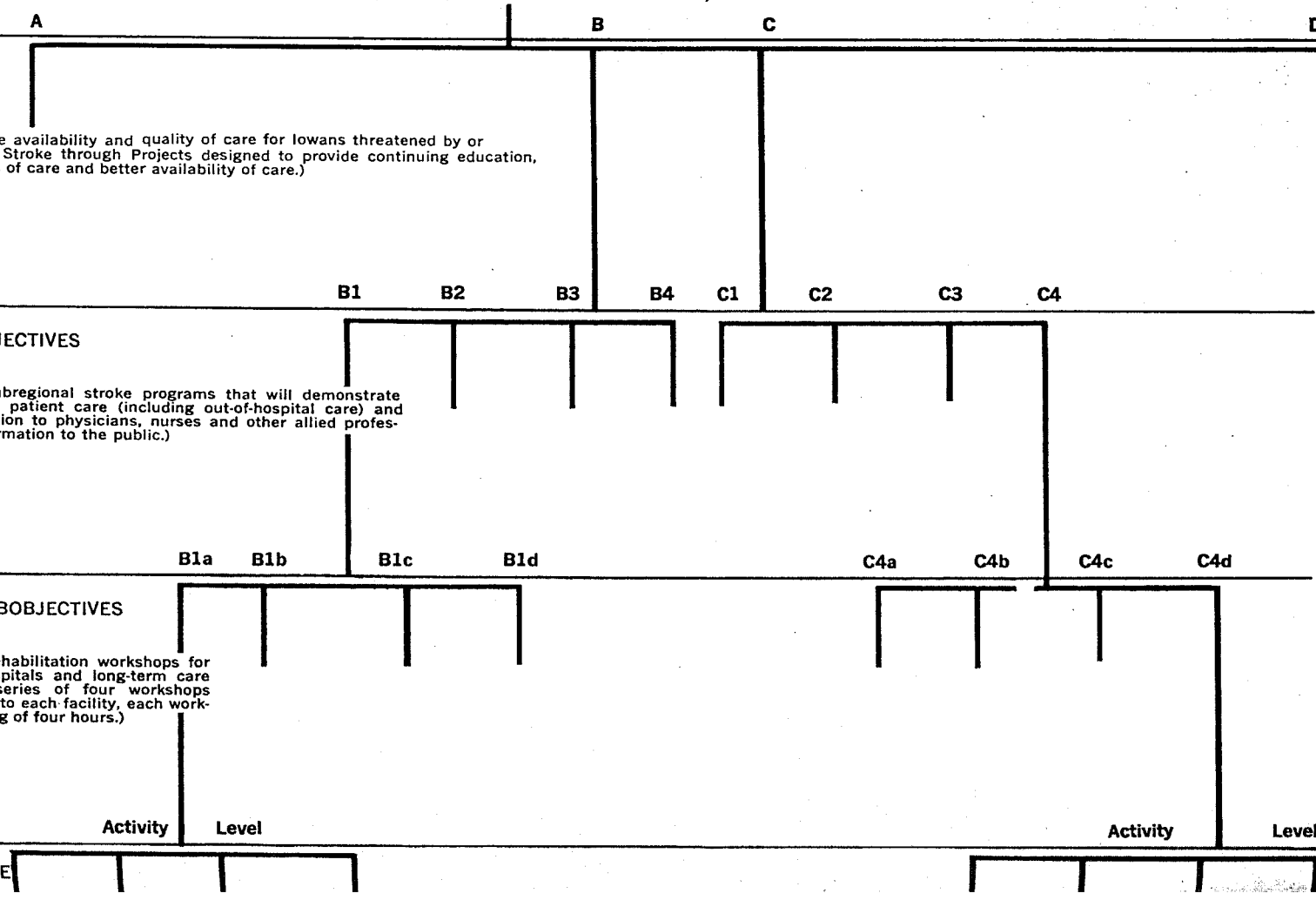
4th Level **B1a** **B1b** **B1c** **B1d** **C4a** **C4b** **C4c** **C4d**

**PROJECT SUBOBJECTIVES**

Example:  
(To provide rehabilitation workshops for nurses in hospitals and long-term care facilities. A series of four workshops are presented to each facility, each workshop consisting of four hours.)

5th Level **Activity** **Level** **Activity** **Level**

**SUBJECTIVE ACTIVITIES**





SLIDE 5—MONTHLY ACCOUNTING BY PROGRAM, PROJECT AND OBJECTIVE

	*	Personnel	Consultants	Equipment	Supplies	Travel	Publications	Other	Total Direct Cost
Program A	*								
Project A1	*								
Objective A1a	*								
Objective A1b	*								
	*								
Project A2	*								
Objective A2a	*								
Objective A2b	*								
Objective A2c	*								
Objective A2d	*								
	*								
Project A3	*								
Objective A3a	*								
	*								
Program B	*								
Project B1	*								
Objective B1a	*								
	*								
Program C	*								
Project C1	*								
Objective C1a	*								
Objective C1b	*								
Objective C1d	*								
	*								
Project C2	*								
Objective C2a	*								
	*								
	*								
TOTAL	*								

At the same time costs are recorded, accurate records of tangible output are maintained, which makes cost analysis an easy task at any time it is needed. Here are examples of how these outputs are reported. (Slide Number 6) (Slide Number 7) There is no uniform method of reporting and these outputs are reduced to different types of units for costing. We probably need more uniformity, but due to the constant changes in many of our programs, any standard form would be obsolete before it was off the press. Each of these reports usually involves several telephone calls to clarify information.

This is an example of the type of cost-analysis report that can be made at any interval and presented to the decision-makers. (Slide Number 8) (Slide Number 9) This particular example includes costs other than those being met by the IRMP and therefore required information not available on a month-to-month basis.

One important evaluation factor that isn't portrayed here is the quality of the output. Output is evaluated for quality in much the same manner that all RMPs carry out evaluation, which includes pre-testing and post testing, attitudinal questionnaires and other techniques. Like all RMPs, we are constantly endeavoring to improve our evaluation methodology.

Of course, it is easy to see that this system isn't comprehensive. Many intangible benefits are unaccounted for. In the presentation of our objective information we attempt to qualify the information, carefully spelling out those probable benefits which are not reflected by tangible output. We cannot ignore these benefits, whether tangible or intangible, form an important part of the analysis.

In the third broad area of activity, we are truly bending—if not breaking—the conceptual rules of PPBS because we are accounting for core activities that it

most cases cannot be related to an end-service. Here is how it works: (Slide Number 10)

Daily time studies are made based on a breakdown of core activities into the functional activities shown. Each of these activities produces a measureable output that does not relate to any precise objective in many cases. The time contributed to operational projects as depicted in the first nine columns can be related to project output and figured into the costs of operating the projects.

The column entitled "other" is for those core staff functions that can be related to an end-service other than that of an operational project. For example, we have a central medical library network that receives a limited amount of attention from core staff members.

The last four columns, entitled "Project Planning," "Data Collection," "Public Information," and "Staff Education," are strictly functions and do not relate to an end-service. However, we have arbitrarily identified tangible output as a gauge to evaluate our core activity.

For example, we can compare the amount of time invested in new project development, which would fall under "Project Planning," with the number of new proposals submitted to our decision-makers. We can compare the amount of time we are spending on a given operational project with that project's output.

SLIDE 6—NURSING WORKSHOPS  
(July 1, 1969 - January 31, 1970)

<i>Capital Division</i>		
<i>Type of workshop</i>	<i>Number of workshops</i>	<i>Attendance</i>
I . . . . .	15	295
II . . . . .	15	301
III . . . . .	11	214
IV . . . . .	<u>9</u>	<u>161</u>
TOTAL . . . . .	50	971
<i>North Central Division</i>		
I . . . . .	16	371
II . . . . .	14	395
III . . . . .	10	199
IV . . . . .	<u>9</u>	<u>151</u>
TOTAL . . . . .	49	1,116
<i>Northwest Division</i>		
I . . . . .	33	810
II . . . . .	26	617
III . . . . .	35	725
IV . . . . .	<u>35</u>	<u>643</u>
TOTAL . . . . .	129	2795

Each workshop is three hours in duration.

SLIDE 7—NURSING EDUCATION CONFERENCES  
(July 1, 1969 - January 31, 1970)

<i>Location</i>	<i>Number of Days</i>	<i>Attendance</i>
Des Moines . . . . .	2	110
Mason City . . . . .	<u>1</u>	<u>25</u>
TOTAL	3	135

HOME SERVICE CONSULTATION  
(July 1, 1969 - January 31, 1970)

July - August, 1969 . . . . .	139 visits	49 patients
September . . . . .	42 visits	23 patients
October . . . . .	28 visits	25 patients
November . . . . .	23 visits	19 patients
December . . . . .	41 visits	35 patients
January, 1970 . . . . .	<u>58 visits</u>	<u>37 patients</u>
TOTAL . . . . .	331 visits	188 patients

Average patient load: 72

STROKE UNIT  
(July 1, 1969 - January 31, 1970)

Total number of patients admitted: 81  
Average patient stay in stroke unit: 12.5 days

SLIDE 7--(Continued)

PUBLIC EDUCATION

	Conferences	Attendance
Northwest Division . . . . .	7	122
Capital Division . . . . .	10	347
North Central Division . . . . .	26	576
TOTAL	44	1,045

Conferences averaged one hour each.

SLIDE 8--COST ANALYSIS FINDINGS

*Nursing Workshops*

1. The cost to the RMP for each nurse who was a student in the workshop was \$1.88 per hour (Student Hours).
2. The combined cost to the RMP and the Heart Association (16 cents per student hour) was \$2.04.
3. The total cost to the RMP, the Heart Association and Heart volunteers (an added 39 cents per hour) was \$2.43.
4. The cost of instruction to the RMP for the workshops was \$40.31 per hour (Instructor Hours).
5. The cost to the RMP and the Heart Association per instructor-hour was \$43.60.
6. The cost to the RMP, the Heart Association and Heart volunteers per instructor-hour was \$52.07.

*Nursing Education Conferences*

1. The cost to the RMP per student-hour was \$3.77
2. The cost to the RMP and the Heart Association per student-hour was \$4.08
3. The cost to the RMP, the Heart Association and Heart volunteers was \$6.15
4. The cost to the RMP per instructor-hour was \$307.70
5. The cost to the RMP and the Heart Association per instructor-hour was \$333.30
6. The cost to the RMP, the Heart Association and Heart volunteers was \$502.66

*Home Service Consultation*

1. RMP cost per visit made to a patient was \$20.86
2. RMP and Heart Association cost per visit was \$22.56
3. RMP, Heart Association and Volunteer cost per visit was \$24.15
4. RMP cost per patient in the program was \$83.44
5. RMP and Heart Association cost per patient was \$90.34
6. RMP, Heart Association and Volunteer cost per patient was \$96.60

SLIDE 9

*Stroke Unit*

1. RMP cost per patient admitted to the stroke unit was \$197.24
2. RMP and Heart Association cost per patient was \$213.07
3. RMP, Heart Association and Volunteer cost per patient was \$306.04
4. RMP cost per patient day in the stroke unit was \$15.77
5. RMP and Heart Association cost per patient day was \$16.24
6. RMP, Heart Association and Volunteer cost per patient day was \$24.48

*Public Education*

1. RMP cost for each individual attending conferences was \$3.71 per hour (Student hour).
2. RMP and Heart Association cost per student-hour was \$4.06
3. RMP, Heart Association and Volunteer cost per student-hour was \$4.90
4. RMP cost per instructor was \$91.64
5. RMP and Heart Association cost per instructor-hour was \$98.79
6. RMP, Heart Association and Volunteer cost per instructor-hour was \$119.11.



There has been no attempt to make a judgment as to whether or not these costs are reasonable. The value of this type of analysis is greatly enhanced when the unit costs can be compared to the same unit costs in similar projects. This information should be shared with other Regional Medical Programs and Heart Associations that are conducting similar projects in the hope that they, in turn, will be able to share similar information with us.

The one classification that can be compared internally between the sub-regions of the Stroke Management Project is Nursing Workshops, since all three sub-regions of the project have identical programs. The results of these internal comparisons, based on RMP costs only, are depicted below.

	Northwest Division	Capital Division	North Central Division
per student-Hour	\$ 1.18	\$ 2.70	\$ 2.24
per Instructor-Hour	\$25.49	\$52.40	\$ 50.97



This information places us in a better position to determine how and in what areas we should be spending our time. Except as it relates to operational projects, this information is not reported regularly to our decision-makers. It is presented at regular staff meetings, where we jointly evaluate how usefully our time is spent and establish work priorities.

Following a true PPBS structure, the entire expense of the core activity would be assigned to project output. In truth, within the core structure we are not evaluating our success on the basis of end-product. We are evaluating means, not ends. However, we are supported by one school of thought which believes that indirect activities should be allocated to a program only when such an allocation would contribute to a better decision.

In summation, the system permits us to:

1. Undertake better cost-accounting for individual projects.
2. Obtain more efficient use of scarce manpower, including staff time.
3. Provide more accurate cost estimates to our decision-makers.

What I have described here is only a start on the construction of a system designed to support our decision-making process with objective information. I believe the system has influenced decisions as those decisions are concerned with alternative ways to do a given job. In all honesty, I can see very little influence on decisions that relate to program alternatives, possibly because we haven't considered that many program alternatives since the IRMP became operational. I think it may have more influence at the end of the current three-year funding period when political influences and obligations will be greatly lessened.

The system is faced with many problems. We need in-depth cost-benefit studies which will carry all the way down to the consumer and will take into account the many economic variables that affect health care. We need to develop better ways to present our information to the volunteer decision-makers.

Presently, we have neither the resources nor the expertise to deal with social costs. Comprehensive costing should also include estimates of cost that are related to changes in other human systems as a result of decisions we make.

We must continue to search for better and more comprehensive ways to quantify services. It is to be remembered, however, that we are primarily a service organization and therefore must be conscious that there is a point of diminishing returns.

Our cost studies on projects would be more valuable if we had cost studies from other regions with which to compare them. Because not everyone is willing to play under our rules, we sometimes feel like the only honest guy in a crooked crap game.

We need a national review and evaluation system that is more consistent in both scheduling and methodology. For example, we have had four fiscal years assigned to us in three years' time.

We need to be permitted to set our own priorities. Presently, while we are setting our own priorities we must try to second-guess what is currently popular in Washington.

Finally, in my opinion PPBS is not a set of techniques so much as it is a set of attitudes. Unless one is really interested in getting the most for the tax dollar, it will not work. Old concepts such as the "budget is a political tool," "the harboring of privileged information," or the "measure of an organization's success by the size of its budget" are concepts which are not compatible with PPBS concepts.

The purpose of PPBS is to bring together the budgeting process with the decision-making process, evaluating both processes on the basis of tangible output. Its intent is to make and keep us mission oriented since we will be ultimately judged on how well we accomplish our mission.

#### FOOTNOTES

1. Samuel N. Greenhouse, "The Planning-Programming-Budgeting System: Rationale, Language, and Idea-Relationships," *Public Administration Review*, XXVI, No. 4 (December, 1966), p. 273.

#### Resource Allocation and the Evaluation Process

CHARLES L. JOINER

#### ECONOMICS, SOCIAL PRODUCTION FUNCTIONS, AND RESOURCE ALLOCATION

##### *Economics*

Economics is the science of allocating scarce resources among alternative uses so as to attain the greatest or maximum fulfillment of society's unlimited wants, i.e., "doing the best with what we have."

##### *Optimum Allocation of Resources*

Classical economics assumes the "rational man" concept. Therefore, if the decision maker then wishes to combine resources to minimize the costs of producing a given level of output; if he knows the resources (inputs)

that can be used in producing the output, and if he also knows the prices for increasing each input (and the increase in output that will result from each input entry), then the way to achieve minimum costs is as follows: the decision maker should use those resources in such a combination that the additional increment in output per dollar spent on each input is equal.

The allocation of resources under the assumptions of classical economics is assumed to be optimized because of the competitive nature of the system itself. Unlike the classical model, many social action programs, including health, involve the allocation of relatively scarce public resources. In addition, there is the need of properly meshing these public funds with private resources for maximum effectiveness for improving or maintaining health. Needless to say, any model constructed for the allocation of resources for better health will have its shortcomings, e.g., the allocation of resources for health means fewer resources available for non-health purposes.

If one considers the health sector as a system of itself, optimum resource allocation requires that the additional benefit rising from the allocation of an additional expenditure (cost) for a particular health problem must be equal to ratios of benefits to costs for other health problems. For a theoretical explanation, additional benefits and costs may be referred to as marginal benefits and costs. Therefore, the optimum allocation of resources toward the solution of various health problems is accomplished when:

$$\frac{MB_a}{MC_a} = \frac{MB_b}{MC_b} = \frac{MB_c}{MC_c} \dots \frac{MB_n}{MC_n}$$

where: MB equals marginal benefits accruing from the implementation of a particular technique or approach for solving the health problem within a series of health problems, a, b, c, ... n.

MC equals the marginal costs resulting from the implementation of a particular technique or approach for solving the health problem within a series of health problems, a, b, c, ... n.

This marginal benefit-cost approach for optimum allocation of resources for the solution of various health problems may also be applied to the allocation of resources among alternative strategies or approaches for the solution of any given health problem. In fact, this benefit-cost approach should be an inherent part of any normative decision making process. However, the application of such a theoretical approach becomes extremely difficult when the decision maker does not know or can not determine precisely the benefits or outputs of a particular technique or approach to the solution of a health problem. It is for this reason that

this paper now turns to the question of social production functions in relation to the political decision process and such problem-solving approaches as PPBS.

#### *Social Production Functions and the Decision Process*

Before one is completely enthralled with the idea of the determination of social production functions and the role of benefit-cost analysis in the allocation of scarce resources, some reflections on the realistic political decision process are necessary. Charles Lindblom<sup>1</sup> has quite adequately described the real political decision process which in some ways appears to be distinctly different from the problem solving approach of PPB. Lindblom states as a first rule of the successful political process, "don't force a specification of goals or ends." The reasoning here is that not only is the specification of objectives intellectually difficult, but also pragmatically harmful. In fact, it could mean that agreement among diverse interests on specific measures may be completely blocked.

For example, the Elementary and Secondary Education Act of 1965, which is considered a landmark piece of legislation in terms of federal aid to education, needed the support of at least three divergent interest groups. The parochial schools saw it as a step in providing financial assistance for parochial school children. A second group saw it as an anti-poverty measure, since the distribution of funds for Title I of the bill was based on the number of poor children in each school district. A third group saw it as a broad beginning of a large program of federal aid to public education. It does seem quite possible that the bill would have been defeated had any attempt been made to secure strong agreement on long-run objectives.

A second major feature of a desirable decision process as seen by Lindblom is its incremental characteristic. The process toward objective attainment should proceed in very small steps because of our inability to foresee the full social consequences of any program and the fact that political decision costs tend to increase as the decisions conflict with values held by interest groups.

The third major element in the Lindblom approach is referred to as the "advocacy" process of reaching decisions. To the extent that advocates of every related interest have a voice in policy making, the self interest motivation will insure that each advocate takes the responsibility for researching the consequences of any action for the value he represents. Obviously, this approach is not idealistic. Instead, it is pragmatic, stresses

<sup>1</sup>Charles Lindblom, "The Science of Muddling Through," *Public Administration Review*, Vol. 19, No. 2 (spring 1959), pp. 79-88.

process rather than substantive criteria. Therefore, by definition, a "good" decision is one which obtains consensus rather than one which meets the requirement of efficiency or effectiveness.

In order to properly relate political values to analytical program decisions involving the allocation of resources, the decision process must include some determination of the social production functions that translate program specifications (input) into program consequences (output). An analogy may be drawn here to consumer preference theory. Economic factors of production—land, labor, capital, and management—are not directly evaluated in terms of consumer preference functions, but only through a process which translates these inputs into outputs. It is the output, or final product, that enters directly into consumer preferences. The process of translating inputs into outputs, of course, assumes knowledge of the production functions involved.

If the analogy is applicable, we need to know the social production functions of health programs. It is at this point that the task of the social scientist becomes more difficult because many of the social action programs of the federal government do not deal with the simple translation of factors of production into commodities, but the production functions are determined largely by institutional or behavioral characteristics.<sup>2</sup> Determination of social production functions involves complicated systems in which institutional, technical and economic factors interact with each other. Therefore, we cannot expect the technical expert to define all of the input-output relationships, i.e., relying totally on physicians to evaluate all health programs or engineers to implement the design of pollution control systems.

It seems imperative that the analysis of production functions in most public programs must take a systematic approach rather than being confined to technical considerations. Many times it is extremely difficult to predict with any real degree of certainty the specific performance of new or proposed social programs. Some of this uncertainty concerning the relationship between inputs and outputs can be reduced via either ex-post evaluation of operating programs or the implementation and evaluation of demonstration projects. Although the process of decision making described by Lindblom of incremental changes has been recognized as an effective means of proceeding under uncertainty, this does not reduce the need for systematic analysis. In some

<sup>2</sup>Charles L. Schultz, *The Politics and Economics of Public Spending*, (Washington, D.C.: The Brookings Institution), pp. 55-76.

instances, evaluation must involve in-depth studies using sophisticated statistical techniques—particularly when the impact of one program is only a part of a much larger program. Feedback of results from operating programs is an absolute essential to program planning, and systematic analysis provides the necessary feedback for decision making and planning.

#### INTRODUCTION OF CONCEPTUAL AND ACTUAL PROJECT EVALUATION PROCESS IN RELATION TO PPBS

It is commonplace to wade through an article on evaluation and find it is like the last ten you read. The mass of articles on evaluation emphasize the necessity for evaluation and they generally state that a conceptual evaluation model should be designed. These evaluation articles stop at this point. I plan to go beyond where others stop and speak to you on a conceptual model designed and tested at ARMP.

In June, ARMP instituted a systematic and in-depth evaluation of all approved projects. This was a first step in total program evaluation and an experiment with PPBS.

The majority of core staff at ARMP were skeptical about PPBS. It was decided that the first two aims of the PPB system should be used to evaluate ARMP projects. These aims are:

1. the careful evaluation and examination of goals and objectives in each major area of activity, and then to
2. analyze the output of a given program (project) in terms of its objectives.

An evaluation model was designed and used with four projects this past summer. The model was found to be adequate for use on these projects. The experience acquired by developing an evaluation model along with the actual evaluation process has led to a more knowledgeable understanding of problems associated with analytic investigation as well as giving an indication of problems linked with PPBS.

#### *Divisions of the Model and Experiences Gained*

##### Project Development

Assumptions: 1. When projects are developed, the alternatives, if known, are brought out and discussed.

2. The project goals and objectives meet program goals and objectives.

Step 1. Determination of the project goals—This first step consists of determining in rather broad and long-range terms what is to be achieved by the project. A

statement of project goals is necessarily broad and frequently long-range, and, for these reasons, a project's goals may not be capable of direct measurement in the short-run. One problem encountered in the evaluation process was that several of the projects did not have realistic goals.

Step 2. Determination and statement of project objectives—Project objectives, as used in this evaluation, are narrow and short-range statements of what the project is to accomplish. Project objectives are derived from and must be compatible and consistent with the project goals. The difficulty encountered here was that often the project objectives were vague (e.g., increase patient care) and had to be rewritten in measurable terms.

Comment: These problem areas have been corrected. Realistic goals and measurable objectives are a part of all new projects. The evaluation process actually begins during this stage of project development. All goals and objectives are being challenged by the evaluation coordinator to make sure they are feasible and applicable to total program goals and objectives.

#### Pre-Evaluation Process

Step 3. Determination of measures of objective attainment—these measures would include, for example, such things as: days, hours, dollars, ratings, ratios, percentages, attitude changes, and patient behavior. Repeatedly, it was found that project directors of funded projects did not know what data to keep and how to record collected data so as to justify the project. There were several reasons for this, one being poorly written project objectives.

Step 4. Establishment of standards—standards, as used in this evaluation, refer to desired levels of attainment. Only through the use of implicit or explicit statements of acceptable and/or unacceptable standards can the administrator decide whether to continue, adjust, or discontinue a particular project. Standards frequently were not written into the projects. This has led to a poor percentage of approved projects for ARMP at the national level (27%). The lack of standards has also made projects difficult to evaluate.

Comment: The problems in steps 3 and 4 are being corrected by a pre-evaluation process. Before any project is written, measures for objectives are agreed upon by all people concerned. During the pre-evaluation process, standards are established. Alternatives to the project are further discussed.

#### Actual Evaluation Begins

Step 5. Collection of performance data—once the desired level of action is decided, the relevant data which will permit the determination of the actual level of performance must be collected. Collection of evaluation data should be an integral part of the on-going project implementation. If steps 1 through 4 are complied with as described above, then actual evaluation can easily be accomplished. It is a matter of inserting data into the proper place. Output studies are important and the type study (e.g., cost-benefit analysis) should be determined when the project begins so that adequate data are available.

Step 6. Comparison of actual performance with standards previously set—This is considered the program (project) effectiveness step. Programs may differ in their effectiveness depending on the extent to which pre-established objectives are attained as a result of activity. Based upon a comparison of actual performance with the standards, the performance will be concluded to have been satisfactory or unsatisfactory. After a determination of satisfactory or unsatisfactory performance has been made, the project administrator has a number of alternatives available to him. If the performance is concluded to be satisfactory, the project may be continued unaltered, or, if the goals and objectives have been met, the project can be satisfactorily concluded. If the performance is determined to be unsatisfactory, the administrator may modify his project objectives and/or standards (objectives or standards are unrealistic), attempt to improve efficiency (inefficient use of resources), or recommend discontinuance.

Comment: It is felt that a seventh step is required between step 6 and the final recommendation. This would be a step for feedback between the evaluator(s) and members of the program (project). Honest communications should take place between the evaluator(s) and the project staff so that apparent results can be discussed. If discrepancies are discovered during these discussions, further study can be made. The evaluator(s) and project members should agree on the results, whether satisfactory or unsatisfactory.

#### Summary and Conclusions

Economics is the science of allocating scarce resources among alternative uses so as to attain the greatest or maximum fulfillment of society's unlimited wants, i.e., "doing the best with what we have."

If one considers the health sector as a system of itself, optimum resource allocation requires that the additional

benefit rising from the allocation of an additional expenditure (cost) for a particular health problem must be equal to ratios of benefits to costs for other health problems.

This marginal benefit-cost approach for optimum allocation of resources for the solution of various health problems may also be applied to the allocation of resources among alternative strategies or approaches for the solution of any given health problem. However, the application of such a theoretical approach becomes extremely difficult when the decision maker does not know or can not determine precisely the benefits or outputs of a particular technique or approach to the solution of a health problem.

In order to properly relate political values to analytical program decisions involving the allocation of resources, the decision process must include some determination of the social production functions that translate program specifications (input) into program consequences (output).

Determination of social production functions involves complicated systems in which institutional, technical and economic factors interact with each other.

The second part of this paper speaks to a conceptual model designed and tested at the Alabama Regional Medical Program. The model was found to be adequate after it was used to evaluate four projects during the summer of 1970.

Divisions of the model are:

#### Project Development

Step 1. Determination of the project goals.

Step 2. Determination and statement of project objectives.

#### Pre-Evaluation Process

Step 3. Determination of measures of objective attainment.

#### Step 4. Establishment of standards

#### Beginning of Actual Evaluation

Step 5. Collection of performance data.

Step 6. Comparison of actual performance with standards previously set.

After a small-scale testing of the first two aims of PPBS, ARMP reported the following benefits:

1. Improved project development.
2. Increased control of funded projects.
3. A better appreciation and understanding of the value of evaluation.
4. An acceptance by the staff that the total program should be evaluated, probably using the PPBS method
5. Development of a more sophisticated decision-making mechanism.

In November, ARMP will continue to experiment with PPBS and will further evaluate its effectiveness. At the present time, however, ARMP is working on other priorities—some of which were determined by the evaluation process described in this paper.

*EDITORS NOTE: Two Appendices to Dr. Joiner's paper are not reprinted in the Proceedings. They are:*

1. *Medical Information System via Telephone (M.I.S.T.) Evaluation Report.*
2. *Reality Orientation Technique Evaluation Report.*

*Both were prepared for the Alabama Regional Medical Program by Edward M. Smith, Ph.D., Research Associate, Bureau of Research and Community Service, School of Health Services Administration, University of Alabama in Birmingham and Douglas Patterson, MHA, Evaluation Coordinator, Alabama Regional Medical Program.*

# SPECIAL INTEREST MEETINGS STATISTICAL MODELS AND OPERATIONS RESEARCH

## Participants

Francis C. Ichniowski—Moderator  
Acting Chief, Systems Management  
Regional Medical Programs Service

David H. Gustafson, Ph.D.  
Assistant Professor  
Industrial Engineering Division  
University of Wisconsin

Vernon E. Weckwerth, Ph.D.  
Director, Systems Development Project  
University of Minnesota  
Member, Regional Advisory Group  
Northlands Regional Medical Program

### A "Weighted Aggregate" Approach To R&D Project Selection

DAVID H. GUSTAFSON, GOPINATH K. PAI,  
GARY C. KRAMER

#### Introduction

There appear to be few formal decision theory procedures for optimally allocating funds among potential projects. One reason for this is the lack of effective methods for assigning a value to each alternate project. With a few notable exceptions<sup>2,3,4</sup> previous project evaluation systems have been either theoretical efforts requiring many modifications before being practical or methodologies lacking the scientific rigor to assure reliability or validity.

Two excellent articles<sup>5,6</sup> have reviewed the research up to 1967 so their efforts will not be duplicated here. Since then, J.R. Miller<sup>3</sup> has suggested some interesting but relatively untested procedures for evaluating alternative projects using an additive model where the criteria are weighted according to importance.

L.P. Hellman<sup>7</sup> has evaluated a value measure for selecting proposals for research grant support. The model he used is based on the Churchman-Ackoff<sup>8</sup> approximate measure of value, modified to satisfy the needs of the National Institutes of Health. The evaluation of each proposal was based on the relative values of the objectives of the funding agencies, the relevance of the proposal's objectives and the probability of success of the proposal's objectives. Proposals with high overall expected values were selected for funding; this model appeared to be superior to the previous method of proposal selection.

Abernathy and Rosenbloom<sup>9</sup> have discussed the pros and cons of parallel and sequential project selection

strategies. A parallel strategy involves simultaneously taking two or more approaches to solving the same problem. In a sequential strategy the best approach is pursued; other possibilities being considered only if the first approach proves unsuccessful. The authors have incorporated the incremental cost of adopting a parallel strategy, the probability of success of each strategy and the cost of failure in a normative mathematical model which selects which strategy to use.

This paper will (1) describe a general project evaluation model, (2) discuss problems with current approaches to implementing the model, (3) propose methodologies to solve these problems, (4) report on the evaluation of some of these methodologies, and (5) suggest areas for further research.

#### The General Model

Complex evaluation problems generally possess five characteristics. First, there are several criteria which are important in evaluating the merits of the projects. Second, the relative importances of these criteria vary from one judge to another. Third, the extents to which these criteria are satisfied are not always directly measurable on an interval scale. Fifth, the criteria are sometimes interdependent.

Recognizing that the overall evaluation is some aggregate of the valuations of individual criteria, we write

$$E = \sum_{i=1}^n W_i \beta (X_i) + \sum_{j=n+1}^{n+m} W_j R_j \quad (1)$$

The  $i$  subscripts are associated with quantitative variables and the  $j$  subscripts are with qualitative variables.  $W_i$  represents the relative weight of the  $i^{\text{th}}$  criterion and  $\beta$



$(X_i)$  represents the utility function associated with the  $i^{\text{th}}$  criterion.  $X_i$  represents the extent to which the  $i^{\text{th}}$  variable is present and  $R_j$  represents the extent to which the  $i^{\text{th}}$  criterion is satisfied. All criteria as well as projects are assumed to be independent.

In order to implement such a model we must (1) select project evaluation criteria, (2) assure independence, (3) establish the relative importance of the criteria, (4) develop scales with which to quantify or categorize the variable, (5) determine for quantitative variables the utility function associated with each criterion, and (6) aggregate the evaluations of all judges.

Such a model has two uses. First, it can be used as an aid in the proposal evaluation process. Technicians can use the model to estimate the relative value of each proposal and report the results to the committee as additional information for their decision making process. Second, it could be used as a guide to proposal modification. The model could predict what decisions would be made by the committee. The proposer could then improve the proposal where necessary. By knowing  $W_i$ ,  $\beta(X_i)$ ,  $R_j$ , and the cost of increasing  $X_i$  or  $R_j$  by one unit, he could select the criteria to give the greatest increase in value for the least cost.

### Criteria Weighting

A criterion's relative importance (weight) should be directly proportional to its impact on the decision making process. Because weights define organizational needs, a set of concisely defined and properly weighted criteria can guide proposers to develop programs to meet those needs. Those who lack this guidance may propose programs of little interest, become discouraged with the process, and be lost as a resource to the organization.

From the proposal evaluator's point of view, criteria weights permit him to more accurately and consistently model the committee's project evaluation philosophy. Proposals are frequently too detailed or numerous to be evaluated by the whole decision making committee so they are normally reviewed by a subset of members and staff. Unless each evaluator knows the relative importance of each criterion, their evaluations will lack consistency.

Some project selection techniques assume that all criteria have equal weight in the decision making process. The success of this approach is directly proportional to the degree to which this assumption is true. Other models estimate weights by using an empirical technique such as multiple regression.<sup>10</sup> The committee rates hypothetical projects described in terms of the criteria. Coefficients are estimated, using the method of least squares, so as to best predict committee decisions. There are two problems with this approach. First, it is difficult to obtain enough data (and therefore degrees of freedom) to yield valid, reliable coefficient estimates.

Second, the regression approach will not improve committee decisions, only predict them, because this method is based on decisions that *were* made by the committee, rather than decisions that *should* have been made. Man is progressively less accurate in evaluating complex problems as the number of criteria influencing his decision increases<sup>11,12</sup>. Hence, the regression approach, as a normative model, breaks down when the number of criteria are large. The decision makers become "cognitively overloaded" and the decisions made may not be the ones they would like to make.

We evaluated a third set of criteria weighting methods where weights are estimated by the committee members. There is evidence<sup>11,12,13</sup> to indicate that under certain conditions, men do this quite effectively. Miller<sup>3</sup> suggests a hierarchical approach to criteria weighting.

### Example

Assume that a list of criteria have been developed in a hierarchical form (Figure 1). All criteria in one column that are connected by lines are *related* in that they are components of one larger criterion in the left, adjacent column. We will refer to each column as a "level". Decision makers are asked to: (1) rank, in order of importance, the *related* criterion in a given level, (2) assign a value of 100 to the most important and values between 0 and 100 to the others so as to reflect relative criteria importance. These weights are normalized and then successively multiplied by weights of *related* criteria at each higher level. In Figure 2, vertical lines represent criteria and horizontal lines connect related criteria. Suppose the first level criteria were ranked II, III, and I and weights of 100, 60, and 40 were assigned. Weights assigned within criteria sets  $(B_1 B_2 B_3 B_4)$ ,  $(C, D)$ ,  $(D_1, D_2)$  are shown in Figure 2a. Next, weights were normalized by dividing each weight by the sum of all weights within a set. The final weight of each lowest level criterion is the product of the normalized weights of itself and the connected criterion at each of the higher levels. Thus, the final weight of criteria  $D_1$  is the product of weights assigned to criteria  $D_1$ ,  $D_0$ , and  $\epsilon$ , in Figure 2b.

While this approach reduces the number of criteria being considered at once, it replaces one bias (assessment error due to cognitive limitations) with another (aggregation error due to multiplication of errors occurring at each level of the hierarchy). As the number of levels increases, the second type of error becomes important.

We compared this approach with a modification (the "ratio method") that appears to reduce both aggregation and assessment errors:

1. Rank the criteria in order of importance.

FIGURE 1.—Project Evaluation Criteria Displayed in a Hierarchical Fashion.

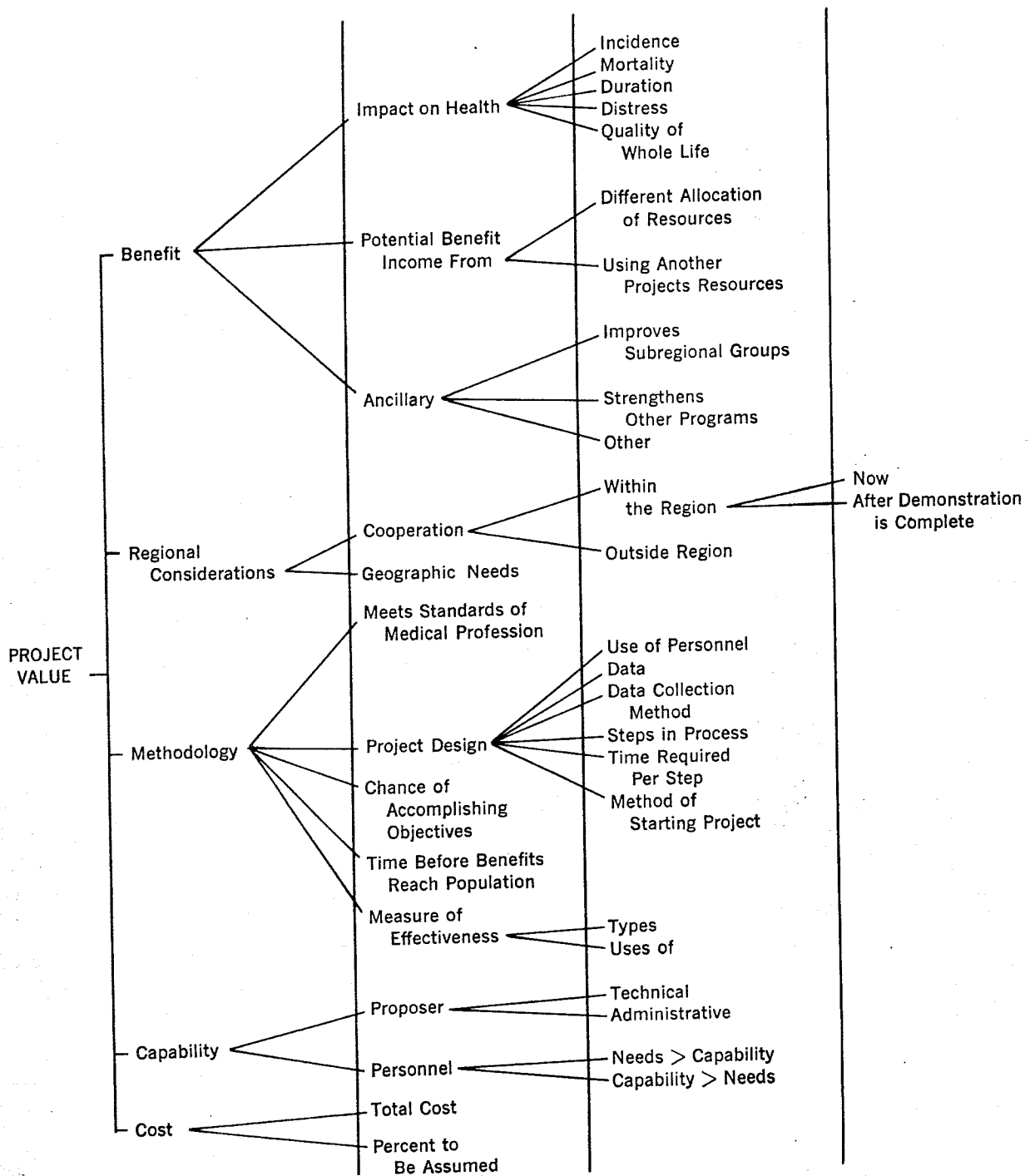
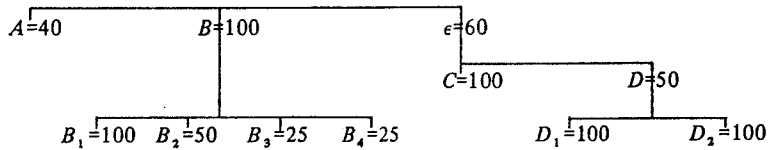
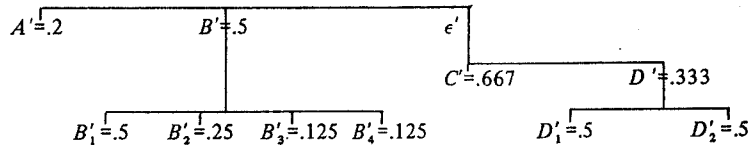


FIGURE 2.—*Demonstration of Hierarchical Method for Criteria Weighting*

2a—Criteria pyramid including criteria weights.



2b—Criteria weights normalized within subsets.



2c—Criteria weights for lowest level criteria.

Criteria	Products	Weight
A'	(.2)	.2000
B <sub>1</sub> '	(.5) (.5)	.2500
B <sub>2</sub> '	(.5) (.25)	.1250
B <sub>3</sub> '	(.5) (.125)	.0625
B <sub>4</sub> '	(.5) (.125)	.0625
C'	(.3) (.667)	.2000
D <sub>1</sub> '	(.5) (.333) (.3)	.0500
D <sub>2</sub> '	(.5) (.333) (.3)	.0500
		<u>Total 1.0000</u>

- Compare the most important criteria with every other related criteria. Estimate how many times more important the top ranked criterion is than each of the other criterion.
- Repeat steps 1 and 2 for a new set of criteria composed of the most important criteria from each set.
- Multiply the weights assigned to criteria in step 2 by those assigned, in step 3, to the top ranked criteria from its set.

*Example*

Suppose for the criteria in 2a ratio weights are assigned to each set as shown in Figure 3a. The new criteria set (A, B<sub>2</sub>, C, D<sub>2</sub>) are ranked (3,1,2,4) and as signed weights of (1:1.5,1:1,1:1.25,1:3). The weights in Figure 3c are obtained by multiplying the weights in sets A, B, C, and D by values of 1/1.5, 1, 1/1.25, and 1/3 respectively. The final normalized weights are obtained by dividing each weight in Figure 3c by the sum of all the weights in Figure 3c. The normalized weights are given in Figure 3d.

Two factors may have caused the superior performance of the ratio method. First, the hierarchical method may yield higher errors because the errors are multiplied rather than added. Second, the ratio method uses an odds estimation methodology while the hierarchical method uses ratings on a 0 to 100 scale. Previous research<sup>11</sup> indicates that odds estimation leads to more accurate estimates of subjective probabilities. Possibly the results extend to criteria weighting.

*Criteria Independence*

Two criteria are dependent when (1) the extent to which one criterion is satisfied is influenced by the extent to which another criterion is satisfied and (2) the utility associated with a given level of satisfaction on one criterion is influenced by the degree to which another criterion is satisfied. When the assumption of criteria independence, postulated in equation 1, does not hold, total project value is no longer equal to the sum of the values associated with the individual criteria.

FIGURE 3. *Ratio Method of Criteria Weighting.*

3a. *Ratio values assigned to criteria in Figure 2a.*

$A = 1:1$	$B_1 = 1:1$	$C = 1$	$D_1 = 1:1$
	$B_2 = 1:2$		$D_2 = 1:1$
	$B_3 = 1:4$		
	$B_4 = 1:4$		

3b. *Ratio weights assigned to new criteria set.*

$A = 1:1.5$	$B_1 = 1:1$	$C = 1:1.25$	$D_1 = 1:3$
-------------	-------------	--------------	-------------

3c. *Ratio weights of all criteria.*

$A = 1:1.5$	$B_1 = 1:1$	$C = 1:1.25$	$D_1 = 1:3$
	$B_2 = 1:2$		$D_2 = 1:3$
	$B_3 = 1:4$		
	$B_4 = 1:4$		

3d. *Normalized ratio weights.*

$A = 0.16$	$B_1 = 0.24$	$C = .20$	$D_1 = 0.08$
	$B_2 = 0.12$		$D_2 = 0.08$
	$B_3 = 0.06$		
	$B_4 = 0.06$		

An additive model with interaction terms may compensate for criteria dependence if enough degrees of freedom can be obtained to accurately estimate coefficients empirically. However, if coefficients must be subjectively estimated, the multi-dimensionality of the interaction term would increase both the number and difficulty of the estimates. While we have very little information about the performance characteristics of judges in weighting multidimensional criteria, we may draw some insights from research into subjective probability estimation.<sup>17</sup> Several researchers<sup>11,12</sup> have shown that men are conservative probability\* estimators and that this conservatism increased with the number of data to be simultaneously considered. Future research should determine (1) if the same problem exists in utility assessment and criteria weighting and (2) the best methods for obtaining these estimates. Until then, the criteria independence problem will have to be treated in some other way.

\*Conservative estimators overestimate the importance of diagnostic data and underestimate the importance of non diagnostic data.

Criteria interdependence has been treated in several ways in project evaluation models. Some approaches<sup>3,7,15,16</sup> assume that all criteria are independent. This biases the evaluations in direct proportion to the magnitude of the interdependencies.<sup>17</sup> Other evaluation models<sup>4</sup> eliminate criteria causing dependencies. Fishburn<sup>18</sup> has suggested a method for identifying such dependencies but there has apparently been no experimental validation of the technique. His method, which uses the concept of indifference between pairs of gambles, is suitable when each criteria has discrete levels and when the number of criteria is small. Unfortunately, bias reduction may be more than offset by the information loss resulting when dependent criteria are discarded. This loss can be reduced by (1) discarding criteria only when there is a high degree of interdependency and (2) discarding those criteria having the smallest influence on project evaluation.

We propose the following untested procedure for discarding criteria:

1. Select and estimate the relative importance of a set of criteria using the procedures suggested earlier; a subset of those criteria will be independent.

2. Select pairs of criteria having a major dependencies. This can be accomplished empirically if data are available. If not, experts can subjectively select those pairs.<sup>19</sup>
3. Remove from consideration those criteria that do not have at least one major pairwise dependency. These criteria can be considered independent.
4. Divide the remaining criteria into subsets having high intradependence but low interdependence by having experts sort 3x5 cards, each containing the name of one criterion, into groups such that
  - a. the extent to which one criterion is satisfied strongly implies or is implied by the extent to which another criterion in that group is satisfied.\*
  - b. the utility function of each criterion in the subset is influenced by the degree to which another criterion in the subset is satisfied.
5. Select the criterion,  $C$ , with the largest number of major pairwise dependencies. We will either discard this criterion or all the criteria with which it has major dependencies.
6. If its weight, as determined in step 1, is less than the sum of the weights of all dependent criteria, discard criteria  $C$ . If not, discard all those criteria having major dependencies with it.
7. Repeat steps 5 and 6 for the criterion having the next largest number of dependencies.
8. Repeat step 7 until all dependencies are eliminated.

### Example

Suppose we have a set of 10 criteria,  $C_1, \dots, C_{10}$ , with weights  $W_1, \dots, W_{10}$  assigned in step 1. Step 2 yielded subsets  $[C_1, C_2, C_4, C_7, C_9, C_{10}]$ ,  $[C_3, C_6]$ ,  $[C_5]$ , and  $[C_8]$ . Step 2 yielded major pairwise dependencies for the first subset as shown below:

	$C_1$	$C_2$	$C_4$	$C_7$	$C_8$	$C_9$	$C_{10}$
$C_1$		X	X	X		X	X
$C_2$	X		X				X
$C_4$	X	X					
$C_7$	X					X	
$C_8$							
$C_9$	X			X			
$C_{10}$	X	X					

\*This method for detecting criteria dependencies was evaluated by Gustafson. He attempted to predict patient length of stay by a Bayesian model that assumed data were conditionally independent. In one case, he acted as if all data were independent. In the other, he used procedure 4a to form conditionally independent subsets of data. The second method predicted length of stay better than the first. This would indicate that the proposed approach may be effective for identifying major dependencies.

$C_1$  has the largest number of dependencies (four) so it is the first to be considered (step 5).  $W_1 < W_2 + W_7 + W_9 + W_{10}$  so  $C_1$  is discarded (step 6).

$C_2$  is the next criteria to be considered (step 7).  $W_2 > W_4 + W_{10}$  so  $C_4$  and  $C_{10}$  are discarded.

Since  $C_2$  no longer has pairwise dependencies it forms a new subset leaving only the dependency between  $C_7$  and  $C_9$  to be rectified.  $W_7 > W_9$  so  $C_9$  is discarded. The new group of criteria subsets is  $C_2$ ,  $[C_3, C_6]$ ,  $C_5, C_7, C_8$ .  $W_3 < W_6$  so  $C_3$  is discarded. The final set of independent criteria is  $C_2, C_5, C_6, C_7$ , and  $C_8$ .

### Criteria Measurement

Measures of the degree to which criteria have been satisfied must be reliable, valid, and easy to obtain. Some evaluation models<sup>15,16</sup> use ordinal values as  $X_j$  entries in some variation of equation 1. These are obtained by ranking projects according to extent to which they satisfy each criterion. Unfortunately, ordinal scale values should not be added<sup>21</sup> because the resulting project scores will be biased in proportion to the degree to which the intervals between project ranks are unequal.

Other evaluation models<sup>3</sup> select only criteria whose values can be added. The important but qualitative criteria are replaced by less appropriate but more easily measurable criteria. In such an exchange, important information may be lost.

As an alternative, we suggest that criteria should be measured on an interval scale whenever possible and otherwise, ordinal scale values should be transformed onto an interval scale using the method proposed by Eckenrode.<sup>22</sup> A set of statements (verbal descriptors) are assigned values on an interval scale which indicate the degree to which a project possessing that descriptor satisfies the criterion. Sensitivity can be increased by increasing the number of descriptive phrases as long as this number does not exceed the evaluator's ability to discriminate. Previous research<sup>23</sup> indicates that men may have difficulty discriminating beyond approximately seven criteria.

In order to test the effectiveness of these two methods, nine of thirteen members of a committee evaluating medical research proposals used the hierarchical and ratio methods to estimate weights for the 40 evaluation criteria in Figure 1. They also rank ordered each of the 40 criteria. This rank ordering was a good approximation of their true feelings because their cognitive limitations were not exceeded. They compared two criteria at a time until the ordering was complete. These rankings were compared, via Spearman Correlation Coefficient, with those derived by the subjective weighting methods.

The results indicate (Figure 4) that the "ratio" method does predict rankings more effectively than the "hierarchical" method. The average Spearman coefficient

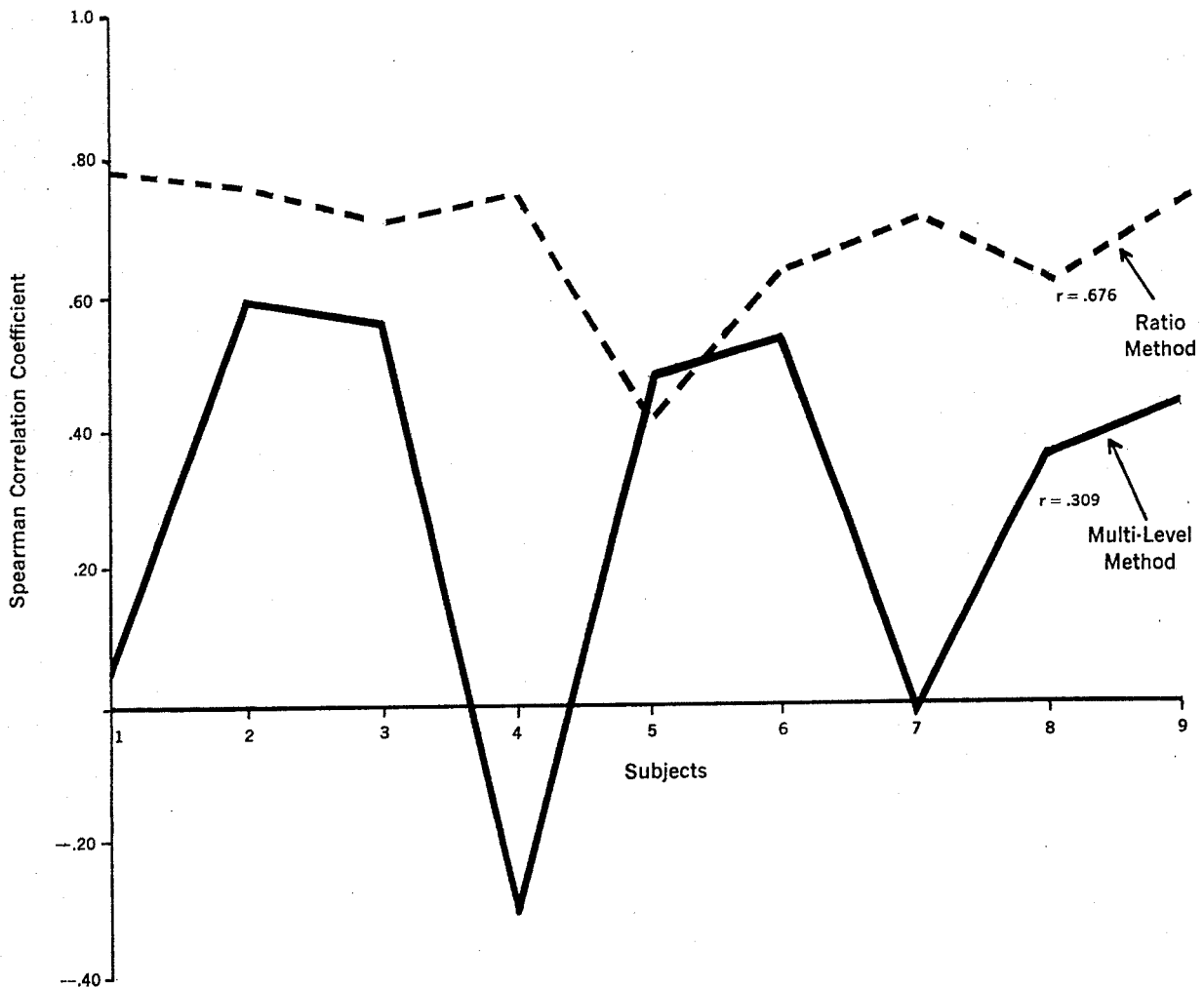
was 0.676 for the "ratio" method versus 0.309 for the "hierarchical" method. The standard deviations of the coefficients indicates that the ratio method has less variation between subjects (0.021) than does the hierarchical method (.295). This implies that the ratio method may more consistently model the decision maker's true feelings about criteria weights.

Inter-rator variability was examined for twenty four qualitative criteria in Figure 1 using a diverse group of twelve health related professionals including engineers, economists, physicians, planners, and hospital administrators. Verbal descriptors were established for the 24 qualitative criteria. Each committee member estimated the importance of these descriptors by drawing lines

from them to an interval scale. For 13 of these criteria the scale went from 0 to 100: for 11 of them, it went from -100 to +100.

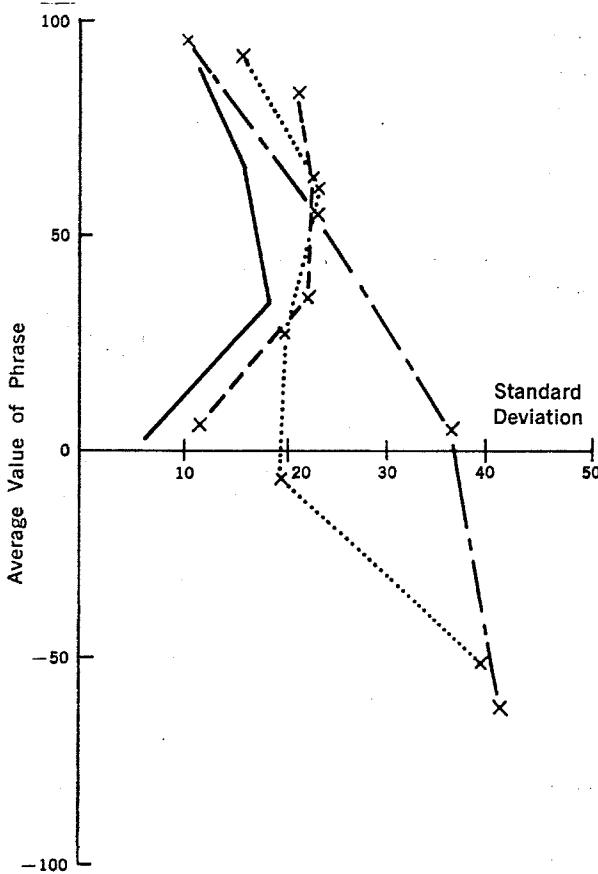
The results (Figure 5) indicate that: (1) The 0 to 100 scale has less overall variability than the -100 to +100 scale. (2) On the 0 to 100 scale, the end point descriptors have less variability than the intermediate descriptors. (3) It would appear that in each case, subjects perceive the descriptors to be approximately equally spaced in importance. This finding is somewhat discouraging because it indicates that subjects may not accurately perceive differences between these descriptive phrases. Group discussion between decision makers may

FIGURE 4.—Evaluation, via Spearman Correlation Coefficient of the Degree to Which Criteria Rankings Were Approximated by Methods for Criteria Weighting.



be one way to improve their perception of the values. (4) The variation between subjects appears to be quite large. This wide variability between subjects may be attributed to individual differences in utility functions. This may be especially pronounced in a group as diverse in background as the one tested. Much more investigation is needed into performance of subjects using descriptive phrases. However, these initial data indicate that subjects can give more than a simple preference ordering to the phrases.

FIGURE 5.—Relation between Value of the Descriptors and Variation between Subjects.



#### Utility Assessment Techniques

Before an additive model can be employed, all criteria measures must be transformed to have the same units of value. One such transformation would be to relate extent of criteria satisfaction to a utility scale.<sup>3,20</sup> When

the criteria being measured have clearly defined end points, there are several utility estimation techniques that may be used.<sup>10,20</sup> When the range of values is not clearly specified, the end point can be approximated by asking experts to estimate the value of the criterion for which they would be very surprised to find a project exceed.

#### Model Modification

If men are conservative estimators of criteria weights, they will not attribute enough importance to diagnostic criteria\* and will attribute too much importance to non-diagnostic criteria. By raising the weight of each criteria to a constant power greater than one, we are in effect, increasing the value of diagnostic criteria and decreasing the value of non-diagnostic criteria. Equation 2 represents such a modification of the weighted aggregate model:

$$E = \sum_{i=1}^n W_i^a \beta(X_i) + \sum_{j=n+1}^{n+m} W_j^a R_j \quad (2)$$

If the value of "a" that maximizes model effectiveness were constant between decision makers, it would be practical to estimate its value and thereby improve the model's evaluation capability. The value of "a" that optimized the performance of equation 2 was calculated in order to investigate this question.

Ten members of the proposal evaluation committee rated on a 0 to 100 scale twelve hypothetical projects, each described by five of the criteria in Figure 1. These ratings were compared to estimates made by the subject's weighted aggregate model (equation 1) where (1) criteria interdependence was not investigated, (2) utilities were assessed by method of order<sup>10</sup>, and (3) criteria weights were established using the ratio method. The results are indicated in Figure 6.

A "committee model" was then developed by averaging the individually determined weights and utility curves. The resulting evaluations of the twelve hypothetical projects were compared, via the Pearson product moment correlation coefficient with the average ratings assigned to each project by the ten members. The

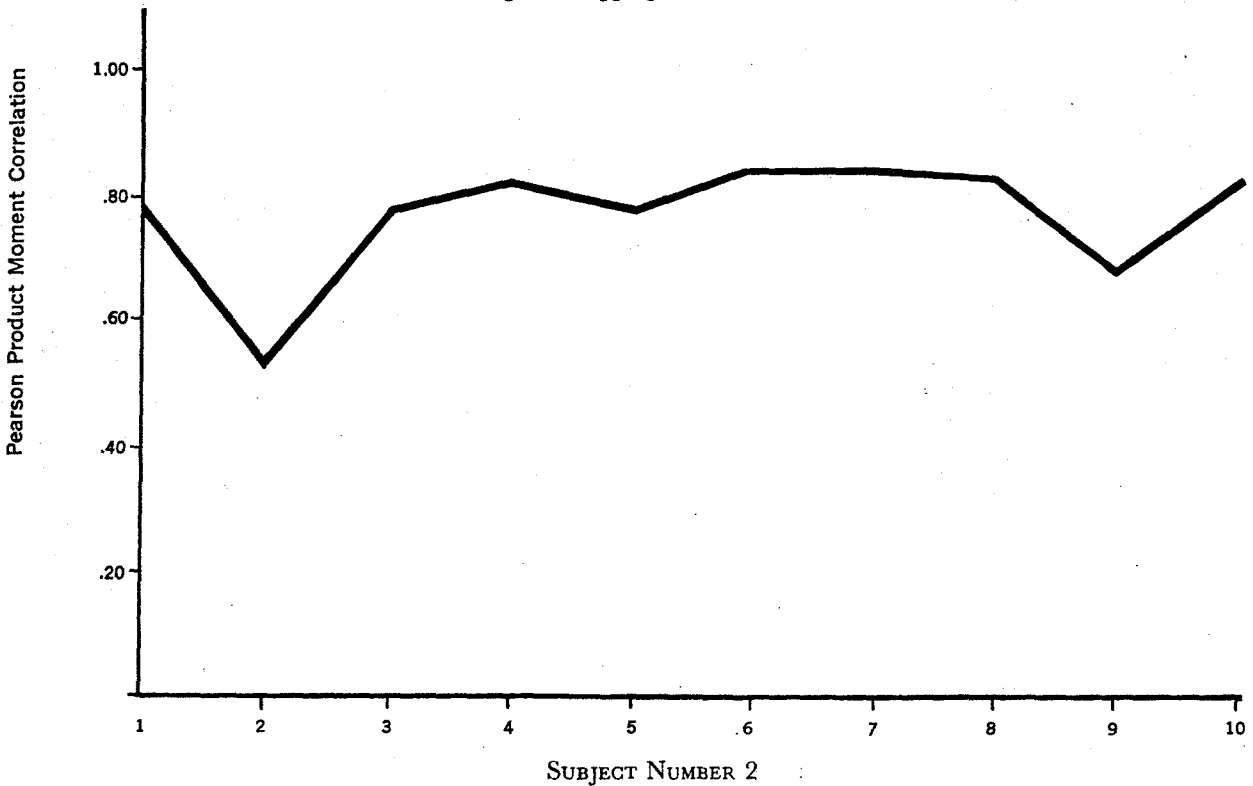
\*Diagnostic criteria are those having a major influence on the rating given a project.

Pearson product moment correlation between model and ratio estimations was 0.9.\*

We next calculated Spearman Correlation Coefficients indicating association between averaged subject ratings and committee model evaluations developed using several values of "a" in Equation 2. Results, Figure 7, indicate that model performance first improves with additional weight being given to more important criteria and then drops off as values of "a" exceed 1.50. This data would lead us to believe that subjects are conservative in weighting criteria and that equation 2 is a useful modification of the weighted aggregate model.

We next investigated the variation between committee members in the optimum values of "a". A significant variation would require separate estimates of "a" for each committee member. This would be a time consuming task for both the committee and the experimenter. Individual evaluation model performance was measured at several values of "a". The results, Figure 8, indicate that there is substantial variation in the optimum value of "a" between individual subjects. Conservatism does not appear in all subjects. In fact, some subjects appear to be radical in their criteria weightings. At the very least, this would indicate that values of "a" for equation 2 must be developed for each

FIGURE 6.—Correlation Between Experimentally Derived Project Ratings and Ratings Computed via the Weighted Aggregate Model.



\*These correlation coefficients are useful only as standards against which to compare evaluation models with "a" ≠ 1.0. The results cannot, for instance, be used to imply that the committee model is an effective predictor of committee decisions because committees do not necessarily operate on a majority rule basis.

committee member if they are to be used at all. This finding is not too surprising when viewed with the results of similar research on subjective probability estimations. The optimum value of a modifier of subjective probability estimates was influenced by the



importance of the criterion under consideration. They also found substantial variation between subjects.

FIGURE 7. A Comparison of Committee Model Performance Using Various Values of "a"

Value of "a"	Spearman Correlation Coefficient
0.6	.908
0.8	.923
0.9	.923
1.0	.922
1.1	.935
1.2	.937
1.5	<b>.955</b>
2.0	.915
2.5	.886
5.0	.702

Fortunately, it does not appear that there is much improvement to be obtained by using equation 2. The last column of Figure 8 indicates for each subject the percentage improvement that could be obtained by using equation 2 rather than equation 1. In 9 of the 11 subjects the improvement is 5% or less. This relatively meager improvement in performance indicates that the additional work required to improve the basic model may be justified only when the projects under consideration will require a large investment.

Further Model Modifications

Equation 1 assumes that all potential benefits will be achieved and that the time required to achieve each of them is the same. Neither of these assumptions is true. Model performance might be improved by considering expected benefits modified by a present worth factor as in equation 3:

$$E = \frac{\sum_{i=1}^n P(S_i | Y_{i1}, \dots, Y_{in}) W_i \beta (X_j) e^{-rt_i} + \sum_{j=n+1}^{n+m} P(S_j | Y_{j1}, \dots, Y_{jn}) W_j R_j e^{-rt_j}}{C(t)} \quad (3)$$

- where  $S_i$  = success of project in achieving benefit  $i$
- $Y_{ik}$  = degree of satisfaction of the  $k^{th}$  factor influencing the success of project in achieving benefit  $i$
- $t_i$  = time required before benefit is achieved

- $r$  = the exponent of the present worth factor relating benefit utility to time for achievement
- $C(t)$  = present worth of project costs. All other symbols the same as in equation 1.

FIGURE 8. Performance of Individual Evaluation Models at Various Values of "a". Circled Value is Best Spearman Correlation for Each Subject.

Subject No.	VALUE OF "a"										% Deviation from opt. for a=1.0
	.2	.4	.6	.8	1.0	1.2	1.5	2.0	2.5	3.0	
1	--	--	.955	<b>.976</b>	.955	.955	.910	.815	--	--	2
2	.771	<b>.778</b>	.764	.764	.585	.705	.705	.655	--	--	24.5
3	<b>.872</b>	.809	.832	.770	.760	.707	.580	.222	--	--	12.5
4	--	--	.895	.910	.910	.910	.910	.924	<b>.951</b>	.937	4
5	--	--	.820	.820	<b>.856</b>	<b>.856</b>	<b>.856</b>	<b>.856</b>	.820	.781	0
6	--	--	.960	.960	.960	<b>.970</b>	.948	.926	--	--	5
7	--	--	.895	.895	.895	.895	.916	.938	<b>.941</b>	.912	5
8	.944	.930	<b>.979</b>	<b>.979</b>	.930	.930	.950	.965	--	--	5
10	--	--	.810	.810	.868	.853	.810	.783	--	--	0
11	.925	<b>.935</b>	.925	.925	<b>.935</b>	<b>.935</b>	.920	.850	--	--	0

The basic concept behind equation 3 is not new but much work is still needed to validate its potential and to develop methods for estimating its parameters. At the same time, there is evidence to indicate that the model is practical.<sup>9,24,25</sup> The research reported here has evaluated methods of weighting and measuring benefit criteria. We have suggested but not evaluated methods for establishing independent criteria. The results of experiments conducted at the research laboratories of Monsanto Company<sup>24</sup> tend to support the hypothesis that R&D planning and control models that are based on subjective probability estimates may reliably be used as an aid in project selection and funding. Other behavioral research<sup>9,10,11</sup> indicates that the posterior probability of project success,  $P(S_{ij}, \dots, Y_{in})$  can be effectively estimated by combining subjectively estimated likelihoods through Bayes' Theorem as follows:

$$\frac{P(S_i|Y_{i1}, \dots, Y_{in})}{P(S_i|Y_{i1}, \dots, Y_{in})} = \frac{P(Y_{i1}|S_i)}{P(Y_{i1}|S_i)} \dots \frac{P(Y_{in}|S_i) P(S_i)}{P(Y_{in}|S_i) P(S_i)} \quad (4)$$

#### Bibliography

1. Friedman, M. and Savage, L.J., "The Expected Utility Hypothesis and the Measurability of Utility," *Journal of Political Economics*, Vol. 60 (1962), pp. 463-74.
2. Cramer, R.H. and Smith, B.E., "Decision Models for the Selection of Research Projects," *Engineering Economist*, Vol. 9, 2, 1964.
3. Miller, J.R. III, "A Systematic Procedure for Assessing the Worth of Complex Alternatives," 1967, Defense Documentation Centre, AD 662001.
4. Schoner, Bertram, *The Selection of R&D Projects*, unpublished doctoral dissertation, Stanford University 1965.
5. Baker, N.R. and Pound, W.H., "R&D Project Selection, Where We Stand," *IEEE Transactions on Engineering Management*, EM 11, pp. 124-134, December 1964.
6. Cetron, M.J., Martino, J., and Roepcke, L., "The Selection of R&D Program Content—Survey of Quantitative Methods," *IEEE Transactions on Engineering Management*, EM-14, pp. 4-14, March 1967.
7. Hellman, L.P., "A Value Measure for Selecting Proposals for Research Grant Support," Operations Research Division, The Johns Hopkins Hospital, Baltimore.
8. Churchman, Ackoff, and Arnoff, *Introduction to Operations Research*, John Wiley and Sons, Wiley, 1957, Chapter 6.
9. Abernathy, W.J. and Rosenbloom, R.S., "Parallel and Sequential R&D Strategies: Application of a Simple Model," *IEEE Transactions on Engineering Management*, Vol. 15, 1, 1968.
10. Huber, G.P., Sahney, V.K. and Ford, D.L., "A Study of Subjective Evaluation Models," Social Systems Research Institute, University of Wisconsin, Madison, Firm and Market Workshop Paper; 6817.
11. Philips, L.D., Hays, W.L., and Edwards, W., "Conservatism in Complex Probabilistic Inference," *IEEE Transactions on Human Factors in Electronics*, HFE-7, 1966.
12. Philips, L.D., *Some Components of Probabilistic Inference*, Doctoral Dissertation, The University of Michigan, 1965.
13. Schum, D.A., "Behavioral Decision Theory and Man-Machine Systems," Technical Report, Systems #46-4, Rice University, Program in Applied Mathematics and Systems Theory, July, 1968.
14. Smith, Lee H., "Ranking Procedures and Subjective Probability Distributions," *Management Science*, Vol. 14, No. 4, December 1967.
15. Mottley, C.M. and Newton, R.D., "The Selection of Projects for Industrial Research," *OR*, 7:6 Nov.-Dec. 1959, pp. 740-751. (Propose Method: Decision Theory, Rating).
16. Hertz and Cerlson, Phillip G., "Selection, Evaluation, and Control of Research and Development Projects," in Burton V. Dean (ed.), *Operations Research in Research and Development*, Wiley, 1963.
17. Mosteller, R. and Wallace, E.J., *Inference and Disputed Authorship: The Federalist*, Reading: Addison-Wesley, 1964.
18. Fishburn, Peter, "Independence in Utility Theory with Whole Product Sets," *Operations Research*, Vol. 13, No. 1, 1965.
19. Gustafson, D.H., Unpublished Doctoral Dissertation, The University of Michigan, Ann Arbor, 1967.
20. Fishburn, P.C., "Methods of Estimating Additive Utilities," *Management Science*, Vol. 13, No. 7, March, 1967.
21. Nadler, G., *Work Design: A System Design Strategy*, to be published soon, G.P. Huber, Chap. 7, General Models—Decision Making, in the above book.
22. Eckenrode, R.T., "Weighting Multiple Criteria," *Management Science*, Vol. 12, No. 3, November, 1965.
23. Miller, G.A., "Magical No. 7 + or -2: Some Limits on Our Capacity for Processing Information," in Alexis and Wilson, *Organizational Decision-making*, Prentice-Hall, 1967.
24. Souder, W.E., "The Validity of Subjective Probability of Success Forecasts by R&D Project Managers," *IEEE Transactions on Engineering Management*, Vol. 16, No. 1, 1969.
25. Edwards, Ward, "Conservatism in Human Information Processing," in *Formal Representation of Human Judgment*, Benjamin Kleinmuntz, ed., Wiley, 1968.

#### Comments on an Evaluation Model for the Regional Medical Program

VERNON E. WECKWERTH, Ph.D.

How generic one wishes to make a model depends on how far one is displaced from the reality of application. The creator of a model in the Ivory Tower can easily assume away the inconsistencies of the world. To the day-to-day doer of what could be called evaluation, there is no way to assume away the problems in the world. Judgment is totally pragmatic. The applied model either represents what is or it is rejected.

As a group, you have been subjected to some high level forms of abstraction in terms of starting points, preliminary strategies, ends-in-view, and implementation with stated intents of transformation of the system. This introduction will, by virtue of that type of presentation, try to be as abstract and obtuse.

You have been told, and by report most of you have acquiesced at least, to the proposition that the RMPs do not form a closed, but an open system. That open system is a seductive proposition. It is as seductive an alternative as many propositions are when the ends-in-view are mundane or repetitive. If the system is one of a static nature – closed, just input, throughput, and output – which is routine, reproducible, repetitive, standardized like a ball-bearing production system, then it is even easier to be seduced.

I propose, however, that the open-ended system embrace is as deceptive in the argument for it as the argument that any living, on-going process like life itself is better than a dead-end. Even the old truism summarized that belief from antiquity – you only have one life to live – you can't live it over again – you are all different. Each RMP is unique and dynamic. For our own mental health, could we believe otherwise?

There are two points to be made:

1. A model is only a model. It can be made sufficiently complex so that it fits within a predetermined degree of closeness to perceived reality so that you choose to believe it and use it, i.e., you choose to believe that the model fits your perception of things rather than concluding that life is a haphazard sequence of chaotic happenstances. It depends on your view of the meaning of change – from what to what in what direction at what rate. In fact, one could play with the words and redefine status quo to be a constant rate of change. What then happens to the obligation to transform the system? It is merely the difference between evolution and revolution. Orderly change with a built-in planning sequence is a necessary part of any dynamic organization. I am concerned that what the “change” model implies is best described as the “rocking chair model” – giving the health field a sense of movement but no sense of direction. Restated, “evaluation of transformation of the system” requires an articulate statement of change from where we are to where we intend to go by a series of defined steps.

2. If the end-in-view is looked at only as input, throughout, output, rather than in the structure of input, content in a context, then I propose that it's the wrong model. I propose that the definer of a closed system has forgotten: the context of uniqueness, that

process is dynamic, that outcomes (and benefits) are what we seek. Change is a means or an observation of means, not an end.

The generic nature and benefits of the model for evaluation proposed here are one of a system possessing six ordered elements:

1. Context - That piece of the world under consideration as it is found at a given point in time. This is the “where” for the RMPs.
2. Content - The inputs of men, money, and material in whatever extant form they are possessed, whether or not they are identified, ordered, or measured. This is the “who” and the “what” for the RMPs.
3. Process - The way the content is put together in some functional, organized way, both in terms of the static, i.e., repetitive closed system meaning like a production process, as well as in terms of the dynamic system of self-modification and directed change. This is the “how and when” of the RMPs.

These three elements are in fact the *independent* variables for any RMP. Each RMP, by its existence, structure, and function, delimits and encompasses at any point in time the dependent elements which are:

4. Output - This is the product produced from content in the process in use within the context of the operation. These are the observable, recordable, reproducible, measurable “why's” of the RMP's using the classical definition of evaluation, i.e., comparing accomplishment with stated objective. These typically form the evidential basis of hard fact observation, on which “output only” evaluation is based.
5. Outcome - These are the time-delayed impacts that demonstrate whether the outputs were any more than just outputs at the points in time. Outcomes (over time) show the time-delayed impacts of output on health states, disease incidence, updated practice, altered organization, complete and continuous care delivery, equalized access, cost effectiveness, etc. These *should* be the “why's” for the RMP, but these kinds of “why's” are either too soft in the data sense or take so long in the time sense that they are only rarely used. The output “why's” are accepted as the basis for funding perpetuation and classical evaluation.
6. Benefit - This is the ultimate “why.” It is also the vaguest and “softest” element in evaluation. It gets at the associated, serendipitous, as well as intended effects that are evident in an altered context. Benefits can be represented in imputed cost

benefit terms or in a gestalt sense of total changes or differences in the context at a subsequent point in time.

Obviously each RMP is unique if one considers a sufficiently large number of items of context. The "wheres" are unique by combinatorial reduction to absurdity.

Each RMP is also unique if one considers the specific combination of the process (how's) chosen. The *how* for each RMP given merely combinatorial structure makes each obviously different from any other.

Given these unique, independent variables in combination the outputs will be by definition unique, depending on how crude or fine one chooses to make the output units.

The outcomes will also be obviously unique, depending, of course, on what time frame is used.

The benefits must of necessity be unique since the context was.

Obviously one can choose or not to be seduced by the age-old proposition that each is different from everyone else, i.e., each RMP is an open, not a closed system. The issue is not that RMP's are open or closed and therefore different but how different. How different must they be so that being different makes a difference? The burden of an evaluation is to categorize, order, measure, and interpret the differences — either relatively or absolutely.

The evaluation issue at hand is answering the simple question, "What social good has the RMP produced?", where in fact the evaluators have the right and the obligation to define "good."

Or restated, what are the outcomes (or benefits) upon which RMP is to be judged? On what basis are they to be held accountable?

If it is to be on the basis of a change or the fancier euphemism, "transformation of the system," there must be a clear statement of what "good" means in terms of changed to what, from what, at what cost/unit of change in what time frame — not just a nondirectional rocking chair model.

If it is to be on the basis of process, then the rate of change and the time horizon must be defined.

One would have to conclude that the goals and purposes of RMP were intentionally stated in the vague way they were because there was no desire to be held accountable or there was no clear *raison d'être* for them. Apparently, it is now becoming necessary to define "good" in terms of the process of change without saying from what to what at which rate in what time.

The framework of the conceptual model represented here which has as its basis a markovian process is a model which may not be explicit enough for day-to-day doing in RMP but the sequence — context, content, process, output, outcome and benefit — is, however, applicable at all levels — be it to projects, to the local advisory or regional advisory groups, to the core staff, to the board, to separate RMP's, or to the RMP as a program.

It should be clear that I believe that evaluation is merely a means of responding to the question of the "social good" of the RMP. It can be answered relatively or absolutely. It is simply a judgment or opinion of the person with the right to decide. This point is made very clearly in the paper, *A Tool or a Tyranny*.

One last comment before the paper: Evaluation is distinct from assessment. Assessment means to produce the evidential base by which statements such as more, less, or equal can be made. Evaluation means to attach such words as good or bad to those assessment findings. It is necessary to be clear on the value judgment meaning of evaluation versus the quantitative meaning of assessment. For example, it is possible that the same level of assessment data could be judged to be "bad" in one context-content-process combination and for the same level judged to be "good" for a different context-content-process combination. Obviously, an evaluation, in my opinion, can be good or bad, better or worse, whether the assessment data is identical in measured quantity or order. This ends the introduction and leads to the delivery of the formal paper which Mr. Ichniowski asked me to discuss with you, weaving into it your questions and comments.

### On Evaluation: A Tool or a Tyranny<sup>1</sup>

VERNON E. WECKWERTH, Ph.D.

Evaluation is a ten letter word - in English. Beyond that statement the only consensus about evaluation is a lack of consensus. This paper is a series of loosely related topics which attempts to give some limited perspective into what evaluation means, how and why it is done and

---

<sup>1</sup>This paper is distributed for general interest. Reproduction in whole or part is permitted if proper credit is given. This distribution neither expresses nor implies approval of its contents by the Project, the University of Minnesota or the Granting Agencies.

in what ways the vernacular use of the term in management relates to the discipline use in research. It highlights four points:

1. There is no *one* way to do evaluation.
2. There is no generic logical structure which will assure a unique "right method of choice."
3. Evaluation ultimately becomes judgment and will remain so, so long as there is no ultimate criterion for monotonic ordering of priorities, and:
4. the crucial element in evaluation is simply: who has the right. i.e., the power, the influence, the authority, to decide.

## INTRODUCTION

A discussion of evaluation will lead to no useful result unless one states at the beginning what evaluation means; why evaluation is being done; to, by, with, and for whom; what is the intended outcome of evaluation; how does one "evaluate evaluation" and who has the right to decide the what, why, where, when, how, and who involved in evaluation.

Evaluation includes within it consideration of approaches, methods, techniques, and uses; a process versus a goal approach; program versus individual objectives; needs, demands, desires, and their interrelationships. It includes objectives versus goals; activities versus accomplishments; inputs versus outputs; outputs versus outcomes, outcomes versus benefits; effectiveness versus efficiency; structure versus qualification; and so forth. It includes the context, the content, and the process; the served and the server; the individual and the group; the quantity and quality; and others. It includes when and where, with or without feedback, and how often. It includes a research versus an administrative meaning. It includes vernacular versus discipline definition. It includes much more than this.

### Dictionary Definition of Evaluation

The dictionary says that to evaluate means "to determine or fix a value of" or "to examine and judge". These two meanings give the first insight into evaluation. The term, "evaluation" has *value* as its root.

Using the dictionary definition, one can separate papers and practice into those to whom value means: 1) a number value, or 2) a value system value. These two groups can each be divided into those who are process versus goal oriented. What is commonly missed is that any element (variable, quality, attribute) that one selects to be included for number value measurement is the result of someone's priority in its selection, i.e, it is of

value in the value system of the one with the right to decide what is to be measured.

All of us know the single most common application of evaluation is to the evaluation of the quality of health care. Quality of care, we know, serves to explain if costs are high, productivity low or demands too great. It will serve here as the example to trace the development of how we arrived at where we are in the Art and Science of evaluation.

### EVALUATION OF THE QUALITY OF CARE

Consider the word quality. It has the same root as qualities. Originally, qualities were selected as the basis for the first quality of care studies. The first question asked upon beginning a quality of care study is, "what is to be included to be measured?" That's where the laundry lists began. Out of that long list, a set was chosen by whosoever had the right to decide. Typically, the qualities were chosen because they either had to be present or were desirable. Thus:

Development one: A list of qualities was presented (a value system value decision) in which merely presence or absence of each quality was recorded.

An array was generated with a laundry list on the left and two columns to check either absent, score it 0, or present, score it 1.

The measure of quality was therefore simply the number of qualities present divided by the total number of qualities. Low quality meant: a proportionately small number of qualities present; High quality meant: a proportionately large number of qualities present.

The first use of evaluation of quality was to make present the qualities that were absent.

As time passed, it became obvious that some qualities were more important than others.

Development two: A weight was attached to the qualities reflecting the importance of each quality.

Obviously, these weights were attached based on who had the right to decide. The array was modified by adding a column of weights.

The measure of quality thus became the sum of weighted presence of qualities. As time passed, these weights became somewhat "standardized" and there developed what we now know as the setting of standards of quality of care. It was a way of saying what qualities had to be present. High value on a quality was reflected in a large weight. Sometimes qualities were judged and weighted so highly that absence was identical to a veto.

Development three: Place a sufficiently large weight on any one quality so that if it were absent the "quality of care" would assuredly be "low".

Quite soon the simple dichotomy, absent or present, was as unacceptable as was the equal weighting. It was natural to expand the measure of presence from 0 or 1 only to 0, 1, 2, 3 ... to as many "units of more-so-ness" as was useful. These degrees of more-so-ness did not have to be whole units or integers. These "measures" tied easily into "standards" since some standards were in fact a level of the degree of presence rather than merely presence or absence.

Development four: Specify a measure of the degree of presence for qualities.

Such a development was conceptually easy to come by, but operationally very difficult to achieve. However, that mechanical difficulty didn't deter the doing of evaluation of quality of care. The procedure merely became a listing of included qualities; the listing of associated weights, and an associated measure of the degree of more-so-ness but combined in some "arithmetical or number value way".

Once that "arithmetical way" was determined, one merely proceeded to specify the distribution of the values and define low and high quality on the scale of that measure.

There were, however, in the 40's and 50's many other forces operating; new knowledge of statistics, probability theory, experimental design, and other measurement technology. People were increasingly dissatisfied with simple arithmetic ways, including the implicit assumptions of *independence* among qualities in the list.

Those faced with evaluation were soon developing sophisticated research designs with fancy mathematical models, formulae, and techniques. The limit functions, interdependence of qualities handled by multivariate correlations, covariance, factorial designs with interactions, simple and main effects plus factor analysis all became involved. In fact, these developments became the life blood of the biostatisticians and the death potions of most of those involved as delivering practitioners - both clinical and administrative.

Development five: Only qualities with experimentally determined measurability, validity, and reliability were permitted to enter quality of care evaluations.

As a result, the evaluation of quality of care developed to such a mathematically sophisticated extent that those who first desired it and created it were bypassed and found that it couldn't be applied on a day to day basis. Hence, evaluation became so detached that now it is not recognized as a part of the ongoing process of clinical management, or program administration, i.e., planning, organizing, assembling resources, directing, controlling, replanning, reorganizing, etc. It is seen as

two completely separate endeavors with the practitioners worse off than before, since "evaluation" must now mean something detached from day to day practise, and in use most likely punative in addition.

#### WHAT CAN THE PRACTITIONER DO?

Every practitioner has taken at least the first steps in evaluation. Each practitioner must determine how sophisticated he wants to get and be prepared to defend where he stops, if he stops short of research design. The steps are simple:

1. Choose the qualities.
2. Attach weights reflecting priorities.
3. Specify measures of degrees of presence.
4. Combine the created array in some functional form(s).
5. Generate distribution(s) of those function(s).
6. Set the cut off points to determine where the quantitative representation concurs with his judgment of desired quality.

He can call in help at any step; develop any number of experimental designs and number value functions, but ultimately that evaluation will boil down to who has the right to decide and who renders the judgment.

#### ACCEPTED OPERATING DEFINITION OF EVALUATION

Dictionary definitions help to give insight into the "whats" of concepts. Operational definitions help to give insight into "how's" of concepts.

The most commonly accepted operational definition of evaluation, the "how", is: *Compare accomplishment with stated objectives*. This is itself a goal oriented definition. The objectives are analogous to the qualities or elements chosen in the quality of care example.

Since the operational definition is so simple - why is evaluation so tough? Let's look first at that operational definition. In it five assumptions are made: 1) objectives are stated; 2) in measurable terms; 3) accomplishments are documentable; 4) in the same measurable terms as the objectives; and 5) one knows what compare means, i.e., what is to be done?

#### WHAT USUALLY LEADS TO DIFFICULTY?

First: Objectives aren't stated. Goals versus objectives are rarely differentiated. Purposes, goals, salutes to mother and country - and lots of other things are usually stated - but not objectives. An analogy may be helpful to distinguish objectives from goals. Consider the sequence,

$1/4, 1/9, 1/16, \dots 1/n^2, \dots$  In this case, that sequence of terms will approach a limit. That limit is analogous to a goal. The individual terms in the sequence are like objectives.

Second: Even if objectives are stated, most of them are not independent. In fact, they frequently are in conflict with each other and rarely would their summation add up to the program goals. Additionally, the state of the art (or science) of evaluation has not developed means of measuring most value system objectives. Thus, our measurement ineptness reflects both our ignorance and our errors.

Third: Even given stated objectives and appropriate measures, we likely can't enable the documentation of accomplishment. Frequently, the measures are too complex or the day to day documentation is either too tedious, or not visibly relevant to the job being done on an ongoing basis. As a result, we substitute approximate measures or frequently just get lost in the data acquisition problems and consume so much time and resources that we judge that documentation isn't worth it - unless it is an experiment in which service is only a necessary evil or a necessary context.

Fourth: In the rare event that evaluation has measurable objectives and documented accomplishment, commonly nobody knows what to do with it! Or if, in fact, someone knows, the comparison will still depend entirely on the judgment of whoever has the right to decide what to do with it.

A facetious and trivial example may help: suppose that an MCH Program has an objective that 75% of all mothers-to-be are to be seen by an O.B. physician before the third trimester. We find that 73% do in project A, and 77% do in project B. Now what? If  $n$  is big enough, the difference may be statistically significant. So what? Is the project with 77% awarded a gold star or more money? Does the project with 73% get a budget cut? In fact, is it not true that since both missed the objective, that both are bad? Why is doing more an ultimate good? After all, the 77%er allocated more resources than should have been to that objective and that project could be "penalized" for misallocation while the 73%er should be given more resources because it was under-allocated.

The overriding question being asked is, is the classical operating definition of evaluation: *Compare accomplishment with stated objective* the end of evaluation? Is evaluation to be only descriptive? Is it merely to tell how it was? If not, is it to include ground rules for translating description into prescription, i.e., administrative action?

## SOME COMMENTS ON MEASURES

Frequently, a quality selected in evaluation has no direct measure or has one which is too costly or tedious to obtain. There frequently, however, are associated or indirect measures which can be used in lieu of a direct measure.

Some measures which are indirect are called proxy measures. This obviously means that they stand in lieu of what is desired to be measured. Frequently, proxy measures in evaluation are used to predict or monitor activities, and are useful because of their high associated though not causal relationship.

For example, the number of individuals using an emergency room in hospitals is associated with the phases of the moon. For administrative purposes of staffing and the provision of service, it is not necessary to know the direct or causal elements. However, if one were to change the pattern of service "demand" it would be necessary to know cause - and the relationship, and not operate purely with proxy measures. Commonly, "Comparison" in evaluation highlights differentials in such proxy measures. Actions are then frequently taken on forces putatively "causal" but to the dismay of the action taker, produce no change because - in medical jargon - he treated the symptom and not the disease.

These experiences further alienate the practitioner and result in his questioning even more, "Why evaluate?"

## USES OF EVALUATION

No attempt is made to provide a laundry list of uses. An attempt is made, however, to fit "evaluation", in the non-experimental design meaning, into day to day operations.

First, we must answer, "For what purpose is the evaluation done?" Regrettably, the answer that would now be given (if honestly ascertainable) is, "The \_\_\_\_\_ law requires it." That is regrettable. In a sense, the requirements in the law reflect a failure on behalf of those responsible for programs to document accomplishment in an orderly, measurable, and articulate manner that met the desiderata of those with right to make laws.

With the legal emphasis on evaluation and the meaning of the term to be the rigid mathematical, numerical, hard fact one, the day to day intuitive or soft data meaning and use, has been both lost and rendered unacceptable.

Evaluation has always been - in the dictionary meaning of the word - present in anyone who was responsible

in his work, and had a personal accountability for his acts. Evaluation is inherent in the *process* of administration - be it clinical or program management.

Anyone who manages successfully either a program or a patient goes through some orderly stages, beginning with planning: that is deciding what is to be done; by, with, for, and to whom, with what material, at what time, in what sequence, at what places, for what intended outcome.

Thus the planning is the *what* step in the administrative or management process with the *how* steps being organizing the what, assembling the resources, directing the delivery and controlling (or supervising, or monitoring) the operation (or performance).

An inherent part of the management process is its evaluation. That examination and judgment in delivery of care is used as feedback to alter the process or treatment of choice in order to replan, reorganize, reassemble, redirect, control, etc., ad infinitum.

Clearly, evaluation has been, is, and always will be, a part of such a management process - be it for a program or a patient.

The similarity in the process can be seen if we move from the individual care of a patient, through a cohort of patients to a program. Consider yourself first as a physician beginning with a work-up. You first chose input facts, i.e., qualities, such as lab tests, signs, asked symptoms, soundings, touchings, etc., plus using the history to assess the patient, derived mentally a set of weights of what's important, arrived at what's relevant (by degrees of presence plus weighted priority), determined a most probable "value" or judgment (or evaluation) and rendered a care plan. You subsequently compared this to what happened to the patient and, depending on the outcome, either altered the care plan or reinforced your confidence in your own medical judgment, i.e., you evaluated on a one case basis.

Consider next a cohort of patients. You look at them as a group. You select another set of qualities (some of which are different from the case specific qualities chosen in the one patient sense) and look at the cohort from a view of those qualities being a set of intertwining degrees of presence and priority. You mentally and numerically measure and then compare the results of the cohort to what is "good medical practise".

At the program director level, you'd look at more than only physician case management for either individuals or for his cohort and include the other health care functional services, living conditions, or what have you, that are qualities of the "program" and go through the same process to determine whether it accomplished what

you stated it would. You have evaluated at the program level.

Although there is a reasonable basis for saying there is a single generic process in doing evaluation, the qualities chosen for patient management are so different from those chosen for program management that the singleness of the process is lost. In fact, because the priorities assigned to the qualities in patient versus program evaluation are so discrepant, conflict has resulted in the whole health care delivery system.

#### HOW TO CHOOSE THE QUALITIES

Since all of us come from rigorous scientific fields, we almost without thought believe we choose qualities based on the facts. What one means by "based on the facts" necessitates some expansion.

For this paper, consider four groupings of facts:

First: Theoretical facts. Starting with *givens* and a set of known theoretical relationships, one by deductive logic can arrive at some qualities which are to be included in evaluation.

Second: Dogmatic facts. Dr. Lebon (that spells nobel backwards - and he has one of those prizes and don't you forget it) says this is a fact - and it is. In general, these are the qualities which those in positions of power, influence, or authority include in evaluation.

Third: Pragmatic facts. Those which are based on astute observations, with data acquired from day to day practise which every intelligent practitioner gathers. These form the basis for selecting another set of qualities. In general, they derive from "experience and demonstrated use ..."

Fourth: Experimental Research Facts: These are the facts derived from research studies which meet the most rigid of experimental design requirements. The resulting qualities are chosen by approaches and methods such as factorial designs, controlled probability selection, or any of the research statistical methods that strikes fear into most day to day practitioners.

From these four fact bases, one can get the qualities to be used in any evaluation schema. It is here also that standards with which we are so obsessed in health care delivery are included.

#### WHY MUST ONE USE EVALUATION?

If one is the perfect clinical practitioner or the perfect program director, his intuitive ongoing soft data system would be "evaluating" without need for a hard fact base. But, since perfection is not a human reality, one must set up a hard fact data system to document



accomplishment. The less prestigious one is, the more subject one is to the "tyranny of hard fact evaluation".

Since one cannot get continuous evaluation, some choices of time intervals must be made - hourly, daily, weekly, monthly, quarterly, yearly, or what have you for ongoing programs. Evaluation of single shot programs are relatively easy if only a "final" evaluation is to be made. One must determine if feedback is to be used - of what kind, and how often. If so, how does feedback fit into a subsequent round of evaluation? Is it now another quality or element? If one does feedback "evaluation" with the intent to alter the program, how does one now evaluate the effect of evaluation?

### APPROACHES TO EVALUATION

No attempt to be either scholarly or complete is intended here. Only three commonly used approaches are included:

First: Very commonly, programs are subjected to periodic review. These "evaluations" are made by a squadron of outsiders. Let us call this the J.D.A. - the judgment day approach.

The big brother squadron, usually called a site visit team, comes in on judgment day. The concern is obvious, "Will one be judged for sins or virtues using the same qualities that one has used to live by?"

Second: Another commonly used approach is one of being reviewed by a hand picked panel called peers. Let us call this the B.R.A. - the bunny rabbit approach. It's title comes from the setting in which Johnny brought a rabbit to kindergarten for drag and brag (Show and Tell). Mary asked if it were a boy rabbit or a girl rabbit. Johnny said, "I don't know". The precocious Mary said, "Since this is a participatory democracy, let's vote."

Although both these approaches have been practised successfully (at least in the evaluation of those with the right to decide), the invalidity is obvious: for the first, one only needs to have the right to choose the qualities and the measures and the weights and the cut off points; and for the second, one merely needs the right to choose the majority of the panel.

Certainly no one could object to those simple requests if the "right to decide" is not the crucial issue in evaluation, and if evaluation does not ultimately become judgment, i.e., the opinion of the person with the right to decide.

Third: The third commonly used approach is the R.C.A. - the report card approach. It is essentially the approach used for evaluation of the quality of care example at the beginning of the paper.

Consider the old fashioned grade school report card. The "qualities" are analogous to items like courses in math, others like art, others like deportment.

By analogy, three groupings of qualities of report card items are apparent:

1. Those that have an inherent measurement absoluteness in them (even though the measure may be arbitrarily defined) like feet, inches, etc. The units have a meaningful metric on the scale. The mathematical formulas work beautifully.
2. Those qualities that have an inherent relative or more-so-ness meaning to them but lack absoluteness such as strongly agree, agree, indifferent, disagree, strongly disagree. Again, the mathematics is reasonably easy to apply.
3. Those qualities that are named or categorizable only. These are those qualities that either have no inherent measure of absoluteness or relativeness or that as yet aren't understood well enough to be measured. It is with these, where real difficulties in the mathematics are found because the weighting is not inherent nor is there a logical way to attach priority values.

Since every program or practise includes all three kinds of qualities, we must, in our wisdom choose, weigh, scale, combine, and then compare to the objectives, i.e., judge the result.

We render an evaluation. So what? We have gone through a magnificently structured and logically justifiable process with bewildering numerological finesse to arrive at the end point - a judgment or opinion of what to do with it.

### WHAT DIFFERENCE DOES IT MAKE?

It makes a difference only if the person in the position with the right to decide agrees.

This formalized ritualistic numerological game called evaluation, is a series of decisions of those with the right to decide and ultimately rests on the judgment of the person who can determine the outcome by:

1. Choosing the doers of evaluation.
2. Choosing the elements for inclusion, and/or
3. having the right to decide what comparison means.

Such evaluative manipulation can occur whenever there is no ultimate criterion which assures a unique ordering of priorities, and the resulting correct method of choice.

### WHAT DOES IT MEAN TO HEALTH PROGRAMS?

1. If one doesn't play the game, or even worse realize what the ground rules are, one may lose the

funding or have funding reduced since it's easy to relate dollars to points scored on a hard fact evaluation index.

2. It's an effective way for funders to mold or shape the program, i.e., dictate the health care delivery system. They need only specify the proxy indicators or elements, their weights and their measures, and attach adequate punishment and rewards so that grantees desiring continued support will allocate the resources to maximize the evaluation index. It's the health care version of "shape up or ship out".
3. Quite clearly, those elements that are easily measurable will get the attention and be assured of inclusion in such an evaluation. I am personally concerned that what is really important in life is inversely related to what is easily measurable.

#### OTHER CONSIDERATIONS

Currently in vogue also is efficiency of health care delivery. One of the chosen qualities is efficiency in every evaluation. The usual operational definition is one lifted bodily from engineering - the ratio of output to input.

The hazard of this measure is clearly from whose perspective output is viewed. From the doer, his activities are always viewed as output. From the receiver, those outputs are always viewed as inputs.

The classical data which allegedly measures output of laboratories, groups of personnel, institutions, etc. such as visits, lab tests, encounters, and the rest are really only inputs to the health of the seeker of service. Even more interestingly, within the sequence of doers, the prior doers' output is also viewed as an input to the next one in sequence. Thus, the lab technician believes he is highly productive because of outputting many lab tests, and the engineering definition gives him a very high efficiency rating. The physician or nurse, however, looks at the lab tests merely as inputs and they in turn, value their visits and activities as the real outputs upon which efficiency should be based.

What is incredible is that none of these measures of efficiency really get at the question to be answered - namely, are any of these inputs or outputs effective in maintaining or altering the health status of the recipient of service.

Clearly, effectiveness must first be defined before efficiency has any useful meaning. It appears that we are producing a health delivery system which is unit by unit approaching 100% efficiency while simultaneously marching toward the other extreme in effectiveness.

By analogy, we are merely counting how many times the bird flaps his wings, without asking, did the bird fly - let alone how far and how high.

Clearly, outcomes as the measures of effectiveness must be the starting point for evaluation before any of the measures of input or output analysis of the efficiency kind have any meaning or usefulness.

#### TWO AIDS TO ASSIST IN EVALUATING HEALTH DELIVERY

In the face of such a bewildering maze of considerations, two simple lists of elements are helpful in retaining ones sanity: The first are the five A's.

In the evaluation of any health care delivery, questions of appropriateness, availability, accessibility, and acceptability to both seeker and server must be answered. These are a dependent sequence. For services can be deemed appropriate yet be unavailable. Or they can be defined to be available, yet not accessible. Or they can be defined to be appropriate, and available and accessible, but still not acceptable to either or both the serving staff or the seeking client. However, overriding these four A's is the one called accountability. It is the essence of the moral contractual agreement made between the seeker when he seeks and server when he serves.

The second list is the generic structure of evaluation implicit in this paper and necessarily a part of the process of evaluation. There must be six interdependent elements to any evaluation undertaking:

First: Context (what, where, when, and who).

Second: Content (program elements being or intended to be provided and why).

Third: Process (how care is organized and delivered).

Fourth: Output (how many times did the bird flap its wings).

Fifth: Outcome (did the bird fly).

Sixth: Benefit (how high and far, with what resources).

Clearly, context, content, and process combine in many ways to produce the output, the outcome, and the benefits.

#### SUMMARY

This paper was intended to give some limited perspective into the what, why, and how of evaluation. It highlighted the reasons for misunderstanding between the hard fact approaches to evaluation and the day to day uses. It is not easy to describe a program even in terms of telling how it was. For ongoing programs it is even

more difficult to tell how it is. It is virtually impossible to tell how it will be, or as some glibly say, how it should be.

It is necessary for each of you as accountable and responsible health devotees to DESCRIBE how it was, but it is more important to structure ways to PRESCRIBE how it will be. This may be the difference

between the SCIENCE (retrospective description) and the ART (prescriptive action) of administration. Hopefully, program evaluation will continue to serve and develop as the management tool it first was, and still intended to be and will not become the program tyranny of the 1970's.

# EVALUATION OF CORONARY CARE TRAINING

## Participants

Rodger M. Shepherd, M.D. - Moderator  
Assistant Program Director, Continuing Education  
Presbyterian Medical Center

Daniele Deverin  
Cybern Education, Inc.

Mariella Larter, M.S.  
Coordinator, Washington/Alaska Coronary  
Care Unit Nurse Training Program

Carol Larson  
Allied Health Specialist  
Continuing Education and Training Branch  
Regional Medical Programs Service

---

### Evaluation of Coronary Care Training: Some Direct Observations of Performance in Hospital Practice

RODGER SHEPHERD, M. D.

The objective of our Intensive Care Training Program is to enable physicians in cadres from small general hospitals to perform certain intensive care skills in their own hospital settings. These skills include: use of central venous catheter, use of intra-arterial monitoring catheter, interpretation of blood gas data, continuous EKG monitoring, airway care, controlled ventilation, cardioversion, and others. The staff of our ICU had visited small hospitals and identified these skills as feasible but underused in smaller hospital ICU's.

The training program is conducted in three phases. During the first phase, the cadre and project clarify mutual objectives. During a second phase, each physician from the cadre undergoes a week-long program of one-to-one instruction at a metropolitan medical center. During a third phase, an instructor-in-residence is maintained in the cadre's own intensive care unit around the clock for 10-12 days.

The direct observations of these instructors have provided valuable anecdotal data on both the project and the resulting student performance:

1. Standardization of Technique: The same single standard technique for insertion of central venous catheter is advocated during each individualized instruction. The mastery of this technique is certified by the instructor. However, the student may not implement this technique in his own hospital setting. It has been observed that the failure of some physicians to support standardized

technique has a disruptive effect and reduces the tendency of other physicians to implement the advocated procedure at all.

2. Availability of Equipment: Standardized technique depends on standard materials. Instructors have observed the lack of certain critical materials or instruments during introduction of a new technique. The attendant frustration during this critical phase may abort or seriously retard the adoption of the new practice in spite of adequately trained personnel.
3. Supporting Services: Interpretation of blood gas data depends on complete confidence in the data. We have encountered one hospital setting where the student's training in interpretation of blood gas data was not implemented until we had rectified certain analytical problems in the clinical laboratory.

### Report on Xerox Study of Eleven National Coronary Care Training Centers

DANIELE DEVERIN

In 1967, Xerox Education Division was contracted by Public Health Service to conduct a 2-year evaluation study of eleven national coronary care training centers.

### OBJECTIVES

The study was designed to fulfill the following objectives:

1. To determine the effectiveness of the training programs in imparting the knowledge, attitudes, and

skills needed for a nurse to perform in a CCU at an acceptable level.

2. To determine the effectiveness of the training programs in developing a high quality of performance in the training graduates.
3. To determine the most effective training program for achieving these aims.
4. To determine the distinguishing qualities and characteristics of a successful CCU nurse.
5. To determine the most effective and reliable methods for the selection of the "best" training applicants.

#### METHODOLOGY

A systematic model was designed to analyze the three interrelated primary spheres of concern:

1. *Input variables*: Trainees' demographic data, education, personality, expectations and attitudes towards CCU nursing, etc.
2. *Process variables*: Training Centers' facilities, approach, curriculum, etc.
3. *Output variables*: Knowledge gained, post-training expectations and attitudes, clinical performance both in-training and on-the-job, etc.

In addition, *Environmental variables* were studied. They consist of the sponsor hospitals' facilities, approach to nursing, etc. that influence both input and output.

The project staff then prepared, piloted, and revised nine data-gathering instruments. A standard personality test, the 16 PF, was also selected. This process involved discussions with PHS contract officers and with various consultants, visits to CCU's, a review of pertinent literature and existing research information, and an analysis of the content to be covered.

In general, data were collected on the trainees before and after training, and at follow-up, between three and four and a half months after training. Data were also collected on the training programs, and on the sponsor hospitals to which the trainees were returning after completion of the program. In terms of the specific problems addressed in this survey, two instruments are of special interest.

The *knowledge test* was especially designed and standardized. The test contained 12 weighted sub-tests, with each sub-test containing a number of weighted items. It was used both before and after training. The *performance checklist* was designed to tap the degree to which the training graduates performed specific CCU nursing functions at follow-up. Together, these instru-

ments constitute the basic evaluative data bank of the study.

The follow-up portion of the study was conducted in two ways: mailed questionnaires were sent out to all graduates of the programs, except as noted below. In addition, other questionnaires, including the performance checklist, were sent to their hospital supervisors. A systematic mail and telephone procedure assured a return rate of at least 90%. In order to monitor the reliability of the mailed returns, and to assess the effect of non-respondent bias, a 10% random sample of the graduates was selected for personal, on-site follow-up visits. The results of these visits confirmed the high degree of reliability in the mailed returns.

*Data-Collection*. The survey period extended from August 1968 to September 1969. In the eleven centers under study, a total of 57 sessions were monitored, for a total of 862 trainees. The 456 sponsor hospitals were all included in the survey.

*Data-Processing*. Standardized procedures were established for handling and coding of raw data. Data-processing was completed at the end of October 1969. A correlation matrix was designed and run on 85 variables.

#### FINDINGS

*Trainees*. The "typical" trainee was female (98%), the mean age for the group was 34 years, and the median was 28 years. About half of the trainees were, or had been married; of these, 60% had families, half of which consisted of 2 or more children. 83% had obtained a hospital diploma, 5% had an associate degree, 18% had a baccalaureate degree and 2%, a masters. Previous coronary care experience was as follows: 17% had worked in a CCU for an average of 8 months, and 36% had worked in ICU/CCU's for an average of 14 months. Most values of the 16 P.F. were close to the normal mean, except on the general intelligence scale when their mean was substantially higher than the mean of any occupational sub-group reported.

*Sponsor Hospitals*. Of the 456 hospitals surveyed, 55% had sent one nurse to training, 27% had sent 2 nurses, and the remainder 3 or more nurses. Hospital size varied considerably: 16% had less than 100 beds, 26% between 100 and 199, 21% between 200 and 299, and the remaining 37% had 300 or more beds. Results obtained before training showed that 27% of the hospitals had a separate CCU and 41% a combined CCU/ICU. These figures increased slightly at follow-up. The most surprising finding of the survey was the number of training graduates still not working in coronary care at follow-up. With a 90% response (N=779)

only 510 nurses were found to work in coronary care, or 65% of the follow-up population. The reason given in 65% of the cases was absence of CCU.

*Trainees' Preparedness upon CCU entry.* Evaluation of objective one was done on the basis of knowledge tests, trainees' attitudes, and supervisors' general ratings.

Out of a possible 220 points, the 862 nurses averaged a pre-score of 127.4 points, or 57.9%, which increased to 74% on the post-test. The variability of test scores at pre and post for the group indicates that instruction had brought about a leveling of test performance.

Trainees' expectations of CCU activities changed markedly between pre and post-training, with a general shift towards a "middle-of-the-road" attitude, indicating a tendency, over the training cycle, for attitudes to become more realistic, which is felt to be a positive and desirable training output. 63% of the trainees found the program excellent and 35% rated it as "good". When asked, after training, whether they would select CCU training again, 63% answered "definitely yes", and 25% "probably".

The following curriculum areas were mentioned at post-training as needing more preparation: Fluid and Electrolyte Balance (53%), Interpretation of ECGs (49%), Recognition and Treatment of Arrhythmias (35%). While these figures may have represented the trainees' anxiety at assuming new responsibilities, similar results were obtained among nurses working in CCU's at follow-up. Fluid and Electrolyte Balance was still on the top of the list, non-coronary complications were mentioned by 38% of the working graduates, basic electronics and interpretation of ECG by 37%.

Suggestions by supervisors regarding possible improvements in the programs agree in general with those of the trainees, the main one being for more stress on the technical aspects of CCU nursing. Rating the nurses' preparation on a seven-point scale, 77.5% of the supervisors selected the two top categories.

*Trainees' Performance in CCU's.* Performance checklists were received by each hospital, one for each trainee. A total of 487 checklists were completed and returned (56% of those mailed). By far the main reason given for non-completion of the form was "CCU not open".

Overall mean performance was rated from "good" to "excellent"; however, while the mean ratings do not deviate significantly from one another, there was a general tendency for nurses to be rated higher for technical, CCU-specific activities than for non-technical, general nursing activities. There is an apparent contradiction between these high "technical" ratings, and suggestions for more program depth in the same areas. It

would seem, then, that the sponsor hospitals view the major function of training as developing technical competence, while general nursing qualities are viewed as inherent in the potential trainee. This will be discussed later under objective four.

A fair comparison between nurses trained at different centers require that some allowance be made for skills the trainees brought with them on entering the program. In all cases, it was found that the centers rated highest in nursing performance had trained the most experienced population, while the lowest ratings were obtained by those centers having trained the least experienced group.

*Model program.* The study failed in providing an analysis of the model program, objective three of the study. Both dependent and independent variables displayed inadequate variance characteristics. Further, the training centers were quite similar, at least on the variables tapped by the instruments. This result was, of course, disappointing, but it should be noted that the basic reason for this failure is the success of the programs in fulfilling the overall objectives.

*Optimal characteristics of a CCU nurse.* The fourth objective was examined from the standpoint of high performance, satisfaction with CCU nursing, and motivation to continue work in coronary care.

Since performance ratings were typically either good or excellent, a detailed study was made of those performance items rated as "deficient" by the supervisors, yielding a picture of what a successful CCU nurse should *not* be, and inversely what characteristics she should possess. The largest number of deficient ratings were found in the broad area of "Communication and Interaction with Staff"; next in line was "Performance of day-to-day assignments" with stress on general nursing competence, and skill in handling and verifying the technical equipment; finally "Communication with the patient and his relatives". Thus, a successful CCU nurse would appear to need excellent nursing skills, an ability to relate well with the members of the CCU team, with the patient and his relatives, as well as technical competence.

73% of the training graduates working in CCU's stated that they definitely wanted to pursue coronary care as a specialty. A number of problems were expressed, however, the great majority stressing staffing difficulties, and lack of support and communication within the hospital in general, and the Unit in particular. A smaller number of nurses also expressed frustration at the occasional "dullness" of Unit work. Successful trainees derived great satisfaction from bedside nursing,

and from the challenge and diversity offered by coronary care work.

Since the follow-up period extended from 3 to 4 1/2 months, long-range tenure could not be ascertained. When asked about their plans for a two-year period, 61% of the nurses stated that they wanted to continue coronary care nursing.

*Selection criteria.* The main criteria used by the sponsor hospitals when selecting potential trainees were: motivation and interest in CCU nursing; stability in present position, and demonstrated excellence in general nursing skills.

The results of the study point to the necessity of providing the potential trainee with a clear perception of what her role will be, prior to selection. They also suggest a need for closer communication between training centers and sponsor hospitals, before, during, and after training.

### Evaluation of CCU Nurse Education in Washington and Alaska MARIELLA LARTER

In July 1969, the Subregional CCU Nurse Education Project of W/ARMP became operational. The goal of the project was to train 873 nurses per year in basic CCU, and to train them in sixteen (16) subregional centers rather than in a "core" or "Seattle based" setting. In the last year, all but one of the sixteen centers has become operational and an additional three communities have become subregional education centers. Each center plans its own objectives, curriculum, eligibility requirements, course length, and teaching methodology.

The plan as outlined in the following pages was developed by the Subregional Project staff in conjunction with the Office of Research in Medical Education of the University of Washington, Charles Dohner, Ph.D., director. From its inception the evaluation was to meet two goals: 1) to evaluate the impact of the project on regional CCU nurse training; and 2) to provide feedback to course instructors on the strengths and weaknesses of their courses and of individuals in them. The evaluation design at present involves measures of knowledge, attitude and skill. A patient care assessment tool is presently under development.

#### KNOWLEDGE TESTING

Practicing physicians and nurses from throughout the region were asked to submit multiple choice questions

relating to a set of regional objectives. A pool of over 800 test items were edited for content and format; the revised items were then rated by their authors as being either "essential", "desirable", or "supplementar knowledge for a CCU nurse. Only those items rated "essential" or "highly desirable" were retained, leaving pool of 250 questions weighted in the following fashion:

CCU Concepts	1)	
Anatomy and Physiology	3)	Summarized as CCU Concepts
The Classic MI	8)	
Diagnostic Tests	5)	
Rehabilitation	8)	
Complications of an MI (excluding arrhythmias)	13)	Summarized as Complications
Electrocardiography	8)	
Equipment and Safety	3)	Summarized as Arrhythmias
Arrhythmias	19)	
Chemical Therapy	13)	Summarized as Chemical Therapy
Other Therapy (i.e. pacing, resuscitation)	19)	Summarized as Other Therapy

#### *Evaluation of CCU Nurse Education*

Test items were randomly assigned to version A or B of the exam. Each exam is equally weighted by content but the individual questions remain different. After field testing on student nurses with no CCU background and on graduates of a USPHS five-week CCU nursing course fourteen items from each version were eliminated. After item analysis of the test results of 200 nurses involved in subregional courses was completed, eleven additional items were deleted from each version.

Test A and Test B now contain 100 items each, and are of equal difficulty according to standard statistical measures. In addition to answering each question with what she supposes to be the correct answer, each nurse is asked to rate her certainty about that answer on a scale from one to three, or absolute certainty to guessing. The computer summary of her scores then computes not only how many questions she answers correctly, but also how many questions she was certain about, and how many which she says she was certain about that she actually answers correctly.

We can thus measure with our instruments three areas of potential change from pre to post course: 1) change in knowledge (right-wrong score); 2) change in expressed certitude and guessing; and 3) change in ability to evaluate her knowledge about CCU nursing. These three

factors are recorded for the overall exam as well as for five content areas within the exam for each student and each class.

The instructor can go over with each student her areas of greatest knowledge or growth in knowledge, her expressed confidence, and whether that confidence is well founded or not. She has considerable data about the continuing education needs of her students at the completion of her course as well.

The regional mean for the pre test (Test A) is 32%. The regional mean for the post test (Test B) is 53%. The reliability coefficients of these exams, when administered to a group of ten or more individuals, range from 0.89 to 0.95.

### ATTITUDE TESTING

A standard semantic differential scale is used to evaluate the attitudes of nurses on ten concepts in CCU nursing. Those concepts include: coronary care nursing; coronary heart disease; cardiac monitoring; cardio-pulmonary resuscitation; doctor-nurse relationships; change in nursing; independent nursing decisions; patient individuality; patient teaching; and emergency situations. One other concept, "death", was eliminated because of possible disagreement about what a desirable change in attitude would be at the completion of a course.

The attitude scales are filled out on an anonymous basis at pre and post tests; results are scored by sub-region. Analysis of results in the first formal evaluation period (September 1, 1969 - February 1, 1970), in which time over 500 nurses participated in Subregional courses, revealed that overall there was not a significant change in attitude from pre to post course. Individual courses showed significant attitude changes in one or more "concept" areas; in discussing some of the puzzling attitude shifts with course instructors, it was learned that the majority of these shifts reflected an attitude consciously conveyed to the students by the instructor.

#### *Evaluation of CCU Nurse Education*

Future planning calls for use of attitude measure at pre course and at six months after course completion. Knowledge and attitude scores have thus far showed no correlation; if this lack of either significant attitude change or correlation between attitude score and knowledge continues, a new strategy to measure this area will be devised.

### SKILL TESTING

A skill test was designed to evaluate the functioning of nurses when presented with simulations of clinical emergencies. The testing involves an evaluation of psychomotor abilities as well as the rationale for initiation of certain therapeutic measures. The skill test is designed to be administered in a mock-up setting using a standard hospital bedside area, an arrhythmia anne resuscitation doll, a bedside monitor, a defibrillator, and standard emergency equipment (i.e. suction, medications).

In initial field testing, nurses suggested the following:

1. that they be in uniform when tested, and
2. that the evaluator "role play" as a new orientee to a coronary care unit, rather than assume a strictly observational and judgemental pose.

Taking these suggestions, a group of 28 nurses from both metropolitan and rural hospitals were evaluated using this tool in their own clinical setting. The range of scores was 8 to 30, out of a possible 32 points. The mean score was 22 points.

The evaluator summarized her conclusions regarding initial use of the tool as follows:

1. Greater consideration needs to be given to the standing orders under which a nurse functions in a given agency; accepted therapy for nurses to initiate varies greatly from agency to agency.
2. No more than one agency can be evaluated on a given day in view of unit pressures, staffing, and patient census; the cost of sending an evaluator any distance is considerable unless other duties can be performed concurrently.
3. It is very difficult to remain neutral even when involved in role playing; there is a constant temptation to correct errors and teach during the testing.
4. The nurses tested need to be thoroughly familiar with all equipment used in the testing situation; thus, hospital equipment or a like brand must be brought to each testing site.
5. Many nurses responded appropriately to situations but for the wrong reasons.
6. A weighting scale needs to be further refined so that there is a greater spread in scores and differentiation between levels of performance.
7. The skill test is an excellent teaching tool but needs further revisions to increase its effectiveness as an evaluation instrument.



## Evaluation of CCU Nurse Education

Future plans call for the random skill testing of nurses at the completion of basic courses utilizing equipment they have used in mock-up drill sessions.

### OTHER EVALUATION TOOLS

A. A personal profile sheet revealing 17 pieces of demographic information about each nurse is filled out at the completion of courses. Correlations are being run between these 17 variables and combinations of variables compared with pre test score, post test score, expressed certitudes, accuracies, and changes in score. Data will be available soon.

B. Use of chi square measure in conjunction with certitude score has been employed by the School of Medicine. A high and significant correlation was found between a low chi square and the overall knowledge of the students tested. This measure is being incorporated into the CCU nursing data analysis.

C. A patient care assessment tool is under development. It is hoped that this tool can be used to demon-

strate changes in patient care as a result of post-graduate learning experiences for nurses.

### Additional Materials Available on Request:

1. Objectives for nurse training upon which knowledge tests are based.
2. Sample computer printout and explanation of data contained in it.
3. Copies of Test A and Test B for review. (Copyrighted material - not to be retained or duplicated in any fashion)
4. Answer sheet incorporating certitude measure.
5. Attitude test.
6. Sample of results of attitude testing returned to course instructor.
7. Skill test.
8. Personal profile sheet.

Write: MARIELLA LARTER, R.N.  
Subregional CCU Project W/ARMP  
180 "U" District Building  
1107 N. E. 45th Street  
Seattle, Washington 98105

### CORONARY CARE NURSING EXAM. . .CONTENT AREAS RELATIVE WEIGHTING, OBJECTIVES

#### CCU Concepts. . . relative weight I

- a: Synthesizes a concept of intensive coronary care in relation to its implications for the professional nurse
- b: Values the necessity for assuming responsibility and self-direction for continued learning in CCU nursing.

#### Anatomy and Physiology. . . relative weight III

- a: Comprehends basic anatomy and physiology of the cardiovascular system
- b: Interprets significant inter-relationships between the cardiovascular, pulmonary, renal, and nervous systems
- c: Interprets significant concepts of stress.

#### Uncomplicated Acute Myocardial Infarction. . relative weight VIII

- a: Synthesizes a concept of coronary artery disease in relation to its implications for professional nursing care.
- b: Develops a systematic approach to the assessment of the individual patient's status upon admission and in subsequent days of hospitalization.

(content areas include: epidemiology; pathophysiology of coronary heart disease; physiologic stress responses; psychology of life-threatening diseases; history and classic signs and symptoms of acute MI; cardiac ischemia as it relates to relief of pain, anxiety, and administration of oxygen; dietary modifications, activity restriction, fluid balance; planning individualized care)

#### Diagnostic Tests. . . relative weight V

- a: Analyzes the major diagnostic tools used in the diagnosis of coronary heart disease in terms of their implications for planning nursing care.

- b: Evaluates the techniques used in the physical and psychological preparation of the patient for diagnostic tests. (content areas include: history and physical; serum enzymes, ESR, WBC, temperature elevation; circulation time; chest X-ray; serial EKG's; heart and breath sounds; vital signs, CVP, jugular veins, urine sp. gravity, I&O; nursing care plans related to scheduling of tests; teaching plans to minimize fear, discomfort, emergencies)

#### Complications of an Acute Myocardial Infarction (excluding arrhythmias) relative weight XIII

- a: Applies the problem solving method to the identification and treatment of the complications of coronary heart disease:
  1. congestive heart failure
  2. cardiogenic shock
  3. acute pulmonary edema
  4. pulmonary-systemic emboli
  5. pericarditis
  6. cardiac rupture
  7. cardiac arrest
  8. extreme emotional reactions

#### Electrocardiography. . . relative weight VIII

- a: Synthesizes basis principles of electrocardiography to serve as a basis for the evaluation of cardiac status of the individual patient  
(content areas include: electrophysiology; hemodynamic vs electrical properties of the heart; depolarization and

repolarization of the myocardium; correlation of the electrophysiology of the heart with the electrocardiographic tracing; basic principles of polarity, amplitude, and configuration of the PQRS in terms of lead axis and cardiac vector)

Equipment and Safety . . . relative weight III

a: Applies fundamental principles of electrocardiographic techniques to achieve maximum effectiveness and safety of electrocardiographic monitoring and twelve lead equipment.

(content areas include: grounding; monitoring capabilities and limitations as opposed to the standard EKG; essential features; purposes and standards of electrocardiographic equipment; interference and means of eliminating it; how to use standard monitoring equipment; safety for staff and patients with monitoring equipment)

Arrhythmias . . . relative weight XIX

a: Applies the problem solving method to the identification and treatment of the complications of coronary heart disease, specifically cardiac arrhythmias.

b: Evaluates alterations in the electrocardiographic rhythm strips and rhythms displayed on the oscilloscope according to their significance to the patient's total condition and their implications for medical and nursing therapies.

c: Develops a systematic approach to the interpretation of arrhythmias.

d: Utilizes the problem solving method in the treatment of arrhythmias.

(content areas include: arrhythmias by site of origin, effect, treatment, and implications for nursing care)

Chemical Therapy . . . relative weight XIII

a: Develops a systematic approach to the classification, analysis of, rationale for, and the nursing implications involved with chemical therapies in the treatment of coronary heart disease and the frequently encountered complications.

Other Therapy . . . relative weight XIX

a: Appreciates the nurses role in the early recognition and treatment of conditions that may precede life threatening conditions.

b: Appreciates the importance of effective habit patterns in the handling of emergency situations.

c: Appreciates the importance of frequent review and continued refinement of emergency procedures.

d: Develops a systematic approach to the identification and treatment of cardiac emergencies.

e: Differentiates the nurses' responsibilities in elective cardioversion and the preventive use of pacemakers, as opposed to the emergency situations involved with these therapies.

f: Utilizes the problem solving method to determine priorities in nursing care in the post-resuscitative period.

Rehabilitation . . . relative weight VIII

a: Develops and communicates a nursing care plan that incorporates preventative, therapeutic, and rehabilitative aspects.

b: Evaluates the patient's CCU experience in relation to his total life situation.

c: Determines implications for the planning of comprehensive nursing care.

d: Values the role of the professional nurse in the health team, especially in relation to her potential contributions regarding the individual needs of the patient and family and continuity of care into the post hospitalization phase.

e: Reviews select basic nursing knowledge and skills in the light of their implications for the patient with coronary heart disease.

(content examples—vital signs, pulses, tracheal suctioning, oxygen administration, respirators, patient positioning, venipuncture, IV therapy and administration, rotating tourniquet, skin care, passive exercising)

Summary of Content Areas and Relative Weighting on both pre and post Tests:

CCU Concepts:	1/100
A - P:	3/100
Classic MI:	8/100
Diagnostic Tests	5/100
Complications:	13/100
Electrocardiography	8/100
Equipment & Safety	3/100
Arrhythmias:	19/100
Chemical Therapy	13/100
Other Therapy:	19/100
Rehabilitation:	8/100

# A SYSTEMS APPROACH TO CORONARY CARE EVALUATION

## Participants

Morton Robins - Moderator  
Acting Chief, Study Design and Analysis  
Staff  
Regional Medical Programs Service

M. A. Rockwell, M. D.  
Director, Rand Health Program  
Rand Corporation

---

## A Study of Coronary Care Unit Effectiveness

M.A. ROCKWELL

This report describes a continuing project conducted by The Rand Corporation for the California Committees on Regional Medical Programs (CCRMP) to measure the operational effectiveness of coronary care units. During the past two years the project, which began as a feasibility study, has become a community action project involving more than 100 hospitals. This report traces the evolution of the study from its initiation up to the present, describes what has been accomplished, and outlines future objectives.

Our study is based on the belief that every CCU should continually monitor its performance. Data should be collected describing patients admitted to the unit, how rapidly they reached the CCU following their onset of symptoms, their clinical course and treatment during their CCU stay, and their clinical course and treatment during their CCU stay, and their discharge status. Collection and analysis of such data is necessary to ensure that the unit is performing effectively.

In 1968, the CCRMP found that most CCUs were trying to collect and analyze such data but many of the units were having problems in their data collection. First, development of the necessary data collection forms and procedures proved to be too difficult for many units. Second, many CCUs soon collected such a large volume of data that it could not be analyzed by manual techniques but required computer methods. Most units did not have access to the necessary equipment and expertise. Third, once the data was collected and analyzed, it was often difficult to interpret because there was no standard against which to compare the results. It seemed desirable to allow each CCU to compare its results with those of similar hospitals. Such comparisons, however, required data collection and analysis procedures to be standardized, a task obviously beyond the capability of an individual CCU.

The CCRMP, aware both of the importance of collecting performance data in CCUs and the difficulties experienced by many units in collecting such data, embarked upon development of a standardized data collection and reporting system for CCUs. In December 1969, a contract was given to The Rand Corporation to develop a prototype system and test its feasibility. Medical guidance of the project was provided by the CCU Steering Committee of the CCRMP.

During the past two years, a prototype data collection form has been designed, tested and revised. On January 1, 1970, a prototype data collection system became operational and participation in the study was opened to any California CCU that wished to participate.

The current system requires that about 100 items of information be reported on each acute myocardial infarction patient admitted to the CCU (only 10 items of information are collected on non-MI patients). The data forms are mailed to The Rand Corporation where they are keypunched. Every three months the keypunched data are processed by computer to produce summary reports. Each hospital receives a 15-page report describing the patients admitted to the unit and the outcome of their hospitalization. Each unit can compare its experience with that of the participating group as a whole.

Preliminary indications are that the data collection system has become an important part of the CCU operation in many hospitals. Although participation in the study is voluntary, the number of participating hospitals reached 120 by June 1970. Thus, about two-thirds of California's CCUs are now involved in the study. In addition, units from the Washington-Alaska RMP, the Northern New England RMP and Missouri either have, or are soon expected to join the study.

We believe that the study has had an important and beneficial effect on CCU effectiveness. First, it has helped some CCU directors improve the operation of

their units by, for example, finding ways of speeding the patient admissions. Second, periodic summary reports have served as a focus for teaching conferences for CCU physicians and nurses. Third, data collected by the system have helped the CCRMP assess the effectiveness of their nurse training program. Fourth, data collected by the system should make it possible to investigate several

ways of reducing the cost of CCU care without compromising its quality. These include: (1) using specially trained CCU technicians to supplement nurses in the units, and (2) using automated monitoring equipment to eliminate the requirement for continuous surveillance of ECG monitors.

# EVALUATION OF INSTRUCTIONAL TECHNOLOGY PROJECTS

## Participants

James E. Dyson, Ph.D. - Moderator  
Director, Continuing Education Division  
Colorado-Wyoming Regional Medical Program

Cecilia C. Conrath  
Chief, Continuing Education and  
Training Branch  
Regional Medical Programs Service

James Barrett, Ph.D.  
Continuing Education Division  
Colorado-Wyoming Regional Medical Program

M. Gene Aldridge  
Continuing Education Division  
Colorado-Wyoming Regional Medical Program

William Engbretson, Ph.D.  
President, Governor's State University

---

## Summary of Session

CECELIA CONRATH

The workshop session on Instructional Technology was developed by the Colorado/Wyoming RMP. The objectives of the session are given below in order of priority.

1. To learn interests and needs of workshop participants for help in evaluation using instructional technology.
2. To help participants learn functions of various types of instructional technology, approaches to evaluation of such technology and relative effects of various approaches.
3. To present information on effective evaluation procedures.
4. To develop an awareness of consultation/referral resources nearby within region and on an inter-regional basis.

The whole idea was to show how questions and concerns can be quickly identified, how resources can be located and used effectively, and to demonstrate that the basic strategy of evaluation grows out of the needs of the participants.

The session opened with a brief statement of the status of instructional technology within RMPS by the Chief of the Continuing Education and Training Branch followed by an outline of the session by Dr. James Dyson, Associate Director of Continuing Education, Colorado/Wyoming Regional Medical Program. A problem census of interests and needs of participants was conducted by Dr. James Barrett of the Colorado/Wyoming staff.

While Dr. Barrett interviewed participants a written recording of the answers was projected on an overhead projector by Gene Aldridge also of the Colorado/Wyoming RMP staff. This enabled a running inventory to be kept in front of the participants as the session progressed.

At the conclusion of the problem census a long distance telephone conference was held with the following consultants:

William J. Paisley, Ph.D.  
Director, ERIC Clearinghouse on  
Educational Media and Technology  
at the Inst. for Comm. Research  
Stanford University  
Stanford, California 94305

Elizabeth Norman, Ph.D.  
Associate Professor of Nursing  
College of Nursing  
Northeastern University  
Boston, Massachusetts 02115

Rick Breitenfeld, Ph.D.  
Executive Director  
Maryland Center for Public Broadcasting  
Owings Mills, Maryland 21117

Gerald W. Gaston, D.D.S.  
OSRMP-CAI Project Supervisor  
The Ohio State University  
Columbus, Ohio 43210

David L. Bell  
Box 488  
Altadena, California 91001

The results were not entirely successful because of the small attendance at the session. This technique is productive with a minimum of 10 and upward in an almost unlimited number. There were only 5 participants and two left early.

Issues concerned with cost effectiveness of different media, adaptability and conversion from one modality to another, and status of evaluation research were brought up during the conference call. Technical prob-

lems, i.e. temporarily losing California participants and poor voice transmission interfered with the reception.

Gene Aldridge assembled kits of material on evaluation of instructional technology and learning theory. Bibliographies on the general field of learning, teaching with films, guides for TV teachers and considerations for judging audiovisual presentation standards were among materials distributed.

# EVALUATION OF PHYSICIAN EDUCATION

## Participants

Daniel Fleisher, M.D. - Moderator  
Director of Health Professions  
Temple University

William B. Munier, M.D.  
Staff Assistant, Continuing Education  
and Training Branch  
Regional Medical Programs Service

---

## Summary of Session

WILLIAM B. MUNIER

The objective of the Special Interest session on Evaluation of Physician Education was to increase the knowledge of the participants about the essentials of sound educational projects. The methodology employed was that of active involvement of participants in deciding on what constituted a sound project. No evaluation of learning was planned.

Specifically, three surrogate projects were presented, one at a time. In each case, desirable and undesirable aspects were listed, as volunteered by the participants following review of the projects. Explanation and analysis of the projects was led by the moderator, Dr. Daniel Fleisher of Temple University, Philadelphia, assisted by Dr. William Munier.

Two of the projects were poorly designed, and contributed the bulk of the undesirable aspects. One of

the projects was very well constructed and contributed the majority of desirable aspects. Following critique of all three, a fairly complete list had been developed of what constituted an effective project. It had been developed by the participants themselves following careful analysis of three projects representative of actual RMP grant requests.

It was felt that the active involvement of the people attending the session was more likely to increase their knowledge than would a didactic presentation. The actual proceedings at the session involved active debate concerning which aspects were good and which were not. Errors in judgement by a given participant — from the moderator's point of view — were quickly lampooned by others. The resulting list at the conclusion of the conference was educationally quite sound. Insofar as no evaluation of learning was planned, the product of the session was good, and all present participated actively, the conference was subjectively judged a success.

# EVALUATION OF MULTIPHASIC SCREENING

## Participants

Donald N. Logsdon, M.D.  
Associate Director, Multiphasic Health  
Screening Center  
Department of Community Health  
Brookdale Hospital Center

Frank R. Mark, M.D.  
Chief, Operations Research and Systems  
Analysis  
Regional Medical Programs Service

---

## Evaluation of Multiphasic Health Testing

DONALD N. LOGSDON

In the chapter entitled Evaluating the Quality of Medical Care by Avedis Donabedian from the recent book *Presymptomatic Detection and Early Diagnosis* by Shark and Keen, the conclusion is reached that "although the assessment of the quality of medical care remains difficult and imprecise, there are several ways in which one may arrive at judgment sufficiently valid for a variety of administrative decisions". Among the ways suggested were "studies of the effect of greater precision and detail in standards on the reliability and validity of judgments (measurement)". As applied to MHT the current operating programs have attempted several evaluation studies which I will briefly describe and comment on.

Dr. Matthew Tayback, in several meetings sponsored by the U.S. Public Health Service in 1967-68, set forth criteria which he suggested for determining the value of Multiphasic Health Testing. He restated the proposition that evaluation should rest on the success of attainment of project objectives, namely, (1) per cent of target sample reached (2) precision and accuracy of individual measurements (Quality Control) (3) yield of screened positives per major procedure (4) per cent of screened positives who make contact with personal physician, and (5) per cent of screened sample with minimum significant benefit in health knowledge due to MHT. Although it is highly pertinent and eventually critical to consider cost-benefit characteristics or end results of MHT, such data will not be forthcoming for several years. In the meanwhile MHT technology needs to be advanced on the principle of its cost-effectiveness and its capability to efficiently process large populations.

Tayback considered establishment of a multiphasic screening (testing) service to be based upon the following operational model.

Therefore, evaluation of MHT projects funded through NCHSR&D should proceed on three levels.

### ACHIEVEMENT OF TECHNOLOGICAL OBJECTIVES

PHS was at that time proceeding on the assumption that MHT is basic to the attainment of a national health objective - periodic assessment of the health status of each adult, 35-69 years of age. The system must undergo continuous improvement resulting in added validity of the health testing and in improved cost effectiveness. Specifically, it was recommended that the achievement of technological advances in MHT be measured by the completion of defined tasks and with time specified end-points. During the twelve month period, January 1, 1969 - December 31, 1969, the following tasks were to be initiated and progress reports submitted by the end of the period. These tasks are not a complete description of the technical problems which need solution.

#### *Glucose Tolerance Test*

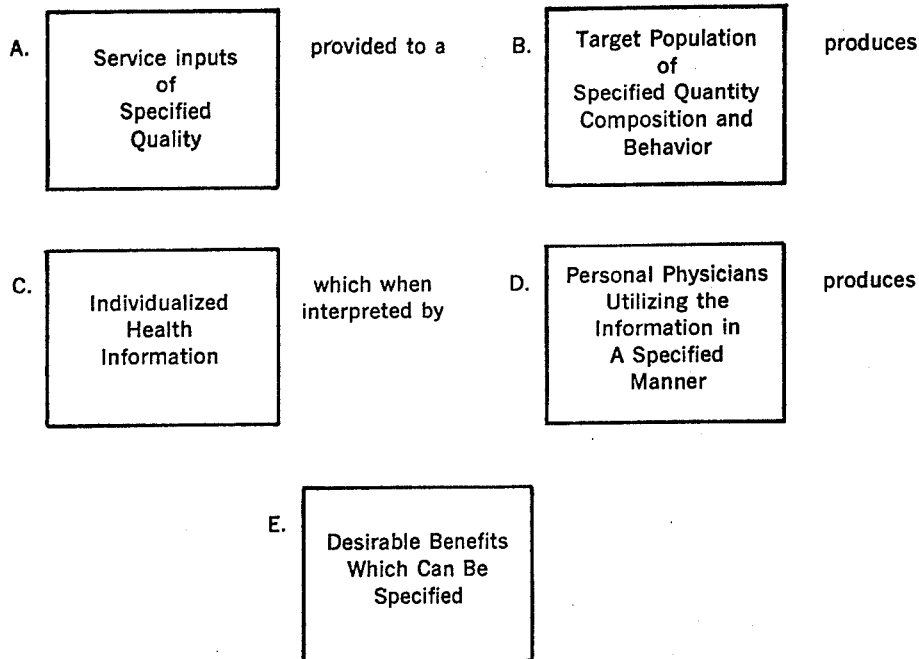
It is imperative to determine the relationship which exists between the result of the abbreviated glucose tolerance test as employed in MHT and standard oral glucose tolerance test as performed in conventional hospital or private laboratory centers.

The effect of time of day on the abbreviated tests must be clarified.

#### *Standardization of Norms*

Interpretation of results from clinical lab tests, including blood chemistry, hematology, and non-lab tests, such as spirometry, by the practicing physician is difficult when the normal population ranges for a specified measurement is not given in the report.





A standard procedure for reporting MHT results should be adopted. The exact measurement obtained should be reported and the normal range for the age, sex and ethnic group category given.

Since the distribution of defined measurement by age, sex, and ethnic groups has not been determined, this should be developed as soon as possible.

#### *Standardization and Documentation of Computer Programs*

Inefficiency (excessive cost) is generated by failure to develop systems which can be replicated with minor adjustments.

Existing computer monitoring of SMAIZ and VCG interpretation needs to be validated with a view towards selection of a standard program so that widespread use can be made of the standard programs with minimum further investment for software development in these specified areas.

#### *Cost Analysis*

Major components of MHT needed to be defined. Each component must then serve as a unit for the determination of cost.

Cost analysis data should be generated within the next 12 months.

In view of the limited staff, this task should be accomplished through a contract negotiated with an interested and competent cost analysis service.

#### *Quality Control*

Each major component of MHT requires a protocol for establishment and control of quality of measurement and test information generated. Each project should develop a manual of procedures in respect to quality control. PHS should then produce a standard manual on quality control and annually update this document.

#### **ACHIEVEMENT OF PROGRAM OBJECTIVES**

Pending demonstration of benefits relating to reduction in disease, disability and age specific death rates, MHT must receive process evaluation on the basis of the attainment of program objectives and the cost-effectiveness of services.

Such process evaluation will be possible by the following strategy:

1. Each project should set forth the target population it seeks to reach with its screening program and should specify the fraction of the target population which it proposes to reach.

2. Periodically (quarterly) demographic characteristics of the screened population were to be reported to PHS. Comparison of 2 with 1 will indicate the extent to which the target population is reached. The minimum set of variables for which information is sought should include age, sex, race, income, occupation, source of regular medical care, utilization of medical care within past six months, date of last general physical examination, and follow-up results.

3. The yield of significant positive findings per examination procedure for age, sex and ethnic groups will permit assessment of the cost per abnormality detected.

4. The patient-physician contact ratio by major screening classification type is a measure of the extent to which significant screening findings receive follow-up exams.

5. A survey of selected classes of the screened population, prior to and following the date of visit to the screening center could provide suggestive clues relative to the health attitude and knowledge of consumers and could provide information of consumer reaction from a single exposure to MHT (a consumer study).

6. A questionnaire survey of physicians, who have received MHT reports, with a view towards determining their attitudes and knowledge of the usefulness of MHT would complement information obtained through 5.

#### ACHIEVEMENT OF LONG RANGE OBJECTIVES

Control of clinically significant chronic disease and prolongation of life are end results of MHT which can be demonstrated only by ambitious research involving careful prospective follow-up of large samples of adults over long periods of time. Such a project is under way at the Kaiser-Permanente medical service in and around San Francisco.

It was recommended that aside from Kaiser-Permanente, no extensive investment of funds should be made at this time to demonstrate long term effects of MHT.

#### PROCESS EVALUATION

The Brookdale Hospital, Multiphasic Health Screening program has been a successful demonstration of AMHT. The questions now to be answered are: What difference does MHT make in the delivery of health services in an urban environment? Can MHT become an effective component in a primary health care system? It was established by PHS support to test the feasibility of operating a MHT program in an "open" medical care system with an adjacent poverty population.

*In order to answer these questions it would be necessary to "close the information loop" by establishing a Follow-up Clinic which would have enabled the program to accomplish the following:*

1. Provision of the necessary follow-up medical evaluation and management for the screened poverty population.

2. Validation of the screening results by comparison with results of diagnostic studies for the poverty population by Follow-up Clinic physicians.

3. Documentation and evaluation of the experiences with this type of health service as compared with the existing health services of the Hospital Ambulatory Care Program. A central record system would enable monitoring of the two types of care.

4. Further utilization of paraprofessional personnel and instrumentation in health care. The use of physician assistants, nurses, technicians plus hardware can be tested.

The above factors are considered important in assessing the difference MHT makes in the delivery of health services. It is recognized that the addition of a Follow-up Clinic would not alone provide an answer to the question of benefits in terms of biologic outcome or end results. However, as the methodology for this type of evaluation is adequately developed, a prospective longitudinal study of morbidity, mortality, and disability could be attempted.

*We began to evaluate MHT at BHC as part of a primary health care system at intermediate points and to determine feasibility of assessing end results.*

In December 1969 a subcontract was signed and work begun for biostatistical retrieval and analysis of the data on the 14,000 screenees processed at the Brookdale Hospital Center MHS program from the beginning in February 1968 through October 1969. Initially, the data were examined in terms of frequency distribution for continuously distributed quantitative variables by age, sex, and ethnic background. Dichotomous qualitative variables were tabulated and percentage positives calculated also by age, sex, and ethnic background. Measurements of central tendency and variation were performed on continuously distributed measurements. This included mean, standard deviations, median, 5 and 95 percentiles. The number and percent of screenees with clinically significant overt and occult abnormalities based on currently acceptable criteria was determined also by the variables of age, sex, and ethnic background. Investigations will also be made for correlation analysis, i.e., history vs. test results, and screening results vs. physician diagnosis. This effort has been successful for Brookdale MHS and should have application to other demonstration programs in MHT.

Problems of data retrieval and analysis include:

1. Quality of input-measurement and keypunching errors.

2. Storage on historical tapes, i.e., completeness and documentation

3. Retrieval - group intervals, criteria of normal, abnormal.

4. Analysis - Mean or median, standard deviation or percentile, test of significance.

An economical evaluation can be approached by cost per test as indicated in the cost finding protocol for the past project year. The SRI method is being tested.

Effectiveness is being evaluated in terms of the yield of unknown and uncontrolled occult and overt conditions detected at the MHT Center. This, of course, is related to the prevalence of disease in the target population. High yields are expected for certain conditions in poverty populations, groups, due to prevalence and the lack of adequate medical care, e.g., rates for hypertension and hypertensive heart disease. The methodological problem of determining "unknownness" can be solved by the use of questionnaire information from patients rather than from M.D.'s. Efficiency is being calculated on the basis of the cost per positive screening test and cost per valid diagnosis. Of course the latter is dependent on adequate follow-up reporting. These evaluation efforts should be performed as the program activities are carried out. Simultaneously, the end-result evaluation is being explored for feasibility in an environment which prohibits randomization into study and control groups for longitudinal study. Present plans could include labeling a sample of the screened poverty population for monitoring over time and comparing their experience in morbidity, mortality, and disability with non poverty population and/or national statistics for the same age, sex, and ethnic group.

For preliminary results for total population see Appendix I.

*Determine the cost of MHT in a Primary Health Care System:*

1. The Brookdale MHSP, as a result of the SRI Cost Finding Study of AMHT has begun to examine the cost of the total program. Information is being collected on the total expenditures for this program through the Brookdale Hospital Center business office. However, it is apparent that a true cost analysis of this program will require the establishment of bookkeeping - cost accounting procedures separate from the Hospital System in order to identify the various costs involved. In addition to the usual items that are included under direct cost, it will be necessary to itemize those costs involved in recruiting the target population to utilize the facility and the follow-up activities.

2. In regard to the latter, the input for a time-effort study has been built into the computerized module for support of follow-up activities. These components must be costed out in a AMHT in addition to those items included in the recent reports on costs of the Kaiser-Permanente MSP in the New England Journal of Medicine. These additional items will obviously increase the cost of AMHT programs involved in motivation and follow-up and the question to be answered is how much.

In addition, a cost effectiveness report can be prepared for the follow-up activities wherein comparison of costs for furnishing the follow-up services using alternative methods will be used with the objective of being able to minimize the resources expended and maximize the number of individuals receiving medical follow up.

*Investigate Consumer and Physician Reaction to AMHT:*

1. During the 1969 project year 50% of the individuals screened resided in the Hospital's core area, and 25% were Black or Puerto Rican. A number of techniques for increasing registration from the hard core high priority areas were tested. Good progress has been made, but it is apparent that "hard" data on the behavior factors are necessary to improve performances beyond this point.

Similar considerations are involved in improving the 70% figure for successful follow up.

The data generated by the screenee process must be evaluated by a physician in the context of his examination of a specific individual. The physician's knowledge of and attitude toward AMHT therefore becomes of central importance.

The staff of the Brookdale MHSC is actively engaged in assisting the research group at Columbia University School of Public Health and Administrative Medicine in the development and implementation of two relevant studies: Consumer Reaction to AMHT #HSM 110-69-212, and Physician's Attitude Toward and Acceptance of AMHT #HSM 110-HSRD-57 (9).

2. The Physician Attitude Study is designed to determine:

- a. What are the social and psychological factors which affect the physician's cooperation with, his acceptance of, and his behavior concerning the MHT at Brookdale, including those factors which facilitate his utilization and acceptance of the service, as well as those factors which are barriers to effective utilization?

- b. What factors differentiate those physicians who actively participate and accept the screening program from those who do not, and what factors differentiate those physicians who change their attitude and behavior concerning AMHS?
- c. How can an automated system such as multiphasic screening be made more useful and acceptable to the practicing private physician?
- d. How do physicians adapt their practices to an automated health testing program?
- e. What inferences can be drawn from this specific study to the more general area of physician's response to automation in medical practice and what impact does it have on medical practice?

In order to perform this type of study the project must actively engage the support of the local Medical Society. This can present a difficulty.

The study of physicians' reactions to automated multiphasic health screening presently provides for an initial and a follow-up survey 10 months later of 1200 physicians in Kings County, New York. The two interviews will determine their attitudes, knowledge, and utilization of the Automated Multiphasic Health Screening Center at Brookdale Hospital in Brooklyn, New York.

The re-interviews were intended to concentrate on changes in attitudes, behavior, and perception of automated screening resulting from exposure to the program. When the study was planned it was anticipated that at the time of the first interview, at least half of the doctors would in the interim become exposed to the Brookdale program and, as a consequence, alter their image of it.

However, results of the first wave of interviews indicate that diffusion of the screening program has occurred more rapidly than anticipated. This fact has bearing on the timing of the re-interviews and in part motivates this suggested modification.

Of the first 712 completed interviews, only 101 physicians have not been exposed to the program (86% were exposed). There is no reason to expect that the remaining interviews will show much departure from this 7:1 ratio. Therefore, we cannot expect dramatic changes between the first and second interviews as a result of contact with the program. Some early results from interviews of doctors follows:

In your opinion, did the summary contain more information than was necessary, just about the right amount of information, or not enough information?		
More information than necessary	123	39%
About right amount of information	133	42%
Not enough information	59	19%
	<u>315</u>	<u>100%</u>

How easy was it for you to follow the general layout of the summary? Was it very easy, fairly easy, somewhat difficult, or very difficult?		
Very easy	111	34%
Fairly easy	112	34%
Somewhat difficult	70	22%
Very difficult	32	10%
	<u>325</u>	<u>100%</u>

In your opinion, should the normal range of results be indicated on the summary?		
Yes	290	90%
No	31	10%
	<u>321</u>	<u>100%</u>

Was the blood glucose test and result clear to you?		
Yes	280	90%
No	30	10%
	<u>310</u>	<u>100%</u>

Was the histogram arrangement of the hearing test results clear to you?		
Yes	186	63%
No	111	37%
	<u>297</u>	<u>100%</u>

How useful was the medical history questionnaire? Was it very useful, somewhat useful, not very useful, or worthless?		
Very useful	34	12%
Somewhat useful	92	34%
Not very useful	92	34%
Worthless	55	20%
	<u>273</u>	<u>100%</u>

What did you think of the fraction arrangement of positive responses by body system? Did you think this was a good way of presenting the medical history information or not a very good way?		
Good way	127	50%
Not a very good way	127	50%
	<u>254</u>	<u>100%</u>

How helpful did you find this reference manual in reading the patient summary? Did you find it very helpful, or somewhat helpful, or not at all helpful?

Very helpful	75	44%
Somewhat helpful	74	43%
Not at all helpful	23	13%

172 100%

Now about what you think *should* be done by screening programs like Brookdale's. Do you think that a screening program like Brookdale should be free of cost to examinees or should there be a charge?

Should be free	152	50%
Should be a charge	152	50%

304 100%

Do you think that a screening program like Brookdale should refer both normal and abnormal patients for follow-up by a physician or only patients with some positive condition?

Both normal and abnormal	241	67%
Positive condition only	103	33%

344 100%

#### *Clinical Laboratory Quality Control Studies:*

After several attempts over two years, there has been relatively little success in providing assistance to the Clinical Labs in AMHT for developing a sufficient program of quality control. The Clinical Chemistry Section, NCDC, has repeatedly demonstrated their interest in providing this support, but various bureaucratic delays have prevented any progress. The problem of assisting these labs remains, and a modest beginning is proposed for the next project year.

This effort would initially consist of a six-month study and evaluation of AMHT interlaboratory standardization utilizing the Brookdale Hospital Clinical Lab as a starting point, and then extending the protocol to include the other AMHT labs. The brief outline that follows describes the activities and resources required:

Study and Evaluation of AMHT Interlaboratory Standardization:

Preparatory efforts — Brookdale AMHT and NCDC through individual and group interaction.

- Develop recommendations for reference methodology, enzyme units.
- Anticipate problems in SMA technology and calibration.
- Design and prepare Multiphasic Text Panel for the elucidation of methodologic, technical, and calibration problems. Check stability of materials.

Pretest in local laboratory. Example: cholesterol study.

- Design general outlines of AMHT internal quality control system: calibration, serum monitor, laboratory responses; design a system of external evaluation.

#### OTHER STUDIES OF MULTIPHASIC HEALTH TESTING

1. HIP - Utilizing MHT to define the health status, practices, and attitudes of a defined poverty population covering a broad age range (12 yrs. +) from an absolute standpoint and relative to a nonpoverty group in the same medical care environment. Action to modify adverse aspects of the health components is to be instituted and evaluation is in terms of change as compared with what occurs in the non-poverty group. An underlying question is whether through the MHT program and activities generated by it, the anticipated gaps between the two groups can be narrowed.

The program expects to begin processing patients in November 1970.

2. Meharry Medical College MHT Project - Evaluation of this project will be performed as part of the study on comprehensive health services by Dr. Sam Wolfe.

3. North Florida RMP, Gainesville, Florida, Dr. Richard Gordon and Co-workers.

In summary MHT is a complex, relatively expensive, experimental system of health services. Evaluation in terms of program effectiveness and efficiency is feasible but the methodology for successful end result or outcome evaluation has yet to be demonstrated for the total system. MHT is adversely affected by two circumstances:

1. It appears too easy and glamorous which is probably the result of over-selling the technological developments, when in fact there are multiple technological problems still to be solved. The major program problems involve the recruitment of the target population and providing adequate follow-up for the individuals tested.

2. The latter relates to the major uncontrollable variable in assessing the value or benefit of MHT and that being the lack of proven therapy for most of the chronic conditions detected.

After struggling with evaluating MHT for several years I usually caution people about trying to implement this system of health services and especially to think through the planned evaluation.

## References

1. Sharp CLE & Keen H. Presymptomatic Detection and Early Diagnosis Williams & Wilkins, Baltimore, MD, 1968
2. Collen et al, NEJM, 283:459-463 1970 and Vol. 280; 1043-1045 1969
3. Neuffield Foundation Report Screening in Medical Care
4. Metropolitan Life Statistical Bulletin, Dec. 1969 Biochemical Profiles
5. Use of "Normal Range" in Multiphasic Testing Files, et al JAMA. Sept. 2, 1968. Vol 205 No 10, p. 684
6. Clinical and Biological Observations on Working Men, Clark, T.W. et al Arch Environ. Health, Vol 19. Nov 1969, p. 700

## APPENDIX 1

NUMBER AND PERCENT PREVALENCE OF CLINICALLY SIGNIFICANT ABNORMALITIES  
ON 13,000 SCREENEES  
THE BROOKDALE HOSPITAL CENTER  
MULTIPHASIC HEALTH SCREENING CENTER  
FEBRUARY 1968 - NOVEMBER 1969

Test	Brookdale Hospital Center		Kaiser Permanente*	
	No.	%	%	Cost
Blood pressure >160/95	4058	31.5	4.1	
Electrocardiogram	3203	25.0	17.3	\$5.90
Chest X-ray	1053	8.8	7.4	6.20
Cervical cytology, III	8	.13		
Visual Distant $\geq$ 20/40	1917	14.9	15.8	1.85
acuity: Near $\geq$ 20/50	817	6.8		
Tonometry:				
$\geq$ 21.9 mm Hg: OD	545	4.7		
OS	659	5.7		
$\geq$ 23.8 mm Hg: OD	194	1.7	0.3	183.00
OS	255	2.2		
Spirometry:				
Pred. FVC <80%	2505	24.1	2.2	14.10
Pred. MVV <80%	2327	22.4		
Audiometry $\geq$ 30 db	3050	28.3	16.2	
Dental: Teeth, poor or bad	1454	11.7		
X-Ray				
Edentulous	1333	18.0		
Alveolar bone loss severe	1439	19.5		
Other X-ray abn.	8785	37.8		
Cytology II-IV	31	.25		

\*NEJM Vol 280 No. 9 p. 459-463

CLINICAL LABORATORY TESTS

Tests	Abnormal limits	Clinically significant abnormalities		
		Brookdale Hosp. Center Total all ages		Kaiser Permanente Total all ages Cost
		No.	%	%
<b>Hemoglobin:</b>				
Females	< 12 gms. %	737	9.5	10.3
Males	< 13 gms. %	249	5.3	3.1
Total		986	7.9	
<b>Hematocrit:</b>				
Females	< 38%	614	7.9	
Males	< 40%	122	2.6	
WBC	< 4 & > 12,000/ cu mm		3.4	2.2
<b>RBC:</b>				
Females	< 4.2 m			
Males	< 4.5 m			
<b>Cholesterol:</b>				
> 95 percentile for age		622	5.0	
Males		235	5.0	
Females		387	4.99	
VDRL	Positive	81	.7	1.5
<b>Urine:</b>				
<b>Culture:</b>				
Females	> 16 col.	679	8.8	3.3
Males	> 16 col.	63	1.5	0.4
Glucose	1+ to 4+	780	6.3	8.2
Protein	1+ to 4+	550	4.4	6.4
Acetone	1+ to 3+	155	1.2	

# EVALUATION OF STROKE – REHABILITATION

## Participants

Jerome Tobis, M.D. - Moderator  
Chairman Department of Physician Medicine  
and Rehabilitation  
University of California at Irvine

Charles M. Wylie, M.D.  
Professor, Public Health Administration  
University of Michigan

Bertram L. Tesman, M.D.  
Assistant Coordinator, Stroke Program  
California Regional Medical Program –  
Area VIII

B. Lionel Truscott, M.D.  
Director, Stroke Program  
North Carolina Regional Medical Program

Philip A. Klieger, M.D.  
Assistant Director for Organizational  
Liaison for Stroke and Rehabilitation  
Regional Medical Programs Service

---

## Evaluating Stroke and Rehabilitation Programs:

### An Overview

CHARLES M. WYLIE, M.D.

At this late stage of the conference, evaluation is no longer an attractive and novel word. The discouraged or bored may suspect the reality of the Turkish proverb: If a stone falls on an egg, alas for the egg; if an egg falls on a stone, alas for the egg. If we fail to evaluate our program, alas for the program; if we do evaluate our program, alas for the program.

To evaluate or not to evaluate - that is not the question for those of us who wish to continue working in RMP's. Society has always advised us to be critical of what we do. The saying, all's well that ends well, reminds us that even centuries ago activities were considered good primarily when their outcomes were good. Thus the salient question is: how can evaluation be a constructive force which improves programs rather than a destructive force for the eradication of programs? It will destroy, for example, if it uses criteria which are so strict that we cannot meet them. It will also destroy if it uses so much of our resources that we have little left to run good programs.

Must evaluation affect us adversely, however? It will if we insist that it be completely free from stress. It will if we expect it to resemble the French view of love, a pleasant diversion between meals, or even more the Swedish view, a pleasant diversion during meals. But evaluation won't harm us if we expect and accept

moderate stress, and use that stress to galvanize us into improvements rather than into fits of depression. This might be regarded as the power of positive thinking about evaluation.

### WHO SHOULD DO THE EVALUATION?

First, a brief word about the site of evaluation. To increase the likelihood of acting on the findings, it seems essential that the effect of RMP's on the national health levels be assessed by those working in the federal office, the effectiveness of regional efforts be evaluated by those in regional offices, while the evaluation of local programs be carried out by local personnel. Too often in their health activities federal personnel evaluate state activities, states evaluate the local picture, locals don't evaluate, and little change occurs. Fortunately, RMP's have learned from mistakes made elsewhere.

The evaluation findings are more likely to be acted on when program personnel evaluate the effectiveness of their own activities. Examining evaluation realistically, however, we must admit that the first priority of the agency staff is to continue the program; program improvement is only a secondary goal, and destroying the program is their great fear. They may often feel that "conventional wisdom" from which the program arose is more important than negative evaluation data. They will correctly add that some decisions must be political and humanitarian, neither of which viewpoints is considered in evaluation.



The likelihood of corrective action may be lower with an outside evaluator, who may have other biases. He may view evaluation as a chance to test theories or methods which interest researchers. He may suspect the evaluation effort, perhaps from bitter past experience, as designed to give the program a legitimacy which it does not deserve. He may suspect further that a critical evaluation will be ignored, or that negative outcomes will be quietly forgotten so as to ensure the growth of future funds. Such events, we may hope, will be rare in RMP's.

In evaluating stroke and rehabilitation programs, our efforts are likely to aim at three levels of information:

1. Changes in resources, including the number or quality of trained personnel.
2. Changes in the activities produced or the work performed by these resources.
3. Changes in the end results of these activities.

Let us consider the strengths and weaknesses of each level of evaluation.

#### RESOURCE CHANGES

RMP funds may improve the quality, quantity, or both, of facilities, personnel, knowledge, or other resources involved in producing stroke and rehabilitation activities. A new hypertension clinic may be supported to prevent stroke, another clinic established for the early detection and treatment of transient ischemic attacks. Health personnel may attend new courses which review, for example, the optimum care of stroke patients. More rehabilitation personnel may be recruited to consult with personnel in home care programs or extended care facilities.

If RMP personnel document that such resources have been changed, but go no further in the evaluation effort (like some annual reports in the past), they imply that these changes will inevitably improve patient care. However, there are too many skeptics among politicians, the general public, and the health professions to expect that such a primitive evaluation, with its possible but still unproved assumption, will go unchallenged. Too many clinics improve the care of small numbers of patients who are already under care, but have no impact on the large burden of neglected disease in the surrounding community. Too many health personnel may fail to act on new information, obtained in courses, or may return to environments in which they cannot apply their new knowledge. Too many rehabilitation personnel must provide minute doses of advice or care to their large caseload of personnel and patients. All of these relate both to the EFFECTIVENESS of what we do (the extent to which we attain our objectives), and to the

ADEQUACY of what we do (how much of the entire problem we are likely to overcome).

Documenting a change in resources is a step which can be swift and cheap; in our concern to "get on with the job," it is only too easy to stop evaluation at this point. To ensure the long term survival of RMP's, however, and to gain information on how our programs may be improved, we must regard this as only the first step in providing more convincing information on the value of stroke and rehabilitation programs.

#### ACTIVITY CHANGES

Many activities are held to be desirable when they seem likely to delay the onset of stroke or improve the function and speed the recovery of stroke patients. An effective change in resources, as described above, will result in more of these desirable activities; we should show that this has truly happened. The process of evaluation becomes more complete and impressive when it shows clearly that the new or improved resources have truly raised the output or quality of activities as well. The steps to collect these data must be planned before the resources are changed. This advance planning makes it possible to contrast the activities before and after the change occurs.

Let us take the situation where an educational program has been shown effective in improving the knowledge of the participants. We wish to show that this change in resource produces a change in subsequent activities. One goal of an educational program may be, for example, to encourage physicians to make better diagnoses on their hospitalized stroke patients. A regional committee of experts or of peers, let us say, has determined the content of the optimal diagnostic examination. The purpose of evaluation will then be to show that physicians taking part in the educational program perform an examination which is closer to the ideal after than before the program. Is such a step feasible? When physicians may frown on taking a test of knowledge and attitudes before and after the educational course, they will not rush to welcome an effort to assess their methods of diagnosis. Compromises may be needed, and we may have to monitor changes in groups of health professionals rather than changes in the activities of individuals.

#### CHANGES IN END RESULTS

Expert committees have been known to err in the past, and a change towards "optimal" care may not inevitably improve the health of the recipients of care. It is

essential, therefore, that some RMP's try to show that health status is raised when a change in resources is followed by more optimal activities. The evaluation of changing health status has been reviewed in an earlier workshop, and it is only too clear that this effort is primitive and difficult. It seems likely, for example, that most outpatient care must be evaluated in terms of activities, since few tangible end results exist.

Case-fatality ratios are high in the acute stage of stroke. An improvement in diagnosis and treatment should be reflected in lower death rates among patients in general hospitals. This will not be the only change in end results, but it is the change which is most readily monitored. Moreover, it is a change which should occur at the same time as the change in activities, and will not be delayed for years after the onset, for example, of effective educational programs. We must have a different time perspective for programs of primary prevention, however, and I shall discuss this in the next section.

In the field of rehabilitation, many measures exist to reflect changes in the physical status of patients. Most indices are based primarily on activities of daily living; they range from those which describe a functional profile of each patient to those which give one overall score to reflect the degree of impairment. Most scoring procedures seem to be repeatable, but little attention has been paid to their validity. The fact that no single method has been used widely may suggest that each has serious inadequacies. Nevertheless, we cannot wait for perfection to occur; it is probably true that any one of a number of indices is better than none at all, and can contribute much to evaluating the end results of rehabilitative care.

#### PRIMARY PREVENTION OF STROKE

Primary prevention of stroke involves those measures taken to prevent the onset of cerebrovascular disease. From the more distant viewpoint, however, cerebrovascular disease is merely a part of the natural course of hypertensive and atherosclerotic cardiovascular disease. Since these conditions begin at a young age, preventive measures before onset are difficult to institute. In practice, therefore, what we label primary prevention is the taking of preventive measures before symptoms begin of cerebrovascular disease.

Probably the technic with the strongest scientific support is the early detection and active treatment of hypertensive disease. How should we proceed to evaluate this effort? We must first form the realistic perspective that primary prevention is a long-term investment. The

cases prevented are mainly those which will develop symptomatic stroke some five, ten, or twenty years later. To expect an immediate and measurable fall in hospitalization rates or mortality for stroke is to expect too much of primary prevention. In its first few years, this program must be evaluated in terms of its intermediate activities and short range goals, the early detection and effective treatment of patients with hypertension. Primary prevention is liable to be wrongly classified as ineffective if we evaluate it by an immediate fall in incidence.

The benefits of primary prevention must be balanced against the costs involved in this process. What must we include among the costs, in addition to the more obvious steps? Certainly we should include the costs involved in diagnosing the false positives, the referrals who are diagnosed as normal by their physicians. Probably we should include the costs involved in diagnosing and treating hypertensives who do not respond to care, or who respond adversely to it. And if we wish to be strict with ourselves, we should also count against the program the cost of diagnosing those who are confirmed to be hypertensive, but who are given no active treatment; reassurance, supervision, and periodic office visits have no magical ability to control the adverse effects of an elevated blood pressure.

#### COMPARISON GROUPS

If evaluation were partly a research activity, producing new knowledge that can be applied to many similar situations, evaluators would have to insist on strict control groups with whom study groups could be compared. Evaluation efforts have the more practical aim, however, of showing whether or not a specific endeavor is reaching the goals which have been set for it. Its generic value has secondary importance; the evaluative study does not have to show that other similar endeavors are likely to be effective. Thus evaluators do not feel compelled to use the rigorous methods and strict controls of those involved in experimental research.

Nevertheless, evaluators must show that activities change and end results improve because of the program being evaluated, and not because of an artifact occurring throughout the region. The evaluation effort must usually involve, therefore, a facility or group of patients which have not received the service being evaluated. Such a comparison group need not resemble the treatment group so closely as it must in an experiment. It must be similar enough, however, to be exposed to the

same extraneous factors which could produce the changes under study. "Before and after" studies become much more successful evaluation efforts when they show that the change occurred only in the group under study and did not occur in a somewhat similar group, perhaps located in a different institution or community.

### CONCLUSION

To seek a graceful end, perhaps I should tell you that around 160 A.D. the Roman emperor Marcus Aurelius gave this advice: "Thou hast embarked, thou has made the voyage, thou are come to shore; get out." At that time, sailors feared to test the effectiveness of their navigational efforts by jumping ashore promptly. They knew only too well the uncertainties and errors involved in sailing in those early years, and feared the unpredictable welcome that might greet them on foreign shores.

In the 1970's, we may still expect some voyagers in the ships of stroke programs and rehabilitation to be slow to leave their vessels for fear that they may have reached wrong and hostile shores; even more reluctance to evaluate the situation may stem from doubts that the vessel has actually left the port of embarkation; and perhaps most reluctance to assess progress will stem from realizing that it takes more than a brisk jump ashore to determine whether we have or have not reached our goals.

#### An Evaluation of a Stroke Program in California

BERTRAM L. TESMAN, M.D.

Area VIII of the California Regional Medical Programs consists of Orange County and, for this specific program, Long Beach. This area incorporates approximately two million people and includes 35 acute hospitals and approximately 75 extended care facilities. To promote effective treatment of patients with stroke, a training program has been set up at Memorial Hospital of Long Beach. Although all disciplines of rehabilitation ideally are involved in stroke, the basic core of the stroke team concept as implemented in Area VIII consists of physician nurse-coordinator and physical therapist. Each hospital in the Area is invited to send these three members of the health team to Memorial Hospital of Long Beach to take special stroke training; back-up teams also can be trained. Hospital administrators also are encouraged to attend the training session. The

physician takes an intensive two-day course; the nurse has three weeks of training; and the physical therapist has two weeks. As of September 1970, seventeen teams, plus selected guests, have been trained in Memorial Hospital of Long Beach.

The medical faculty to train these stroke teams includes specialists in all aspects of the stroke problem. The paramedical faculty includes all standard rehabilitation disciplines, i.e., physical therapist, occupational therapist, nurses, speech therapist and social service workers.

After completion of the training program the core returns to its own institution to utilize the team approach and to train fellow workers in the methodology. As a result of this experience, the team members have improved not only their own expertise but also their awareness of the techniques of the other disciplines in dealing with stroke problems.

The stroke team training divides stroke care and rehabilitation into three phases. The first phase, Phase I, provides the supportive care to the patient until his vital signs have become stabilized. This includes passive range of motion exercises, proper positioning and meticulous skin care. The second phase, Phase II, consists of a multiphase patient evaluation and implementation of an active rehabilitation regimen designed to meet the individual's specific needs. The last level of care, Phase III, essentially is a continuation of the second phase, but emphasis is placed on the post-hospital needs of the stroke victim.

The nurse-coordinator is the catalytic agent among the various modalities in the stroke team. She visits and assesses each new patient in her facility, initiates Phase I at the physician's request, assists in developing the patient care plan with the attending staff and demonstrates proper care techniques when indicated. In addition she is prepared to complete forms which are intended to elicit data for the stroke registry in Area VIII.

The physician is the medical coordinator of the stroke team who is responsible for leading the patient care conferences. He serves as moderator at staff meetings when stroke data at his particular hospital are reviewed and analyzed. He will be available for consultation about the team approach to care of stroke patients for other members of the medical staff at his facility if it is requested.

The physical therapist is responsible for a continuing assessment of all the stroke patients in the hospital and he helps establish their active rehabilitation programs. He also is available to all staff members for consultation.

Presently, one year after the team training was initiated, an assessment of the stroke teams in Area VIII reveals that only one hospital has an active program. We would like to discuss some of the difficulties and obstacles we have identified as a result of the evaluation and we also would like to discuss our resultant plans for increasing the number of effective stroke teams in this Area.

The problems we confronted when attempting to initiate the program were numerous. One year elapsed between the time Area VIII submitted the grant and funds finally were available. This posed a recruitment problem for us. Although I had visited every acute hospital in this Area and discussed the program with administrators, by the time the project was funded many changes had occurred in all levels of personnel. Therefore, most of the commitments for placing staff in the training program were no longer valid when the course actually began; so, again, we have to begin a recruitment program. Moreover, as a result of the change in fiscal policy in Medical and Medicare funding, there was a marked curtailment of available monies to extended care facilities in our Area. This not only makes it economically impossible for them to send staff for an extensive training program, but also limits their ability to provide optimum rehabilitation in their own facility.

Although the team concept in rehabilitation is not totally new to the field of medicine, it is a new approach in many of the hospitals in this Area. Because of the emphasis placed on the active involvement of all team members some of the physicians reacted to the program with diffidence. Also, many of the nurses felt uncomfortable about suggesting the proper level of care to the doctor as the patient's physical needs changed.

Analyzing all of these difficulties, we believe we now have some practical solutions. First, a follow-up faculty is being organized to aid and supervise the already trained stroke teams in their own institutions. This follow-up team will consist of a nurse and appropriate therapists to aid and help organize the individual stroke teams within the hospitals. They will remain in an acute hospital for approximately two to three months until the training of all personnel has been accomplished, team conferences and other aspects of the team approach are underway and the total team feels confident in their activities. They also will discuss the entire program at staff meetings to orient the physicians in the new rehabilitative techniques. In this way we hope to stimulate the physicians as well as the hospital personnel to institute the team approach to stroke care. At the conferences, which will be on a weekly or bi-

monthly basis, the personnel from the surrounding extended care facilities will be invited. It is hoped that personnel in the facilities will become more aware of complete stroke rehabilitation and also that the physicians on the staff of the acute hospital will become cognizant of those extended care facilities which are willing to cooperate in giving better care to their patients on discharge from the acute hospital.

We also hope to develop a mobile van unit which will transport a stroke team to the various extended care facilities in our community in an attempt to introduce the phases of rehabilitation that we have been teaching. We hope that this demonstration pilot project may serve as a model for other communities to augment rehabilitation care where it is not available.

In addition, we have instituted a stroke volunteer training program. Ten volunteers have begun a two-month intensive training program utilizing a carefully selected faculty representing all disciplines of stroke rehabilitation. These volunteers will function in a capacity to aid in the resocialization of the stroke patient and, whenever possible, will assist him in his rehabilitation program under the guidance of the special therapist following the patient's discharge from the hospital.

In 1969 the Collaborative Community Stroke Survey was begun in seven counties throughout the United States in an attempt to gather pertinent epidemiological data concerning stroke throughout our country and compare various separate areas. Orange County became involved with this study and we hope to use this data to help us evaluate our stroke program concepts. The mobile van team also will be recording their efforts with patients and comparing them with a control group to see if a coordinated team can aid and improve rehabilitation care in extended care facilities.

We shall begin a follow-up study on stroke patients this Fall utilizing a form which was developed by a committee of members of all health disciplines involved in the delivery of comprehensive stroke care. It was designed to extract the following kinds of information: the patient's functional condition, types of medical care and rehabilitation being rendered, social and economic conditions, special needs of the patient and his family. Follow-up visits will be made by public health nurses from the Visiting Nurse Association of Orange County from a random sampling of stroke patients six months after their episodes, then again at twelve and eighteen months.

It is our feeling that the level of acute care to the stroke patient has improved in our Area as a result of the stroke team training. However, we have also made many

mistakes in attempting the introduction of the stroke team as we have designed it. An analysis of our work has given us approaches to solving problems relating to the stroke team. The assessment also has helped us seek new and innovative methods of meeting the health and rehabilitation needs of the stroke patient beyond the walls of the acute care facility.

### North Carolina Comprehensive Stroke Program

B. LIONEL TRUSCOTT, M.D.

#### OBJECTIVE

To offer the stroke patient increased opportunities for early diagnosis and treatment, early hospital discharge, and continued follow-up through a community stroke program.

#### *Development of the Program*

*Identification of Subobjectives.* The objective must be reached as a result of accomplishing subobjectives, and these must be (a) realistic within the limitations of personnel and time of the average community hospital and the area it serves, and (b) subject to measurement. The major subobjectives thus identified were:

1. A community health team for comprehensive management of the stroke patient: from diagnosis through follow-up.
2. Professional health personnel knowledgeable in the most advanced methods of diagnosis and treatment of stroke.
3. Increased availability of manpower trained in rehabilitative techniques.
4. Guidelines for high quality, uniform, total management of the patient.
5. Consultative support for communities lacking in specialized personnel.
6. An evaluation mechanism to determine the extent to which the subobjectives and activities had been achieved.
7. Feedback of data to community, for measuring impact of program and identifying needs.
8. Part-time Executive Secretary to administer all activities.

*Activities.* The activities to accomplish each of the above subobjectives were:

1. Development of an organizational framework for a community stroke program, with clearly defined areas of responsibility: Local Stroke Program Committee with Subcommittees (In-Service Edu-

cation, Discharge Planning and Follow-up, Area Resources Development, and Public Education.)

2. Development of a Basic Training Course for Stroke Teams and of an In-Service Education Program for other professional health personnel of the community.
3. Development of an In-Service Training Program for paramedical personnel to make them knowledgeable in rehabilitative techniques.
4. Development of guidelines (organizational, medical, nursing, and rehabilitative)
5. Coordination with State Board of Health Physical Therapy Consultants and with Medical Centers for consultative support to the community.
6. Development of a system to identify the accomplishments, problems, and breakdowns. (hospitalization forms, discharge planning forms, follow-up reports, etc.)
7. Computerization of appropriate data and retrieval for feed-back to community health personnel.
8. Determination of qualifications and procedures for obtaining a local, part-time secretary.

*Program Design* It was not considered feasible to involve each community in the planning process of such a complex program. In consultation with practicing physicians and resource personnel from the three medical centers and the State Board of Health, the Project Staff accomplished the above activities. To ensure that all necessary steps were completed in correct sequence for maximum efficiency, a time-sequential work plan was developed according to the Program Evaluation Review Technique (PERT).

#### *Establishment of a Community Stroke Program*

1. *Community Approval.* (a) The aims and procedures of the Program are explained to a few interested physicians. (b) The interested physician or physicians appoint an ad hoc Steering Committee representing all deliverers of health care; Project Staff describes details and responsibilities in the local program. (c) A permanent Local Stroke Program Committee is formed, and chairmen of Subcommittees appointed. (d) Members of In-Service Education Subcommittee ("Stroke Team") are selected by the Program Committee.
2. *Education and Training.* (a) Stroke Team attends a 4-day Basic Training Course. (b) Project Staff and Consultants conduct two In-Service Education sessions (2 hours each) for community physicians,

and nurses. (c) Project Staff and Consultants conduct 5-6 practical sessions in rehabilitative techniques (positioning, transfer, ambulation) for nurses and physical therapists. (d) Community hospital nursing staff, with aid of training aids loaned by Program and help of Project Staff as needed, conduct 3-4 practical sessions for licensed practical nurses, aides and orderlies. (e) Project Staff helps plan periodic continuing education sessions.

*Implementation of Community Stroke Program*

*Admission of Patient*

1. Nurse notifies Secretary
2. Secretary notifies: Project Staff and State Board of Health Physical Therapy Consultant.

*Evaluation and initial orders*

1. Nurse and physician record admission clinical data, consultation and laboratory requests on form 1b.
2. Physician writes Stroke Admission Orders.

*Treatment of Patient*

1. Guidelines of Management followed.

*Discharge Planning Conference*

1. Secretary notifies Project Staff and Conference members of date.
2. Conference held.
3. Copy of Discharge Plan sent to Project Staff.

*Patient Discharged*

1. Secretary notifies Project Staff of date of discharge and of first follow-up.
2. Forms 1a and 1b completed and sent to Project Staff.

*Follow-Up*

1. Project Staff and physician receive follow-up evaluation reports.

*Some Features of Evaluation*

*Basic Training Course:* Evaluation by participants

*In-Service Education*

1. Evaluation by participants.
2. Pre-and post-session testing.

*Hospitalization Data*

1. Date of admission
2. Clinical and administrative data (Hospitalization Forms 1a and 1b)
3. Date of Discharge Planning
4. Discharge Plan
5. Date of Discharge
6. Date of first follow-up

*Follow-up Date:* Periodic follow-up reports

*Computerization and Retrieval of Data*

*Feed-Back to Community*

1. Periodic visits
2. Annual Workshop

*Summary of Results*

Improvement of, and accessibility to the health delivery system is apparent in the following brief summary:

1. *Community Stroke Programs* presently involve 22 hospitals and 8 nursing homes, with follow-up conducted by 19 county health departments. Over 915,000 people reside in the counties with local stroke programs.
2. *Education, training, and more effective use of manpower* participating in local programs:

M. D. . . . .	125
R. N. . . . .	390
P. H. N. . . . .	103
P. T. . . . .	18
L. P. N.s and Aides . . . . .	314
Others . . . . .	55
Total . . . . .	<u>1,005</u>

3. *Altered and improved patterns of care* are indicated by gradually increasing precision and completeness of clinical and laboratory evaluation, institution of early rehabilitation, more organized discharge planning, and systematic post-hospital follow-up. Some pertinent facts, from the hospitalization forms used in this program, illustrate changes after the start of a local program: (These figures are based on 122 pre-stroke program and 145 post-stroke program patients.)

	<i>Pre-stroke program cohort</i>	<i>Post-stroke program cohort</i>
<i>Patient evaluation</i>		
1. Blood pressure . . . . .	71%	96%
2. Type and speed of onset . . . . .	70%	88%
3. Side, severity of weakness . . . . .	59%	72%
4. Functional ability . . . . .	46%	63%
<i>Use of Multitests</i>		
1. Electrocardiogram . . . . .	27%	51%
2. F.B.S./2 hr. p.p. sugar . . . . .	39%	63%
3. Other (skull x-ray, etc.) . . . . .	18%	27%

<i>Patient evaluation</i>	<i>Pre-stroke program cohort</i>	<i>Post-stroke program cohort</i>
<i>Treatment</i>		
1. Stroke admission orders . . . . .	7%	71%
2. Rehabilitation begun within 48 hrs. after admission . . . . .	0%	22%
<i>Mortality within 48 hours . . . . .</i>	24%	16%
<i>Discharge planning done . . . . .</i>	49%	61%
<i>Scheduled, follow-up care to date . . . . .</i>	None	100 pts.

*Measurement of Health Status* (side and severity of weakness, functional abilities, etc.) at admission, discharge, and at 3-month intervals thereafter is presently available on approximately 200 patients treated according to the Guidelines of Management. These data are now being retrieved for evaluation.

*Reduction of hospitalization costs.* Comparison of pre-stroke program cohorts with post-stroke program cohorts indicate that the latter have a reduced hospital stay of over 4 days (approximately \$200 less per patient).

*Future Plans*

1. Consolidating gains of participating communities.
2. Stroke Prevention and Surveillance.
3. Training additional manpower through new programs.

# EVALUATION OF CANCER REGISTRIES

## Participants

Charles R. Key, M.D. - Moderator  
Assistant Director for Cancer  
New Mexico Regional Medical Program

George Linden  
Chief, California Tumor Registry  
State Department of Health

Charles R. Smart, M.D.  
Director, Intermountain Tumor Registry

Abraham Ringel  
Public Health Analyst, Operations  
Research and Systems Analysis  
Regional Medical Programs Service

---

## Use and Evaluation of Cancer Registries

ABRAHAM RINGEL

Service-oriented cancer registries are organized and operated primarily to assist physicians and patients in the care of the latter. This is accomplished most directly with periodic letters to physicians, and sometimes also to patients (with the physician's consent) to ensure routine surveillance of the disease. Thus, one measure of the effectiveness of a registry is the increasing percentage of successful medical follow-up of patients over time. The advantages of medical follow-up are also reflected in the increased diagnosis of additional primary malignancies and recurrent cancers in the early stages of the disease.

Additional services may take the form of periodic comparative reports to physicians to evaluate the diagnosis and management of cancer in the community and in the separate hospitals. Patient information by age, race, and sex by cancer site and histologic type, by extent of disease (stage), methods of diagnosis, treatment modalities, and survival may lead to improved understanding and management of the disease in the community. For example, the data collected by the registry may be used to determine the trend in the diagnosis and survival of patients with various sites of cancer. This information may also be helpful to hospital administrators in the development of strategies for optimum operation of their institutions, as well as to community planners to determine priorities and the allocation of resources for facilities, equipment, and manpower.

However, it must be emphasized that statistical reports without analysis and interpretation have little

value. Most physicians and other users of registry data do not have the time or background to evaluate statistical data.

A subsidiary value of a cancer registry is its effect in the preparation of complete and accurate medical charts. One way to measure this would be to evaluate the completeness and accuracy of various items in medical charts prior to the initiation of the registry, with medical charts completed after the registry was organized. Comparisons of the information recorded concerning diagnosis, extent of disease, pathology, and therapy for the same sites in the two periods might show significant changes for the better.

Examples of measures to determine the effectiveness of cancer registry programs are:

1. Improvements in the medical follow-up of patients in each of the participating hospitals;
2. Improvements in the proportion of cases microscopically confirmed in the participating hospitals;
3. Improvements reflected in the earlier diagnosis of cases by anatomic site;
4. Changes in the length and/or quality of survival, by age, sex, race, and socioeconomic group for each type of cancer;
5. Improvements in the completeness of reporting by participating hospitals;
6. Improvements in the completeness and accuracy of abstracted cancer cases (quality control);
7. The schedule of participation and compliance with agreed upon procedures and definitions by participating hospitals;
8. The utility and value of the central registry in intramural and community programs of professional and public education.



## Alternative Methodologies for Evaluation of Registries

GEORGE LINDEN

Let me first express my appreciation for being invited to participate in this Regional Medical Program special session on evaluation of cancer registries. Let me also make clear that I am not in any way an expert in program evaluation. I am here today because of my background and experience in the organization, operation and use of a central cancer registry. My first impulse, when Mr. Ringel invited me to participate, was to back off as fast as I could; all I could think of was "I don't know how to evaluate cancer registries." But Mr. Ringel is very persuasive. He accepted my statement and then went on from there to convince me to participate in this session.

Evaluation itself is not new to me. My training as a statistician and my position as Chief of California Tumor Registry for more than fifteen years have forced me to be continuously aware of the problem of evaluating what I was doing or attempting to do. Most of it has been informal—the one formal evaluation having occurred when I first joined the Registry staff. The Registry, which had been operating for seven years, underwent a thorough evaluation of its activities. This resulted in the deletion of many items which were originally thought to be "nice to know about" and some which were important but not obtainable and also involved some basic changes in procedures which made the Registry more efficient and better able to meet its goals.

Our purpose here is to discuss means of evaluating cancer registries which have been developed as part of the Regional Medical Program activities. Any such evaluation must, of course, go back to the purposes and the goals for which the registries were established. These will differ among the various operations and each registry will have to be evaluated in terms of its own precise purposes and goals. There is, however, a common goal that underlies the activities of all cancer registries and all cancer programs and that is the benefit to the cancer patient. The primary question therefore becomes: "What effect does the registry have on the cancer patients?" Since the survival of cancer patients is usually the focus of our measurements, the question can be narrowed to: "What has the registry done to improve the survival of cancer patients?"

It is precisely this question which led to my initial reaction of pulling back and saying I didn't know how to evaluate cancer registries. Can we prove that the activities of the registry actually led to the increased survival

of cancer patients and that this increased survival would not have occurred without the activities of the central cancer registry? I can assure you that this is a very difficult hypothesis to prove directly and conclusively.

This does not mean, however, that evaluation is impossible; cancer registries can be evaluated by dropping down to a lower order of evaluation. There are many areas in our work and personal lives where complete scientific proof of a given hypothesis cannot be obtained but where the preponderance of evidence leads us to what we regard as a reasonable conclusion and action can be taken on the basis of that conclusion. For example—one of the goals of most central cancer registries is to provide data and information that is useful to the medical community in its cancer educational activities. These data provide a resource for the physician in describing and analyzing his experience or his hospital's experience and can also be used as a basis for clinical and other studies. I would not want to take on the task of proving conclusively and scientifically that the use of the data for medical education did in fact assist the cancer patient. Conversely, however, there would be little disagreement with the assumption that the continuing education of physicians who are diagnosing and treating cancer patients will help the patients with cancer. If we can accept this as a reasonable assumption, we can then say that one of the goals of a cancer registry is being met when we provide physicians with these data. The next step is to decide whether the registry is in fact providing such data, and here we are on much firmer ground. We can review the activities of the registry and determine whether the registry has or has not provided such information for the use of the medical community. We can go one step further and try to determine whether in fact these data are being used and how they are being used by the medical, hospital, and public health community.

Another example of the evaluation of a cancer registry on what I call a secondary level has to do with the following: Can we prove conclusively that medical follow-up increases the survival of the cancer patient? Obtaining such proof may be possible (it certainly would be difficult to do) and I've heard the statement challenged. I think I am a reasonable person. I think it reasonable that medical follow-up of cancer patients will result in longer survival than the survival of patients who receive no medical follow-up after their first course of treatment. I am willing to accept this assumption and therefore would accept an increase in the level of medical follow-up of cancer patients as evidence that the registry activity had been beneficial (being also hard headed, I would want to see evidence showing that it

was the activity of the registry which had increased the level of medical follow-up).

We are on much firmer ground when we consider an item of evaluation which the American College of Surgeons is planning to introduce as a requirement in their cancer program—that is, the quality of survival. I think it would be easier to prove and certainly easier to accept the hypothesis that continued medical follow-up is beneficial to the well-being of the cancer patient.

The point of these remarks then is that while we may not be able to prove directly and conclusively that a cancer registry increases the survival rates of cancer patients, we can on reasonable grounds show that the successful operation of a cancer registry does benefit the cancer patient.

The Regional Medical Programs are a comparatively young operation and the cancer registries organized as part of this program even younger. Most of the registries were organized during the last couple of years. It may therefore be a little premature to attempt to evaluate them in terms of the final goals for which they were established. It may instead be necessary to look at the registries in terms of their developmental goals during their early organizational years. It is obviously not possible to judge a registry in terms of publication of five-year end-results if it has only been in existence for two years. The factor of time therefore becomes important. How long has the registry been established? If it has been in existence for only one year or two years, what goals were specified for completion within that one or two year period?

Evaluation will therefore probably be in terms of technical goals. Is the registry system itself organized and operating? Does the registry have properly trained staff? Have the details of the operating system been worked out? Have they been documented? Is the system actually working? Are the various parts of the system, hospitals, physicians, etc., cooperating fully?

Has the registry developed suitable forms for obtaining the original data plus a handbook of instructions for those who are charged with obtaining the data? How good are the data being entered into the system? Does the abstract or other form on which the data are entered accurately reflect the patient, his cancer and his treatment? What educational means are being used to insure that the personnel in the hospitals abstracting the data are trained and knowledgeable? Are workshops being conducted to assist these people? Has a program of quality control of data been instituted? Is there review of incoming records and independent abstracting of a sample of cases to insure a high quality of data?

Is the coding and classification of the various pieces of information entered into the registry system accurate? Is there any check on the quality of coding? Is the data processing system working as it should? Are the data being processed accurately and on time? Are the computer programs for processing and retrieval of data functioning properly? Can data be obtained quickly and at minimum cost?

How current are the data? Are the hospitals reporting cases early enough or are they lagging behind in their abstracting? How good is the follow-up system? Is it working as originally planned or are there difficulties in carrying it out? What proportion of patients are actually followed? What proportion of patients are followed medically?

These are some of the questions which you will be concerned with in evaluating the effectiveness of a registry operation during the first organizational years.

Although the accumulation of information on survival may take a number of years, it is still possible for the registry to feed information back to the hospitals, the medical community, and the individual physician (if this is part of the reporting process) during the early years of the operation. It can fairly early provide basic information on the demographic characteristics of the cancer patients, their cancers (site, histologic type, stage, etc.) and treatment. Information on stage of disease can provide an estimate of the level of early diagnosis of cancer. This can be used to support a program to improve the level of early diagnosis and bring patients to treatment earlier.

I don't want to exaggerate the output that a registry can produce in its early years. A registry's usefulness increases with time and the early years are a time of limited output. What is most important of course is that the community and especially the physicians be informed of the progress of the registry and be the recipients of early output of information. This is an important point for evaluation. Has the registry produced any data? If so, has it been disseminated to the medical, hospital and lay communities? How has it been used?

I would like to take a few moments to stress the documentation of activities carried out by a central cancer registry. At the beginning of my talk I mentioned the informal evaluation which occurs almost continuously. On occasion it becomes very immediate and important. We were asked, several weeks ago, to provide documentation regarding the value of our activity to the Department's program. I was told, at 3:30 P.M. on a Thursday, that the documentation was to be ready before noon on Friday. We have, during the many years

of operation, developed material which can be readily used for documentation. I wrote a very short statement regarding our program and attached to it two of the documents which we had developed. One of these was a Progress Report which is compiled every six months. We originally started this to keep track of and to evaluate our own activities. We have since found it useful in many other situations. A copy of this report is available for observation on the table. The other document I attached was a list of publications which the Registry has produced. I believe this kind of documentation is extremely important in evaluation of a cancer registry.

The Progress Report for January-June 1970 shows the number of cases received during the first half of 1970 and the total as of June 30. It also shows the status of current follow-up efforts, including the number of cases in active follow-up, how many died, the number actually followed and how many were medically examined. The report also contains a detailed description of the requests for data which were completed during this six month period. There were a total of 57 such requests and I believe that the listing of the individual requests constitutes evidence useful in evaluation of the activities of the registry. The annual reports which we provided to hospitals during this period are also described. There is a section on the Hospital Data Books which we developed this year for each of the participating hospitals. The Data Books provide a comprehensive and clearly presented account of the cancer experience of each hospital and are a solid example of the usefulness of a central cancer registry. A copy of the Data Book is also on the table. The Progress Report also includes a description of a number of studies in which we were involved during this period, a listing of two new publications, a description of the future plans of the Registry, and an account of our activities with the Regional Medical Programs in California. It also covers a proposed central cancer registry in Los Angeles County; hospital consultation, training and lectures carried out by the Registry staff; the activities of the Alameda County population based Cancer Registry; work performed under contract with the National Cancer Institute's End Results Group; the Third National Cancer Survey; and a list of visitors to the Registry. The Progress Report has developed from very modest beginnings to a very useful tool for orientation and for documentation of the activities of the California Tumor Registry. Other evidence and documentation of the Registry activity is available for your review on the table; there are also sign-up sheets if you want copies of any of the material.

A few words about what one writer of the material I received in preparation for this session called "dynamic evaluation". I agree wholeheartedly that the evaluation process should not be static. The placing of a value on any part (or all) of the registry system should be followed by the inquiry, "Does the evaluation indicate that changes are necessary to improve the situation?" If so, the evaluation should at least indicate the necessary changes and possibly initiate action to make the changes. Maybe we should propose that a future evaluation be made of the effects of the present evaluation. Was it worth the time and effort? Did it really result in an improved registry program?

What I have said today is certainly not exhaustive in terms of evaluation of cancer registries, but I hope that the combination of your own discussions on program evaluation and our presentations and discussions here will make it possible for you to evaluate the activities of your own cancer registry program.

#### Methodologies for Evaluating Effectiveness and Value of Registries

CHARLES R. SMART, M.D.

Incidence and Epidemiological Registries study the differences in geographic, racial, religious, environmental, social and economic groups seeking etiological factors leading to prevention.

End Results Registries study survival to determine national baselines and to monitor change in survival rates.

The Clinical Cancer Control Registry has in the past been hospital based and patient oriented, attempting to control cancer through encouraging life-time interval follow-through examinations on all patients having had cancer. Through the regular follow-up examination it is hoped that recurrences and second primary cancers will be discovered at a time when they can still be cured. This type of registry also seeks to serve as a self-evaluatory and educational mechanism for both the hospital staff and individual physicians.

While these various registries are emphasizing one phase of the problem or another, their functions greatly overlap and their goals can be summarized under the headings of *service, education* or *research*.

In October 1965, Congress passed Public Law 89-239, known as Title IX - Education, Research, Training, and demonstrations in the fields of Heart Disease, Cancer, Stroke, and Related Diseases, encouraging through

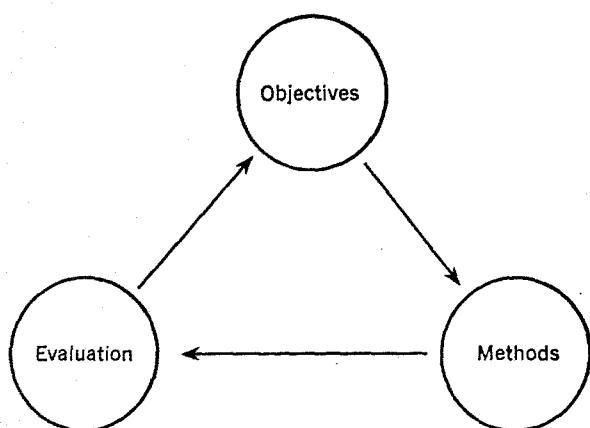
grants the development of cooperative arrangements among medical groups and institutions in making available to their patients the latest advances in the diagnosis and treatment of these diseases. This bill has given rise to many new clinical cancer control types of registries to assist in clarifying the local problem with solid facts, thus allowing logical planning of needed programs and assuring a greater continuity of re-examination of cancer patients. We shall concentrate upon and describe methods of evaluation of this type of registry.

### OBJECTIVES AND EVALUATION

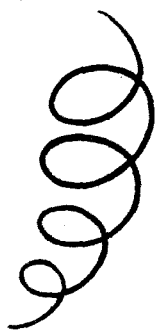
Analysis of goals and objectives must precede evaluation.

1. Decide upon the goals you intend to reach at the end of the program
2. Select the procedure, content, and methods which are relevant to the objectives
3. Carry out the program
4. Measure or evaluate the performance according to the objectives or goals originally selected.

### PROGRAM DEVELOPMENT



### DYNAMICS OF PROGRAM EVALUATION



Accentuate the positive  
Eliminate the negative  
(PROGRESSION)



Eliminate the negative  
but alone leads to stereotyping  
(STATIC)

Cancer Control depends upon both Physicians and Patients. The physician is busy and must find the interaction with the Cancer Control helpful and satisfying — the re-enforcement must be meaningful. Knowledge, skills and attitudes developed on the part of the physician will be transferred to better patient care. Evaluation feedback from the physician and from patient care should be utilized in adjusting the program's methods or objectives.

By modifying the evaluation measurements used in educational programs (attendance, opinion, gain in knowledge, change in behavior), one can develop the following parameters for the evaluation of cancer registries:

1. Participation
2. Opinion or Attitude of the physician
3. Improvement in life-time interval follow-through examinations
4. Improvement in patient management
5. Improvement in patient survival

### EXAMPLE OF PROGRAM EVALUATION

#### Utah Tumor Registry

In October 1966, the tumor registry of the Intermountain Regional Medical Program was formulated on paper. The registry was an integral part of a comprehensive cancer control program involving clinics, seminars, telephone - radio - TV programs, etc. The general concepts are depicted in the diagrams on the following page:

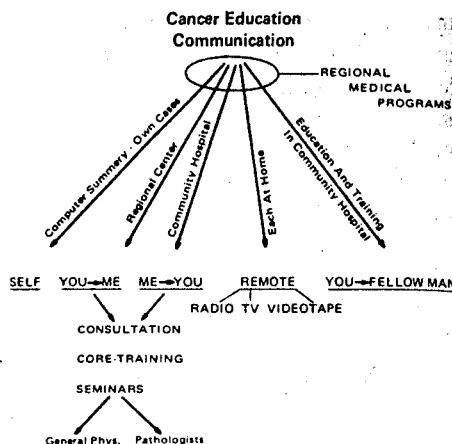
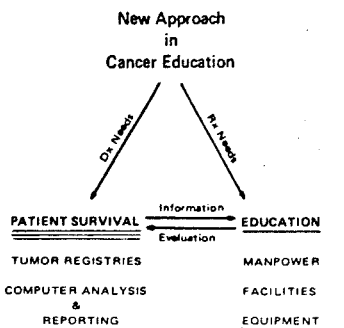
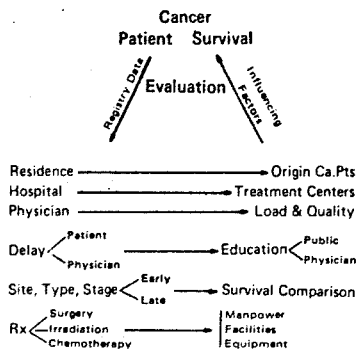
The objectives of the Cancer Control Registry were:

1. To survey and to establish local baselines.
2. To provide local practicing physicians with accurate, meaningful feedback.
3. To save lives through the systematic follow-up of all cancer patients.
4. To identify deficiencies and design operational projects accordingly.
5. To evaluate operational projects.

At first the importance of No. 3 was not completely appreciated. It now heads the list.

#### Methods

1. Gain the support of the medical profession, hospitals, health department, cancer society and other interested health agencies.
2. Enhance presently existing hospital tumor registries by providing:
  - a. Meaningful listings of their patients' data
  - b. Survival reports by site and stage



c. Computer written follow-up letters to their physicians on all living patients not reported within the past year.

d. Automatic updating from death certificates, readmission to other hospitals, etc.

e. Public Health Nurse tracer on all patients who could not be found by the hospital tumor registry secretary.

3. Divide and conquer cancer by providing each physician with his own patients' data as derived from multiple hospital registries, enabling him to evaluate his own cancer practice - patient follow-up, treatment and survival. Listings of current medical references and state and national survival rates are included on the physician's computer report.

4. Merge and analyze the data from the entire state, allowing planning, evaluation, education and lost patient follow-up on a regional basis.

a. Medical society articles are published in the *Rocky Mountain Medical Journal*

b. Cancer society - developed rural cancer survey clinics as a result

c. American College of Surgeons Study Committee on Cancer

d. The State Health Department

e. The Regional Medical Program - for evaluation of operational projects.

### Evaluation

For the purposes of this subject we will deal only with the Utah data, although the entire registry now serves six states and is known as the Rocky Mountain States Cooperative Tumor Registry. Many of the present innovations in this registry were contributions from the other five states. While the Utah Registry did not officially exist until April 1967 when it received funding from the DRMP, patient data from all 44 hospitals were collected back to January 1, 1966, and in those four

hospitals which had registries, data were included going back to January 1, 1957.

### Tumor Registry Report of Accomplishments

April 1, 1969 - Sept 1, 1970.

	IRMP & Utah registry	Six state registry
New cases . . . . .	7,707	19,134
Follow up letters . . . . .		49,287
Cases followed up by public nurse. . .	319	
Dead . . . . .	42	
Referred to physician care . . . . .	176	
Lost to follow up . . . . .	73	
In process . . . . .	28	
Patient listings sent to physicians . . .		2,445
Hospital print outs . . . . .		230
Outputs for special research . . . . .		61
Outputs for articles in <i>Rocky Mountain Medical Journal</i> . . . . .		8
Training sessions for registry workers . . . . .		4
Tumor registry handbooks printed and distributed . . . . .		650
Cumulative cases in registries . . . . .	40,488	51,915

Since in evaluation one is primarily interested in determining whether goals and objectives have been accomplished we will look at the results rather than how they were achieved.

Objective No. 1. To save lives through systematic follow-up of all cancer patients.

1. How many cases have been registered? Eliminating benign tumors there were 28,996 cancers registered as of September 9, 1970 in the Utah Registry. Eliminating further those registered in more than one hospital, etc. (the non-analytic cases), we are left with 23,183 analytic cases which were treated by 774 physicians in 44 hospitals. This is about 2500 new cases per year. Based on the population of 1 million people in Utah, one would

expect 2,850 new cases per year, or 12.3% more than were registered. No attempt has been made to pick up the cases diagnosed and treated in physicians' offices, which make up a high percent of all skin malignancies. This is evidenced from the fact that the three leading dermatologists in the state each have fewer than three cases recorded. By checking the death certificates and cross checking with the radiation therapists and pathologists of the state and considering the number of patients diagnosed and treated out of our state, it would appear that we are missing 15 - 20% of the malignancies in the Utah Registry.

2. What fields of medical practice treat the most cancer patients in the registry?

Field of practice	No. Doctors	No. patients	No. patients/Doctor
Surgery	138	7,748	55
G.P.'s	252	3,417	14
Int. Med.	122	3,059	25
Radiation Rx	5	2,676	535
OB GYN	80	1,897	24
Urology	23	1,778	74
Neurology and Surg	17	563	33
ENT	26	358	14
Pediatrics	43	294	7
Orthopedics	35	196	6
Ophthalmology	26	127	5

### 3. Physicians Non-Participation in Cancer Patient Follow-up Program.

(Non-participation = over 10% or 5 patients not reported 2 yrs.)

#### By County of Residence

	Non / total	Salt Lake Cty	Weber Cty	Utah Cty	Other
G P's	8 / 257 ( 3%)	2 / 77	0 / 26	2 / 37	4 / 117
Surg.	14 / 138 (10%)	11 / 87	1 / 20	0 / 10	2 / 21
OB-GYN	11 / 80 (14%)	10 / 54	1 / 14	0 / 7	0 / 5
Urol.	2 / 23 ( 8%)	0 / 15	2 / 4	0 / 2	0 / 2
Ortho.	2 / 35 ( 5%)	1 / 26	1 / 5	0 / 3	0 / 1
Neuro.	2 / 17 (11%)	2 / 14	0 / 3	0 / 0	0 / 0
Ophthal.	2 / 26 ( 7%)	1 / 13	0 / 5	0 / 5	1 / 3
ENT	1 / 26 ( 3%)	1 / 19	0 / 3	0 / 2	0 / 2
Int. Med.	4 / 122 ( 3%)	2 / 89	0 / 20	2 / 9	0 / 4
Radiation	0 / 5 ( 0%)	0 / 3	0 / 1	0 / 0	0 / 1
Ped's	1 / 43 ( 2%)	0 / 25	0 / 8	0 / 2	1 / 8
	47/772 ( 7%)	30 / 422 ( 7%)	5 / 109 ( 5%)	4 / 77 ( 5%)	8 / 164 ( 5%)

Since the above data depict the Non-Participation Rate the overall participation rate according to the criteria set forth is 94%.

4. A telephone interview was carried out with 102 physicians' office staff to try to determine the manner in which the follow-up letters were being used and if they were helpful. The offices were chosen according to the size of community and the number of letters they were receiving. 23% of the offices felt the letters were helpful as a reminder of those patients not returning for re-examination. 85% of the letters were completed by the physician, 11% by the nurse and 4% unknown. Only 8 offices had a system that would contact the patient if he failed to keep an appointment, and most offices depended upon the patient to return at the time of his appointment but have no fail-safe mechanism built in. A number of physicians known to be strong supporters of

the tumor registry were used as a control; yet two thirds of their secretaries said they knew nothing about it or gave a negative response.

5. In view of the questionable validity of the above study a survey questionnaire was distributed at a scientific meeting of the Utah State Medical Association. Fifty physicians were in attendance (some out-of-state guests). Forty-one questionnaires were completed showing 73% of the physicians felt the registry had been of value or assistance to them. Of the 27% that felt it had been of no value or assistance, three did not treat cancer patients, two had not received any reports, two did not use their reports, two were internists, and one physician

Number of Letters per Office

		1-10	11-49	50+	
Community	3,000	Yes 1 No 8	Yes 2 No 9	Yes 1 No 3	4/24=17%
	3,000 to 10,000	Yes 2 No 6	Yes 3 No 7	Yes 6 No 5	5/24=21%
	10,000	Yes 5 No 16	Yes 6 No 17	Yes 6 No 9	17/54=31%
		8/38=21%	11/40=27%	7/24=29%	24/102=23%

saw no advantage in the program. One wanted simplification and remuneration.

6. Are the individual physician's cancer registry listings of his personal cases of any value in the saving of patient's lives?

a. These listings are simply a compilation of the most significant data which the registry has on all of a doctor's patients. It supplements the tumor registry follow-up letters in reminding the physician of patients who have not returned, but in addition summarizes his experience with specific kinds of cancer. A prominent surgeon in Salt Lake City said, "Before I received my personal computer listing of my cases, I could not remember a single case of cancer of the stomach that was living. I found that I have five cases still living. I operate with a different attitude on patients with cancer of the stomach now!" One of the busiest urologists in our city said, "Before I received my computer listing I thought carcinoma of the prostate was a pretty benign

disease. I was amazed to see the large percentage of my patients who have died rather rapidly of this disease."

b. In assessing our success or failure in any venture it is important to have standards with which to compare our performance. In the physician listings, medical references have been provided pointing out current thought on the latest in the diagnosis and treatment of that particular disease. I have been of the opinion that these articles are little used and are likely a waste of effort. In considering discontinuing them and utilizing the space for a summary analysis of the states experience with that site or for use in communicating educational messages to physicians, a cry went up from the Wyoming Cancer Registry. They found that their reprint service increased 300 fold following the distribution of physician listings. The last time we alternated every other site with summaries of our state's experience with that cancer. A questionnaire in the form of a post card was prepared and sent on September 9th with the physician's reports.

Value Of Reports To Physicians  
(post card questionnaire)

Type of Practice	Questionnaire			Percent Reporting	
	Nos. Sent	Nos. Returned	Percent Returned	Some or Frequent Value	Little
GP's .....	257	59	23%	74%	20%
Surg. ....	138	42	31	90	11
Int. Med. ....	122	38	31	57	39
OB-GYN. ....	80	15	19	80	20
Peds. ....	43	10	23	50	40
Ortho. ....	35	4	11	100	0
ENT. ....	26	10	38	100	0
Ophthal. ....	26	3	11	66	33
Urol. ....	23	4	17	75	25
Neuro. ....	17	0	0	0	0
Radiol. ....	5	0	0	0	0
<b>TOTAL</b>	<b>77</b>	<b>185</b>	<b>24%</b>	<b>75%</b>	<b>22%</b>

The questionnaire indicated that 82 did not use the references, 58 used them once or twice, 39 three or four times and 17 over five times. 106 desired the future use of both references and site summaries, 54 summaries alone and 8 references alone, 16 indicated they wanted neither.

c. The addition of State and National Survival Figures is essential to the physicians' report in order that he might have standards for comparison as well as for developing a better understanding of the natural course of that specific disease site and type. The latter will benefit him in future decisionmaking regarding patient management. One of the problems in our present reports is that we are comparing absolute with relative survival — which will have to be corrected.

7. In attempting to evaluate (measure) whether we are saving lives through systematic follow-up of all cancer patients, we have been skirting the issue and measuring the methods rather than the objective. There are a number of studies showing where second primaries are nearly as curable as the primary and where isolated metastases have been cured in as high as 25% in some selected series and where subsequent therapy to recurrent or metastatic disease has sometimes resulted in cures. Rather than trying to document each of these cases in the registry, it should suffice to demonstrate that because of the registry, patients are now being followed and re-examined who otherwise would not be. The Governor of the American College of Surgeons, Utah Chapter, is a man of long experience in the surgery field. He thought the tumor registry was just a statistics-gathering tool, for which he had neither time nor patience and persistently threw all of his follow-up letters in the waste basket as soon as they arrived. He agreed to try the experiment of pulling all of his records and filling out the reports on those he had seen within the past year and then calling the others in for re-examination. Several months later at a surgical staff meeting he presented his experience with carcinoid tumors of the small bowel. Following this some figures from the hospital tumor registry were referred to and the doctor interrupted the meeting to say, "I thought

the registry was nothing but a bunch of busy work, until I tried the experiment of calling the patients I had not seen for re-examination. Many of those I thought should have been dead were alive and many that I thought should be alive had died, and some patients were able to have further treatment and hopefully would be cured." He said, "The importance of this program is not statistics but GOOD PATIENT CARE!" He encouraged all present to try the same experiment.

If we accept the word of the office secretaries before discussed, then 23% of offices receive *some* help in follow-up from the registry. If we accept the opinion of the physician questionnaire at the State Medical Meeting, then 73% and it helpful. If we look at the percentage of doctors participating in the follow-up program then 94% is the figure. The real question is how much do these letters help or how much could they help in a diligent conscientious doctor's office?

One can see that even in the most diligent doctors' offices cancer patients will be lost unless there is some type of fail-safe technique which will call to the attention of the doctor that the patient has not been in. In the best offices the number of patients who will be thus called in for re-examination will be in the range of 10 to 35%.

Objective #2 of providing practicing physicians with accurate, meaningful feedback has been discussed and evaluated through the post card questionnaire and the response for reprints, etc. In addition to the above, articles are being published every other month in the Rocky Mountain Medical Journal by various medical societies. These articles are attempting to answer the questions: where are we? where should we be? how can we get there? Sixty-one requests have been filled in the past year for special studies for physicians and hospitals.

Objective #3 should now be the development of local baselines. This has been accomplished through our computer summaries and survival curves which are run every 3 to 6 months on the entire region and on each state. In addition a special matrix run will summarize all information which we have on the computer for a specific site or for all states grouped together.

#### *Survey Of Four Physician's Practices*

<i>Phys.</i>	<i>Total Pts</i>	<i>Dead</i>	<i>Alive</i>	<i>Not rep. 2 years</i>	<i>No. letters</i>	<i>Pt's called in</i>
A . . . . .	2,588	1,461	1,077	26 2.4%	385	110 (28%)
B . . . . .	709	270	436	1 0.2%	125	12 (10%)
C . . . . .	244	129	115	1 0.9%	28	6 (21%)
D . . . . .	137	45	92	2 2.2%	43	15 (35%)



Objective #4 of identifying deficiencies and designing indicated operational projects is in constant action. A cancer of the head and neck survey was conducted in Price, Utah, because of the increased number of these carcinomas in this area. Also a special study by the American College of Surgeons study committee in Utah is investigating carcinoma of the stomach in this area due to increased incidence as identified by the tumor registry. The American Cancer Society undertook cancer of the breast and cervix surveys (detection clinics) throughout the rural areas of Utah because these are the

first and second most frequent malignancies in this state. A special investigation of cancers of the lip is presently being done due to the poor survival rates observed through the registry.

Objective #5 of evaluating operational cancer projects is under way to see if specific malignancies are getting the best primary and palliative therapy. The Salt Lake area is being compared to the Ogden area and both of these are compared with the Southern Idaho area where we have put on continuous monthly cancer clinics and seminars.

# EVALUATION OF REGIONAL ADVISORY GROUPS

## Participants

James P. Harkness, Ph.D. - Moderator  
Deputy Coordinator, New Jersey Regional  
Medical Program

David A. Pearson, M.P.H.  
School of Public Health  
Yale University

Homer Hagedorn, Ph.D.  
Arthur D. Little, Inc.

Paul E. White, Ph.D.  
Division of Behavioral Sciences  
Johns Hopkins School of Hygiene and  
Public Health

Stephen L. Maxwell  
Judge of the District Court  
St. Paul, Minnesota and  
Chairman, Northlands Regional  
Advisory Group

Rhoda Abrams  
Assistant Branch Chief, Evaluation  
Branch, Office of Program Planning  
and Evaluation  
Regional Medical Programs Service

---

### Regional Advisory Groups as a Factor in the Regionalization Process

DAVID A. PEARSON, Ph.D.

It is important to examine three key phrases used in the title of this paper. The first is "regional advisory groups." The implication is that RAGs have a major policy-making role which allows them to both affect and effect regionalized activities. Such is the case, as each RAG is charged with responsibility for planning and operating each Regional Medical Program (RMP). Further, each RAG has a specific role in developing program objectives and for reviewing, guiding and evaluating ongoing planning and operating activities. Concerning regionalization, Regional Medical Program Service guidelines require that "the Advisory Group, through the grantee, must submit to the Division of Regional Medical Programs an annual statement giving its independent evaluation of effectiveness of the regional cooperative arrangement (regionalization) established by the Regional Medical Program."<sup>1</sup> Clearly, the RAG has responsibility in the regionalization process and, [as stated in the above quote] it must make an *independent* evaluation of regionalized activities. Logic, rather than official guidelines, dictates that RAG evaluation is *independent* of the RMP Coordinator or his staff.

The second term needing examination is the word "factor." As used in the title, it implies that the RAG is but one of various elements which combine to promote or retard regionalization. Obviously, there are other factors of an extrinsic and intrinsic nature. A decade ago, Masur wrote about the regionalization aspects of the Hill-Burton program stating that "factors of medical economics, civic pride, institutional autonomy, and professional chauvinism have retarded the initiation and development of coordinated hospital systems."<sup>2</sup> The Hill-Burton goal of coordinated hospital systems and the RMP goal of cooperative arrangements contain many of the same concepts and principles of regionalization. If there is objectivity, it will be found that Masur's comments about Hill-Burton's regionalization apply un-  
nervingly to Regional Medical Programs. Therefore, each RAG, in keeping with its charge to produce an independent evaluation must ask such probing questions as:

1. what regionalization activities have we planned?
2. how many have been implemented?
3. how are they progressing?
4. of all our RMP activities how many are truly "regionalized" as opposed to "regional"? To this extent the term "regional" implies simply those RMP activities that happen to take place in a

---

<sup>1</sup>USDHEW, Health Services and Mental Health Administration, *Guidelines, Regional Medical Programs* 1968, p. 9.

<sup>2</sup>Dr. Jack Masur, "Regional Planning Cannot Remain a Paper Pattern," *Hospitals*, 34, January 1, 1960, p. 48.

geographic area or region, and the term "regionalized" implies very definite well-defined, *co-operative* activities.

These questions lead to the third phrase which needs explanation—"the regionalization process." It might have been better to use the phrase "regionalization concept" rather than the "regionalization process," as "concept" means an idea whereas "process" implies an ongoing, continuous activity or development. Clearly in the continental United States we do not as yet have a regionalized approach to the delivery of medical care; nor do we truly have a definite formal regionalized approach on a national basis to any component of health services, be it continuing education or hospital planning. At present, we are more in the world of applying regional concepts or ideas to the field of health.

Many individuals are familiar with earlier efforts at applying the regional concept to the broad field of health and medical care; the Bingham Associates Program, the Rochester Regional Health and Hospital Council, the Albany Regional Hospital Program of Albany Medical College, and the regionalized program of medical care in Puerto Rico stand out as "benchmark" efforts. In the United States, these programs were the initial application of the regional concept. They were, to a considerable degree, based upon an earlier phase of conceptual development, a phase which began in serious fashion in England when the Report of the Consultative Council on Medical and Allied Services (Dawson Report) was published in 1920. This report contained a recommendation for regionalizing the delivery of personal health services. The characteristic regional format of a medical center as a base facility, a community hospital as a district facility and a health center in an outlying area with a two-way flow of service and education between the institutions had its modern day origins in this 1920 report.

Individuals, public and private committees and commissions have been influenced by this report and its format for regionalization. Similarities are found in the 1932 final report of the Committee on the Costs of Medical Care, reports by the Senate Subcommittee on Wartime Health and Education in the mid 1940s, the report of the Commission on Hospital Care in 1946, the Commission on Financing Hospital Care, and in such Federal government efforts as the Ewing Report and the Magnuson Commission. The writings of such individuals as Graham Davis, Joseph Mountin, and John Grant contain a philosophy similar to that of the Dawson Report.

One must not gain the impression that all these reports, individuals and programs defined or im-

plemented regionalization in the same manner; they did not. Perhaps our present inability to state specifically what regionalization means in the health field comes from these varied approaches. At least three models result from a historical analysis of health regionalization: (1) patient care; (2) planning and coordination; and (3) a continuing education model. The patient care model is more or less a composite; however it preceded the development of the other models. The Dawson Report, the report of the Committee on the Costs of Medical Care, and the regionalized health service program in Puerto Rico fall within this category. The patient care model is characterized by such program operations as: (1) direct service patient care; (2) regional planning; (3) coordination of services and facilities; (4) post-graduate or continuing education; and (5) clinical and administrative consultation. The regional scheme has such characteristics as: (1) a coordinated network of comprehensive regional health and medical care services; and (2) cooperative relationships between local, district, state and (frequently) national planning agencies. The "coordinated network" results from a process of *integrating* services through cooperative efforts which are directed at relating spatially separated health care resources and activities to one another within a defined service area.

The planning and coordination model has similar program operations, but excludes direct patient care services. Whereas the regional scheme of the patient care model is formal and somewhat rigid, the planning and coordination model contains voluntary relationships between local, district and state planning councils, plus voluntary relationships between facilities within a given service area. Another distinction between the two regional schemes is that the patient care model has relationships among all health services resources, whereas the planning and coordination model limits its concern primarily to facilities. Examples of this model include various hospital planning councils and the report of the Commission on Hospital Care.

The continuing education model is derived from the numerous programs of postgraduate education developed and administered by medical schools beginning at the conclusion of World War II. Although this model justifiably belongs in a discussion on the evolution of health regionalization, such programs are "regional" only in the sense that relationships exist between a medical school and certain hospitals within a defined area, or certain organizations offer educational activities for the health manpower of a given geographical region. The only program activity is continuing education and,

usually, it is difficult to pin-point characteristics of a regional scheme.

If we in the health field are objective, we will recall that application of regional concepts had been done to other industries, and with greater success, before we "found" this approach. Such being the case, and through the study of various types and forms of application, it is not difficult to develop certain essential elements which characterize a regional scheme that can apply generally to any activity. The regional concept is not static; it is flexible and its application varies depending upon the type of activity. Therefore, in this author's mind, the essential elements of regionalization in the broad sense can be divided into two categories: structural elements and functional elements.

Structural elements are common among all regional schemes, whereas functional elements vary and relate specifically to the activity which is to be regionalized. The three common structural elements are as follows:

1. the region must be demarcated so that the area is economically and spatially defined;
2. there must be a single organizational structure that involves the complete region wherein administration is undertaken by a single agency with authority to undertake its responsibilities through centralization of policy and decentralization of activity or programs operations; and
3. there must be a single financing mechanism which supports completely the entire regional activity.

The functional elements specific to health regionalization are, again in this author's eyes:

1. direct service patient care;
2. maintenance and improvement of professional, technical and administrative practice;
3. rational planning, coordination and integration of resources and services.

Unless a scheme contains these six essential elements, it falls short of the ideal.

Obviously, Regional Advisory Groups have had a major impact in determining both the regional scheme and regional activities of RMPs. What is the result, then, when a comparison is made between RMP regional activities, the three models of health regionalization and the ideal model? Comparing RMP activities with the patient care and ideal models results, generally, in a failing grade for the operating RMPs. One must agree with Bodenheimer's<sup>3</sup> evaluation that RMPs have not

generated "comprehensive regionalization." He states that "comprehensive regionalization would provide a mechanism for allocating resources, including manpower and facilities, among all health institutions in a region and it would link the region's central and peripheral institutions in order to facilitate patient referrals, flow of patient records, consultation by specialists and generalists, and continuing education."<sup>4</sup> Clearly, Bodenheimer's use of the term "comprehensive regionalization" parallels the functional elements of the patient care and ideal models, and clearly the operating RMPs and their RAGs have failed to implement comprehensive regionalization.

Have RMPs had any successes, and do they deserve a passing grade for any implementation of regionalization? Clearly, this time, the answer must be yes. Although the following information is based on 1969 data, the current percentages are about the same. Almost a year ago, the Division of Regional Medical Programs had approved 536 projects: of this total number 55 percent were in continuing education and training, 26 percent were demonstrations of patient care, 11 percent were concerned with planning, coordination and evaluation, and 8 percent were in the area of research and development.

The author does not have first-hand knowledge of all 536 projects; however he has visited various RMPs out of professional interest and as a consultant for the Division of Regional Medical Programs. Further, progress reports and other descriptive materials have been perused on a number of funded RMP activities. From this composite, the author has a comfortable feeling that his knowledge about RMPs and their activities is representative. This being the case, it is a fair judgement to state the RMPs in general have addressed themselves to various aspects of regionalization, or "cooperative arrangements," as stated in the Federal guidelines.

Over half of the funded RMP activities are in the area of continuing education, and many involve cooperative arrangements among institutions, agencies and other resources in the regions. Cooperative arrangements are more characteristic of patient care demonstration activities where, frequently, there exists a coordinated effort of patient referral and a flow of patients and patient services between institutions. Examples of regional cooperation are *not* difficult to find among 8 percent of the total funded activities in the area of research and development. Not all of the planning, coordination and evaluation activities (11 percent of the

<sup>3</sup>Thomas S. Bodenheimer, "Regional Medical Programs: No Road to Regionalization," *Medical Care Review*, 26, December 1969, pp. 1125-1166.

<sup>4</sup>*Ibid.*, p. 1155.

total) are concerned with cooperative arrangements because many such projects support RMP core staffs. On the other hand, various projects can be isolated in this area where there are definite efforts to identify the characteristics (health, economic, social, demographic, etc.) of a region; further, numerous projects address themselves to coordination of facilities, manpower and other resources in the regions.

Federal guidelines explain that the terms "cooperative arrangements" and "regionalization" are synonymous, although they also state that regionalization can connote more than regional cooperative arrangements. In support of a more broad connotation for regionalization beyond the limited idea of cooperative arrangements, the guidelines list several other facets: linking patient care to research and education; sharing of resources; coordination among and between public, private and voluntary health agencies and organizations. This broadening of the term approaches the functional elements of the patient care and ideal models discussed above.

If this distinction is only partly true then there is confusion. A personal opinion holds that the Federal guidelines explain the broad facets of the regionalization concept but emphasize only one segment ("cooperative arrangements") of the regional process. However, "cooperative arrangements" can exist, and many did, before PL 89-239. Continuing education relationships and cooperative arrangements have existed between medical schools and community hospitals for years; similarly between health agencies and organizations and the health manpower within given geographic areas. Formal and informal cooperative arrangements for patient care also existed prior to RMP. As must be obvious by now, regionalization means more than developing cooperation between the resources of a given area. Further, it means much more than undertaking activities for persons, institutions or agencies simply because they happen to be located within a defined geographic area.

The delineation of a region is one of the easier aspects of regionalization in the health field. Relatively tried and true economic, epidemiologic, and demographic techniques can be applied to designate a particular geographic area as a district, region or health service area. Given a qualified and capable staff, the RAG can delegate delineation responsibilities. More serious questions and problems involving policy come about as to what is to take place within a geographically and economically delineated area. It is here that each RAG has not only significant responsibility, but a major effort which transcends mere planning to include an appraisal

of performance or operations; each RAG plans and evaluates.

Cooperative arrangements are the distinguishing characteristic of the Regional Medical Programs' approach to regionalization. To go beyond this approach toward the patient care and ideal models may possibly go against PL 89-239 Section 900c, which requires that RMP activities cannot interfere with the patterns or the methods of financing patient care or professional practice or with the administration of hospitals.

Each RAG, as one of the contributing elements in the regionalization process, must determine whether the patient care and ideal model runs afoul of this requirement. There appear to be no official regulations or guidelines to determine how far an RMP can go without interfering with existing patterns.

Each RAG will undoubtedly face this challenge in the near future, not necessarily from Regional Medical Program Service itself but from outside influences such as organized community and consumer groups, technological, scientific, and organizational accomplishments and progress in the health field, and proposals for national health insurance.

What will become of Regional Medical Programs when one or a composite of existing legislation and proposals for national health insurance is enacted into public law? Have the RAGs made any evaluation of such existing proposals and legislation? Is regionalization and the organization of health services included within these activities?

Currently, there are about 10 to 15 legislative proposals for national health insurance. A number of these contain specific components which are directed at changing the organization, delivery and financing of health services. Health legislation during the 1960s contained wording to prevent change. Regardless of personal feelings about whether the nation's health service system should or should not be changed, it is obvious that the authors of health legislation in this decade see things far differently than those of the past decade. The boldness of the current proposals to change the "system" should provide some evidence of the necessity to alter existing practices. The prediction is made that the patterns of organizing and financing health and medical care services and programs will change within the next five years. Therefore, it is not whether change will come, but rather the extent, scope and type of change.

The composition of each RAG represents significant leadership in each of the regions. Collectively within each region, and collectively from a national standpoint,

the RAGs can have a significant impact on the direction of public policy in relation to health services. The nation is obviously in a transitional period regarding all of health services—its provision, financing, education, research and planning. Now, more than ever, is the time for objective and independent evaluation of the regionalization process in each of the planned and operating Regional Medical Programs.

The underlying reasons for the generic approach to regionalization involve principles of optimal allocation, distribution and use of resources, maximization of output, and the concept of providing goods and services to persons within defined locales. Application of the regional approach to any field or activity is done because it holds promise as a way of achieving balance.

Regionalism results out of a basic need or requirement for a more structured approach. In our society, this need is caused by complexity and the related necessity to obtain benefits which have economic or social value; i.e. the need for greater efficiency which, in turn, is related to prudent allocation of such resources as manpower, equipment, facilities, money and natural resources. The organized approach to obtain greater efficiencies is planning, and it is here that the process of planning and the concept of regionalization come together. This is why Comprehensive Health Planning Programs and Regional Medical Programs should work together.

In general, regionalism is seen as the natural outgrowth of progress, and is associated with better service to the consumer and maximization of output for the producer. The regional concept is viewed as a tool of logic to reduce certain intangible factors to more understandable components. Those of us in the health field have an opportunity to apply this concept. In this respect, regionalism provides the basis for scientific planning and operation in the health field. Herein lies both challenge and opportunity for each regional advisory group.

### **Regional Advisory Group Basis for Evaluation**

PAUL E. WHITE and VAN HOVE

In considering the question of RAG evaluation it is necessary to distinguish between evaluation and research. Although evaluation may involve research, it is essentially a process of comparing scientific aspects of reality with preconceived norms. The norms reflect values and are usually expressed in terms of priorities we

set for our behavior, both individually and collectively. These priorities also take into account the likelihood of their being achieved and represent in effect a selection of ends from a number of alternatives. Research may serve several purposes for us. It may indicate the various alternatives open to us and it may help us to decide on the feasibility of achieving them, but the final selection of ends is governed by our values, i.e., what we feel is desirable. Some organizations are fortunate in having consensus among their members' values. Others are less fortunate and are not able to decide on priorities. Research cannot create values. It can, however, facilitate their application.

Once values are explicit, priorities can be set and evaluation is possible. Evaluation is a process of determining to what extent goals have been realized, and, at a secondary level, why a degree of success or failure has come about.

Two caveats should guide our consideration of evaluation. One is that the common practice of assessing a chaotic situation in search of a measure or measures to justify a program is not, by our definition, evaluation. This procedure is more a process of post hoc rationalization or of documentation to provide legitimation. A second point is that once indices of success have been devised, they sometimes can be the tail that wags the dog, while, ironically, no longer reflecting the achievements originally desired. The measures or indices, in effect, lose their validity. An example of this occurs in the field of rehabilitating the handicapped, where the measure of "number of patients processed" has often led to the rejection of persons requiring longterm treatment and the acceptance of persons with negligible handicaps. In this case, the type of "score" has displaced the direction of the program.

One must therefore periodically reassess the validity of measures one chooses and also must not allow them to become the *criteria* for selecting a course of action. This is the problem of the means becoming the ends and of the "rigged game." In such situations, the criteria may lose any meaningful relationship with desired ends.

For these reasons, evaluation seldom can rest upon single measures. The meaning of each measure must be ascertained periodically, and the validity of a measure or index can be ascertained only by its interrelationships with other indices. For example, the number and kinds of organizations represented on a RAG tells us very little about what we really wish to know when we ask about the representativeness of the RAG. We must, through research, determine the consequences of various forms

of "representativeness" on the activities and achievements of the regions.

The research we have been conducting is not evaluative. We have been studying the behavior of selected RMPs in an endeavor to understand some aspects of why they behave as they do, not whether they behave well. In the course of our research, however, we have employed methodological techniques and gotten some results that have significance for evaluation.

Let us now consider RAG evaluation. Evaluation generally is expensive and should therefore be done only with a clear purpose or purposes in mind. In the case of the RAG, a reasonable purpose would be to reveal its shortcomings in given respects and to correct them. It is assumed that someone is interested in and has the power or sanctions to correct these shortcomings should they exist.

Intensive evaluation of a single local RAG is probably not advisable for several reasons. One is that the workings of the local RAG are already well known to local participants. Another is that the value or import of possible changes in RAG structure or function can be ascertained only by comparing the characteristics and functions of a number of RAGs. Research findings from the study of a number of RAGs can provide us with the means, i.e., methods and criteria for evaluating particular RAGs without undue effort or expense.

We shall focus on three major functions which RAGs can perform and which are likely to be valued. We shall call these representation, legitimation (within the region) and decision-making on two levels: one, decision making with regard to setting explicit policy and, two, decision making with regard to particular tasks. In evaluating these particular functions, baseline data are not necessary, for, being themselves aspects of RAG activities, they obviously did not exist prior to the organization of the RAG. The principle task in evaluating these functions is in determining, (1) whether or not they are, in fact, carried out and to what extent, and (2) the validity of the indices we employ to measure these functions. Essentially, the problem is one of whether the indices measure what we believe they measure.

Representation is an interesting aspect of the RAG. A frequently used measure is membership of various organizations and professions in the RAG. Yet the meaning of representation is more complete than this. We imply by representation, not only membership per se, which may be token, but involvement as expressed in *interest* and *meaningful role*. In our research, therefore, we have compared the characteristics and activities of several RAGs in an effort to understand the relation-

ships among these characteristics. Because our data collection was concluded only recently, the findings we report are tentative and may not be borne out by a more complete analysis of the data. The tentative findings do permit some interesting speculation, however.

One tentative finding is that RAG membership tends to reflect the target populations of the RMPs in terms of grant applications approved and possibly in terms of core staff activities as well. In areas with large metropolitan areas, those RAGs whose membership reflect only state level organizations (with the exceptions of hospitals) tend not to be involved with urban health problems. One tentative explanation for this is that state level organizations overrepresent rural or suburban interests.

Beyond membership per se we have documented attendance at meetings of the representatives of various organizations and professions. An example of attendance in one RAG is given in Table 1. Inspection of a number of cases such as these will permit us to detect trends in particular RMPs in relation to such variables as funding, grant awards, core staff activities and RAG functions. (In our analyses, state and local organizations will be separated.)

A provocative finding which requires further investigation before being interpreted is that as the proportion of physicians on the RAG increases, their attendance decreases. Conversely, as the proportion of lay people increases their attendance increases.

We are at present considering attendance as a measure of member's interest and are investigating factors conducive to greater or lesser attendance. One factor logically related to attendance or interest is the function of attendance for the member. Several of these functions may be enumerated. Narrow organizational interests may be furthered by procuring money for the organization, by monitoring the distribution of funds among competitors, and on another level, by keeping abreast of changes in the inter-organizational field. The setting of policy for achieving collective goals is another function. This policy may range in scope from establishing criteria for grant approval and core staff activities to deciding on particular proposals as they arise without explicit criteria for their determination being stated.

Comparisons of attendance with the dispersal of funds and with the explicitness of goals and the consistency or departures from the goals of fund allocations will allow us to assess various kinds of involvement in the RAGs. Tables II and IV illustrate a content analysis of the meetings of six RAGs. These data represent a preliminary attempt to characterize meeting

topics on a scale from narrow to broad interests. We shall in our analysis try to relate these different concerns to other RMP characteristics.

In relationship to our discussion of the analysis of decision making, which has been the major focus of our research, we should like to consider briefly the question of RMP legitimation. The RAG may play an important part in legitimation. Interestingly, while legitimation, which we shall define as "the perception of the activities of an organization as reasonable and useful," is vital to the survival of the organizations, the bases of legitimacy are generally not well understood. We have used two measures of RMP legitimacy, (1) consensus on RMP functions and (2) submission of proposals for grants. The RAG is important in this respect for we are comparing consensus (a) among RAG and RMP committee members with (b) consensus among organizations involved with RMP, and (c) among organizations that are not involved with RMP. From these comparisons, controlled statistically for other RMP characteristics, we hope to learn whether and under what conditions RAG membership contributes to RMP legitimacy.

Our analysis of decision-making is perhaps the RAGs most important function. It reflects representation in its fullest sense and is possibly conducive to RMP legitimacy in the community (contingent, of course, on the range of organizations on the RAG). We have several measures of decision making. One is a measure of who is perceived as influential in decision making. (See Tables V and VI.) These assessments will be compared with actual decisions made with respect to grant applications and core staff activities.

Another measure is provided by the course of grant applications from submission to (1) rejection or (2) approval by the RMPs. Characteristics of the applicant organization, the proposal, RMP staff, RAG and committee composition in relation to the acceptance or

rejection of applications will permit us to make inferences about the decision making process and about the role of the RAG. Core staff activities will similarly be analyzed in terms of RAG and committee involvement in decisions affecting the staff activities.

What kinds of questions might be answered by such analyses? We can illustrate with our finding on regional responses to the RMPS. The RMPS has two major means of communicating policy to the regions: (1) through directives and (2) by its dispersal of funds. Our analysis of project applications indicates that the regions respond little to directives, while, on the other hand, the national level's awarding of funds in particular areas stimulates the submission of project applications in those areas. The implications of these findings for policy are obvious.

Other questions to which we hope to have answers include, "What are the consequences of RAG membership for the dispersal of funds in terms of recipients and programs? What is the effect of frequency of meetings on RAG attendance, of RAG functions on programs? Is, for example, a RAG that actively screens applications, as evidenced in the selective rejection of project applications, related to a consistent regionalization policy?"

The types of questions we are posing reflect a concern with understanding the unintended or unanticipated consequences of organizational policy, structure, and activity. This paper is intended to illustrate the use of measures and their interrelationships to discover organizational processes and outcomes. We have selected for consideration three aspects of the RAGs: representation, legitimation, and decision-making. Although our research is not evaluation, we trust the findings will have implications for evaluation by indicating the validity of given measures and facilitating corrective action, once norms and desired ideal consequences have been decided upon.



Table 1.—RAG Composition and Attendance by Institution of Affiliation and Profession of Members

	RMP #6							
	Year 67		Year 68		Year 69		Year 70	
	C <sup>1</sup>	A <sup>2</sup>	C	A	C	A	C	A
<i>Affiliation:</i>								
Federal Agency .....	.03	1.00	.03	.00	.03	.75	.03	
State/Local Ag. ....	.06	.38	.06	.62	.06	.75	.08	
Heart/Cancer Vol. ....	.06	1.00	.08	.87	.08	.92	.08	
Other Vol. Org. ....	.06	.25	.09	.25	.08	.50	.05	
Phys. Org. ....	.03	.75	.03	1.00	.03	.75	.08	
Other Prof. Org. ....	.06	1.00	.09	.75	.08	.70	.13	
Hospital .....	.13	.94	.14	.65	.17	.87	.21	
University .....	.06	.75	.06	.75	.08	.66	.08	
Non-Affil. ....	.48	.60	.46	.45	.38	.48	.28	
Total .....	(31)	.69	(35)	.56	(36)	.65	(39)	
<i>Profession</i>								
Physician .....	.38	.71	.34	.76	.38	.75	.39	
Administrator .....	.06	.88	.08	.50	.08	.66	.07	
Nurse .....	.03	1.00	.05	.86	.05	.87	.07	
Health Prof. ....	.03	.75	.03	.50	.03	.50	.05	
Non-Health. ....	.50	.59	.50	.48	.47	.53	.41	
Total .....	(34)	.68	(38)	.59	(40)	.64	(41)	

<sup>1</sup>Composition: Proportion of RAG members in that category

<sup>2</sup>Attendance: Average attendance by the members in each category

Table 2.—Scale of RAG Topics According to Narrower (1) or Broader (7) Conception of the Role of the RAG

Topic	Value
Structure of RMP .....	3
Application and review procedure .....	2
Specific proposals .....	1
Staffing of RMP .....	5
Staff function and duties .....	5
Budgets .....	6
(Other) individual committee function and/or activities .....	7
Goals and priorities of RMP .....	4
Housekeeping .....	1
Relation w/Washington .....	6

TABLE 3.—*Frequency Distributions of Matters Discussed in RAG Meetings*

Matters Scaled According to a Narrower or Broader  
Conception of the Role of the RAG

Meetings 66–68

Meetings 69–70

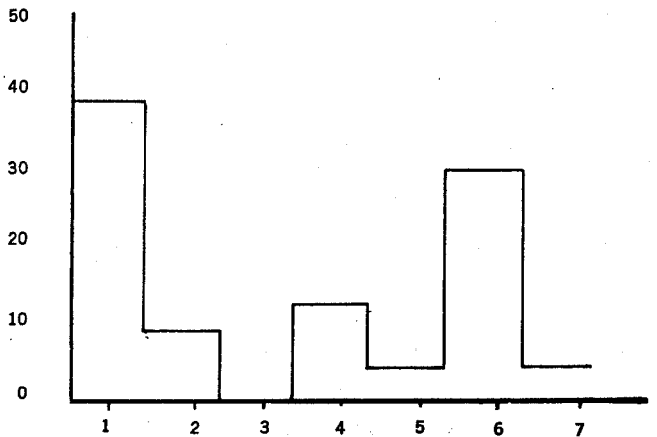
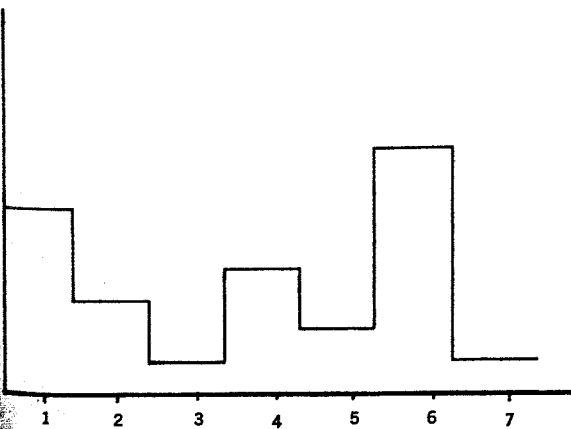
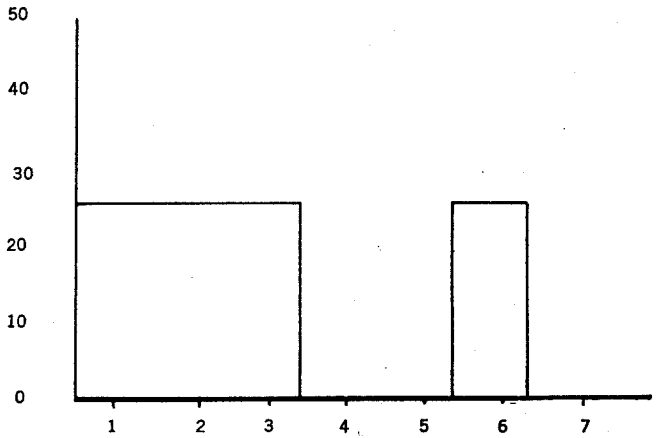
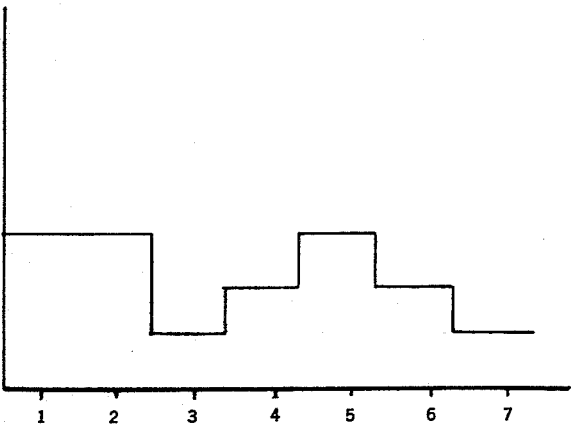
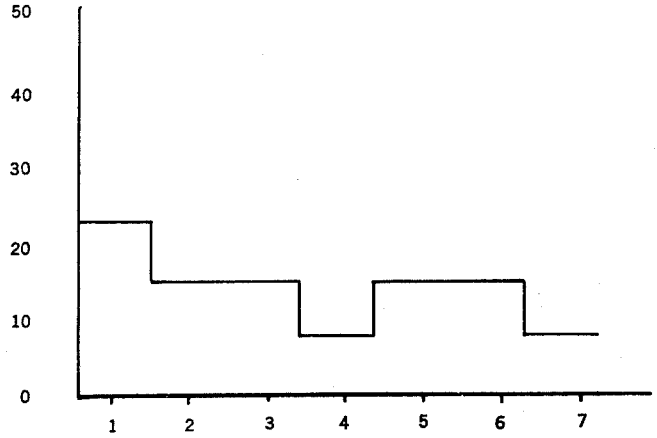
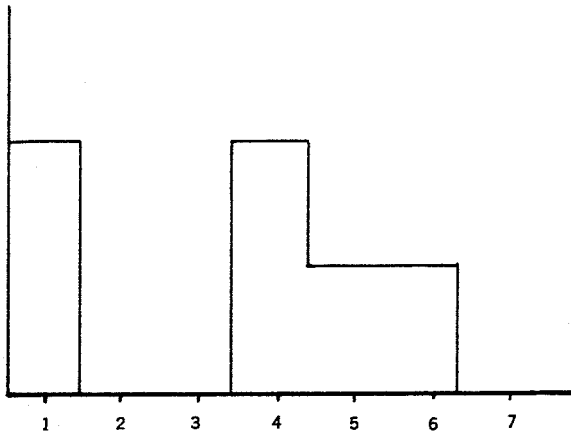
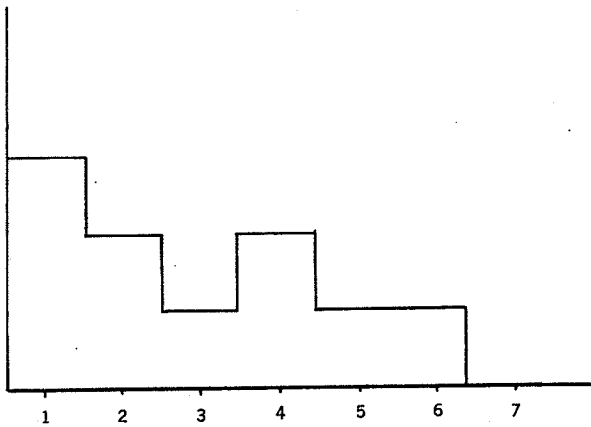


TABLE 3.—*Frequency Distributions of Matters Discussed in RAG Meetings—Cont.*

Matters Scaled According to a Narrower or Broader  
Conception of the Role of the RAG

Meetings 66–68



Meetings 69–70

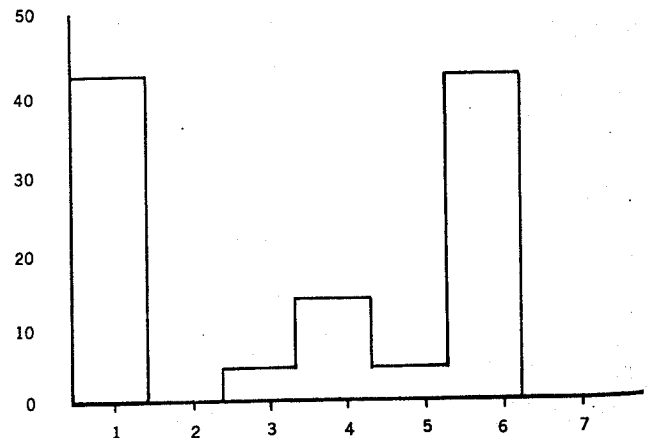
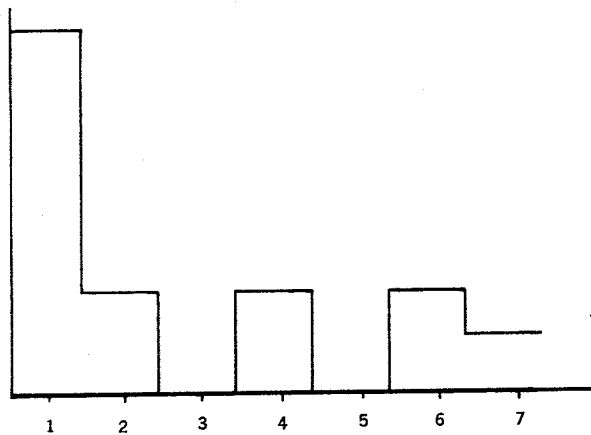
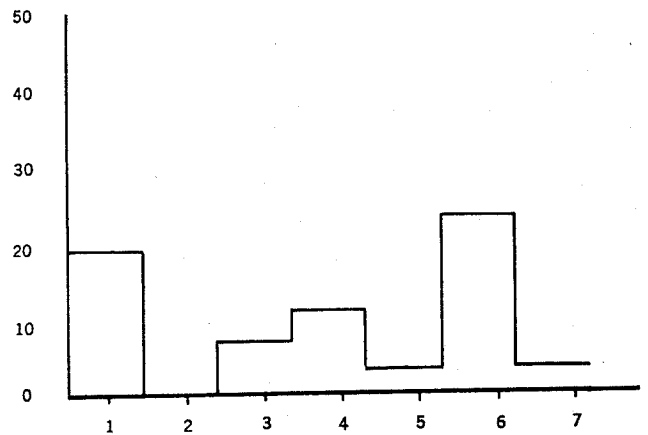
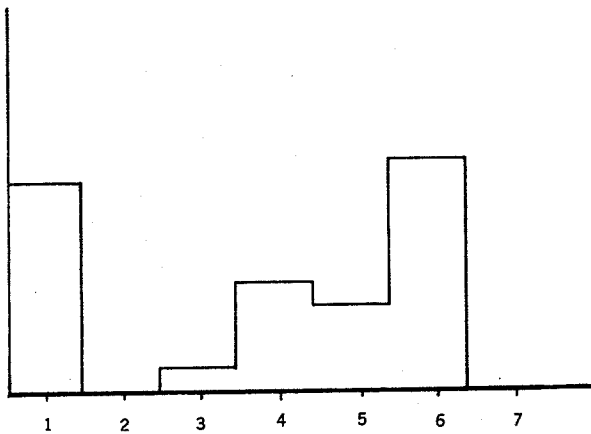
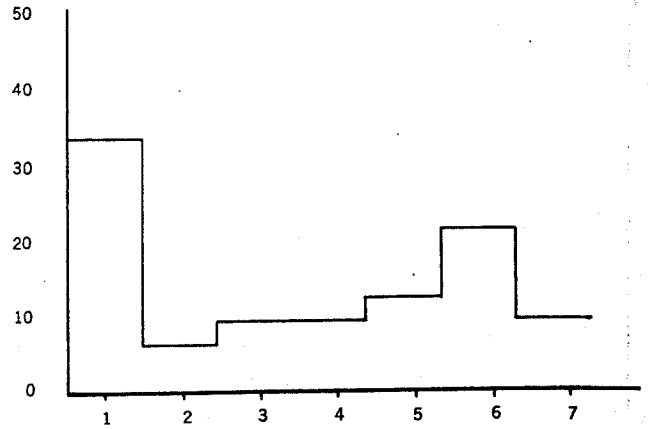


Table 4.—Mean Scores on Topics Scale

	66.68	69.70
1. ....	3.5	3.5
2. ....	2.5	3.8
3. ....	3.6	3.0
4. ....	3.8	3.4
5. ....	2.9	3.6
6. ....	3.9	4.0

Table 4.—Mentioned as Influential Part of RMP Organization by Principal Staff and Committee Chairmen

	RMP No. 1			RMP No. 2			RMP No. 3			RMP No. 4			RMP No. 5			Total		
	S	C	T	S	C	T	S	C	T	S	C	T	S	C	T	S	C	T
Staff ...	.66	.13	.36	.20	.22	.21	.42	.45	.43	.55	.83	.62	.50	.42	.46	.49	.38	.48
RAG ...		.13	.07	.40	.22	.28	.32	.09	.23	.20		.15	.25	.14	.20	.25	.12	.20
Board ..	.33	.25	.28				.05	.04	.07					.14	.07	.05	.10	.07
Steering Com. .		.25	.14				.11	.09	.10				.25	.14	.20	.07	.10	.08
Categori- cal ...				.40	.55	.50	.05	.27	.13					.14	.07	.05	.21	.12
Other ...		.13	.07							.25	.17	.23				.09	.05	.07
Outside .		.13	.07				.05		.03							.00	.05	.02
N . . . .	(6)	(8)	(14)	(5)	(9)	(14)	(19)	(11)	(30)	(20)	(6)	(26)	(8)	(7)	(15)	(57)	(42)	(97)

S= Staff responses  
 C= Committee Chairmen responses  
 T= Total responses

Table 5.—Coordinator and RAG Chairman Mentioned as Influential by Principal Staff and Committee Chairmen

	RMP No. 1			RMP No. 2			RMP No. 3			RMP No. 4			RMP No. 5			Total		
	S	C	T	S	C	T	S	C	T	S	C	T	S	C	T	S	C	T
Co-ord RMP .....	4	4		1	2	3	11	4	15	14	4	18	4	4	8	34	14	48
RAG Chrmn .....	1	1		1	1	2	2		2	1	1	2				5	2	7
# Respondents .....	4	5	9	6	8	14	12	10	22	16	6	22	5	4	9	43	33	76

S= Staff responses  
 C= Committee Chairmen responses  
 T= Total responses

# MEDICAL CARE EVALUATION

## Participants

Glen E. Hastings, M.D. — Moderator  
Coordinator, Nassau-Suffolk Regional  
Medical Program

John W. Williamson, M.D.  
Professor, Department of Medical Care  
and Hospitals  
Johns Hopkins School of Hygiene and  
Public Health

Lawrence M. Witte  
Senior Assistant Health Services Officer  
Office of Program Planning and Evaluation  
Regional Medical Programs Service

### ABCD Strategy on Patient Care Assessment\*

JOHN W. WILLIAMSON, M.D.

One of the most important questions to be asked of any health care system is: WHO needs to learn WHAT to most improve health status of the population receiving care?

Previous study has demonstrated there is an important relationship between patient care assessment and education that might provide a framework for answering this question.<sup>1</sup> Systematic investigation is needed to identify educational objectives that specify the individual who needs to learn as well as the goals to be achieved in the learning process. The doctor, nurse, administrator, patient or general public might each contribute important elements of change to achieve needed improvement. After "instruction," the same methods of inquiry used to identify the problem can be reapplied to evaluate the impact of educational effort exerted to solve the problem. Finally, from this second evaluation, new objectives can be identified for repeating the educational cycle, if warranted, to achieve further improvement.

Systematic application of this approach requires a priority list of health problems to be studied. Methods for developing such a list to encompass conditions involving the most preventable (or remediable) impairment are described in a subsequent publication.<sup>2</sup> Given such a health problem of high priority, a strategy is then needed to identify preventable impairment not being

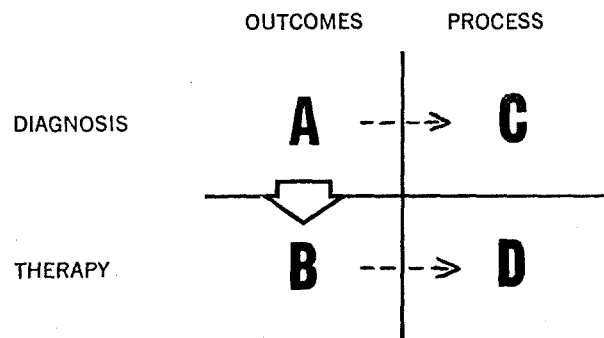
prevented by current medical care. This paper will describe and illustrate such an approach as applied to patient care research; it will be referred to as the "ABCD Strategy."

### DESCRIPTION

Those who have faced the task of evaluating patient care and have contemplated the hundreds of variables that can be measured have probably recognized why a systematic approach is needed. The ABCD Strategy was designed to help identify those variables which might have the highest probability of effecting significant improvement in the health status of a target population.

Figure 1 describes the elements of the strategy. Areas A and B represent the outcomes of care; C and D represent the processes of care that are associated with those outcomes. It is important to note that these areas are lettered in the order of their priority for evaluation and *not* in their chronologic sequence in clinical practice.

FIGURE 1.—ABCD Strategy of Patient Care Assessment



\*Development of this paper was supported by Research Grant PH 43-68-948 from the National Center for Health Services Research and Development, Health Services and Mental Health Administration and the Milbank Memorial Faculty Fellowship Fund.

—Area A represents *Diagnostic Outcomes*, the conceptual base required by the physician to formulate therapy and prognosis. It could be as simple as a single symptom or laboratory result (e.g. “cough” or “hyperuricemia”) or as complex as a disease diagnosis together with major treatment or prognostic considerations (e.g. Lobar Pneumonia due to pneumococcus in a non-compliant patient allergic to penicillin).

—Area B represents *Therapeutic Outcomes*, the effect of treatment on the health status of the patient (e.g. whether the patient lived or died, remained ambulatory or was bedridden, returned to “work,” or remained dependent, etc.).

—Area C represents *Diagnostic Process* or the procedure carried out to formulate the conceptual base symbolized in “A” (e.g. history taking, physical examination, ordering laboratory tests, analyzing data, arriving at a diagnostic synthesis, etc.).

—Area D represents the *Therapeutic Process* involving planning, implementing and evaluating therapy (e.g. prescribing, operating, instructing the patient, follow-up, compliance of the patient, etc.).

The strategy depicted in this figure suggests that the areas most important for assessment are Diagnostic Outcomes (Area A) and Therapeutic Outcomes (Area B). The data obtained from these assessments may indicate whether subsequent effort to study and improve patient care process (Areas C and D) is warranted. Note that the arrows leading to C and D are dotted; this is to indicate that if the outcomes are within previously agreed upon standards, further study of care process can be deferred for that particular problem in favor of an outcome study of the health problem with the next highest priority. Finally, it should be borne in mind that subsequent process study might possibly reveal the outcome criteria or standards were unrealistic rather than indicating remediable deficiencies in patient care.

To implement this strategy, *outcome criteria* need to be developed and compared to *outcome measurements* to determine whether *process study* is required. If process study is indicated, *resulting action* should then lead either to revision of the outcome criteria or eventually to improved care outcomes.

*Outcome Criteria:* Many sources of information can be used in synthesizing such criteria or standards. For instance, to determine the maximum acceptable one-year case fatality rate for patients with a given health problem, data from any of a number of sources might be used: general mortality or actuarial statistics; mortality studies of populations similar to one’s own; previous mortality studies of one’s own population; peer

estimates. Our work and that of investigators like Beverly Payne indicate that the latter source (peer estimates) may offer the most practical basis for setting standards.<sup>3,4</sup> Naturally, all sources of information need to be considered in developing a final synthesis. Also, formulation of such criteria should be prior to and independent of outcome measurements activity.

*Outcome measurement:* To measure diagnostic outcomes represented in Area A, it is important to determine the proportion of the population requiring care for a given health problem who *do not receive it* (false negatives) and similarly, the proportion of those receiving care for the same problem who *do not need it* (false positives). To measure the therapeutic outcomes represented in Area B, follow-up study is important to investigate the patient’s resulting functional condition. If the follow-up interval is sufficiently long as to ensure stabilization of health status, each patient can be classified by level of maximum overall impairment. In this study, the following six levels were used:

1. No impairment
2. Measurable impairment or risk (though asymptomatic)
3. Symptomatic (though working)
4. Not at “work” or “major life activity” (though ambulatory)
5. Bedridden
6. Dead

*Process Study:* Comparing measured findings with established criteria reveals whether detailed study of medical care process is indicated. It is recommended that 95% confidence intervals about measured findings be used, especially if one’s sample is less than one hundred patients. (For example, if a maximum acceptable case fatality rate were set at 5%, a measured rate of 10%, with confidence limits of 4 to 15%, would not be significantly different from criteria and process study would not be indicated.) Finally, the specific objectives and methods of process study will vary widely according to the content and seriousness of the outcome deficiency leading to such inquiry.

*Resulting action:* As a result of process study, the direction and priorities for action to improve outcomes should be revealed. If improvement were not found indicated by such study, then the outcome criteria would seem to require modification. If the criteria proved accurate, then improvement of the health care system would be necessary to correct those factors found causally related to the deficient outcomes. Later, a repeat cycle of criteria, measurement, comparison, and

possibly another process study, would be done to evaluate the effect of the preceding effort to improve.

## ILLUSTRATION

### *Assessing Diagnostic Outcomes (Area A)*

*Example 1:* Area A study of Urinary Tract Infection management in a community hospital in the Midwest was done by prospective examination of over 6,000 consecutive admissions. This investigation was conducted in collaboration with the Department of Medicine of a nearby State University School of Medicine. *Criteria* were independently established by group judgment, taking into account the sensitivity and specificity of methods for detecting urinary tract infections (UTI's) and the implications to the patient of a "missed diagnosis" or a "misdiagnosis." The maximum acceptable percentage of false negatives was set at 15%. The maximum acceptable number of false positives was set at 20%. *Measurement* revealed that 265 of 6,145 consecutive non-new born admissions probably required management of UTI; 187 of these patients did not receive this care from the regular hospital staff, yielding a false negative rate of over 70%. Of 110 patients diagnosed by the hospital staff as having UTI, 32 had urine test results negative for pyuria and/or bacteriuria, yielding a false positive rate of 29%. *Process Study (Area C)* in this example was indicated for two reasons since maximum acceptable criteria for both false negative and false positive diagnoses were exceeded by measured findings. The process study consisted of testing the physicians with a series of brief simulated patient vignettes (describing a patient's history, physical and laboratory findings) requiring a decision to treat or not to treat for UTI with antibiotics. It was found that the doctors tested would usually prescribe antibacterial treatment for all patients who complained of classical UTI symptoms regardless of the results of urinalysis or urine-culture. Chart reviews and follow-up study confirmed this finding in actual practice. Clinical studies have established that many patients with UTI will not present with classical symptoms, and that many patients with such classical symptoms as burning and frequency of urination will not have bacterial infections. It seems possible that a practice of diagnosing only patients with overt symptoms could account for many, if not most, of the false negatives and false positives found. It was clear the improved outcomes would depend upon the physicians learning to utilize urine test results in the diagnosis of these infections. *Resulting action*, in this instance, included several meetings of the medical staff with the

faculty from the State University in continuing education programs related to the diagnosis of UTI. When subsequent study revealed little or no improvement in performance, the physician staff, nurses and administration solved the problem by instituting routine bacteriologic screening (by smears and cultures) of urine from all admissions to this hospital. This procedure includes follow-up verification of bacterial infection of patients with positive screening test results, thereby effectively reducing both false negative and false positive results well below diagnostic outcome criteria.

*Example 2:* Area A study of UTI was also carried out prospectively in the Medical Out-Patient Clinic of the same State University mentioned above.<sup>5</sup> *Criteria* of "maximum acceptable" outcomes established for the Community Hospital study were applied here: false negative, 15%; false positive, 20%. *Measurement* of diagnostic outcome was accomplished by an independent Study Team who examined 133 consecutive new patients admitted to the Medical Clinic. They obtained from each patient a detailed history and urine specimen for urinalysis and culture. After receiving this special workup, patients were admitted to the Medical Clinic for routine management by the regular clinic staff. Over three months later, the patients' charts were examined and recorded results were compared with the findings of the Study Team. Of 18 patients requiring management for UTI, according to the Study Team, 10 were missed by the clinic staff. Although there were no false positives, it is interesting that the upper limits of the 95% confidence intervals about the proportions of 0/8 and 10/18 are the same as the upper limits of the proportions found in the Community Hospital. In other words, with 95% probability, it is possible that in both institutions, there could be as many as 77% false negatives and 36% false positives. *Process Study of Area C* was clearly indicated in this instance. The Study Team identified one or more major UTI screening indications in 108 of 133 consecutive new patients admitted to the Medical Clinic. (Examples of major screening indications are: previous UTI's, recent pregnancy, history of pelvic surgery, catheterization, renal calculi, etc.) The Medical Clinic staff found and recorded such indications in the charts of only 69 patients and followed through with the indicated screening in only 31 of these. To determine whether the problem was a matter of "not knowing" or "not doing," each staff member completed a written examination (similar to speciality board exams) to test his knowledge of urinary tract infection diagnosis and treatment. The average score was 83%; subscores indicated that these physicians had adequate knowledge

of UTI screening indications and diagnostic requirements. The educational problem identified here was of a different nature than that found in the Community Hospital, but had equally serious implications for the patients concerned. It is surmised that continuing education, consisting of formal courses, information, and admonitions to improve, would probably be as ineffective at the University as it was in the Community Hospital described above. *Resulting action* has consisted of discussion of the problem in one staff meeting of the Department of Medicine. Unfortunately, unlike the persistent and ultimately successful effort in the Community Hospital, to our knowledge, no action has yet been taken to solve this problem at the University.

*Example 3:* Area A study of heart failure was conducted at a City Hospital in the East by a group of full time internists who were also faculty members of an internationally recognized private school of medicine. They were interested in applying these same principles to the study of patients in the Emergency Room of their institution. The first cohort they investigated consisted of 113 consecutive patients suspected of having acute coronary artery disease. *Criteria* were based on peer judgment. The staff identified 5% as the maximum acceptable level for both false negative and false positive diagnostic results. *Measurement* of diagnostic outcomes was based on retrospective chart review since sufficient information about critical positive and negative findings was available in the chart of nearly every patient. The findings indicate a false negative rate of 3% and a false positive rate of 0%. *Process study* of Area C was not indicated since criteria were not exceeded. *Resulting action* has not been indicated or undertaken.

*Example 4:* Area A study of diastolic hypertensive patients requiring heart failure management was conducted prospectively by screening approximately 2,000 consecutive medical patients who visited the same City Hospital's Emergency Room. *Criteria* of maximum acceptable rates of diagnostic outcomes were established by the staff at 5% for false negative and 10% for false positive diagnoses. *Measurement* of diagnostic outcome was carried out on each medical patient found to have diastolic pressure of 110 mm HG or greater on the routine blood pressure examination. When such a patient was discharged from Emergency Room care, a member of the Study Team took him to an adjacent room for an independent workup for cardiac failure, which included a history, physical examination, EKG and chest film. From this study, 98 patients were found to require management related to heart failure. Only 8 of these 98 (or 8%) were missed by the regular Emergency Room

staff. *Process study* of Area C was not indicated since the 95% confidence interval about the 8% included the 5% maximum acceptable level. *Resulting action* was taken in this example because the medical staff was sufficiently disturbed by the finding of 8% false negatives. They carried out an educational program on heart failure diagnosis for the Emergency Room staff. This implies that the maximum acceptable criteria established by the staff may have been too high.

Having illustrated the assessment of Area A (Diagnostic Results), using methods which employ the independent diagnostic assessment of a cohort of patients to identify percentage of false positive and false negative diagnoses, attention will now be given to illustrative examples concerning assessment of Area B (Therapeutic Results).

#### *Assessing Therapeutic Outcomes (Area B)*

*Example 1:* Area B study was conducted by a one-year follow-up of 75 patients having management needs related to heart failure among the 113 consecutive Emergency Room patients in the above City Hospital suspected of having an acute coronary occlusion. *Criteria* of the maximum acceptable case fatality rate set by the medical staff by peer judgment techniques were 30%. *Measurement* of therapeutic outcome on one-year follow-up revealed that 23 (31%) had died. *Process study* of Area D was not indicated since the 95% confidence limits about the measured findings did not exceed criteria. *Resulting action* has not been necessary.

*Example 2:* Area B study of the 75 acute coronary suspects having "heart failure" management needs included the follow-up of the 46 patients who were alive and ambulatory one year following their coronary care unit admission. *Criteria* were set by staff judgment. The maximum acceptable proportion of patients ambulatory at the end of one year who were not back to their major social activities, e.g. at "work," was 20%. *Measurement* indicated that 17 (37%) of these 46, had not returned to "work." *Process study* of Area D seemed warranted because of the marked discrepancy between the standard established and the measured findings. Study revealed that 12 of the 17 individuals who had not returned to "work" had not had a coronary occlusion at the time they were in the coronary care unit. These patients, although suspected of having an acute coronary occlusion, proved to have other explanations for their symptoms. Of these 12, ten had been leading active lives prior to admission to the Coronary Care Unit. The fact that more than two out of three of the patients who had not returned to their previous major activity and had no



discernible organic reason for that level of impairment seemed an important finding. *Resulting action* consisted of staff review of this problem and the conclusion that "work" evaluation studies of patients discharged from the Coronary Care Unit are indicated.

*Example 3:* Area B study of the 98 patients with heart failure management needs related to diastolic hypertension was also based on a one-year follow-up. In this example, mortality was the therapeutic outcome studied. Complete information was obtained from 87 of these patients one year following their Emergency Room visit. Nearly half of the group was below 50 years of age and most were active working-class individuals. *Criteria* were established by first referring to standard mortality rate for Maryland, adjusted for age, sex, race and a diagnosis of hypertensive heart disease.<sup>7</sup> These statistics indicated that of the 87 followed, only two should be dead within a year. Individual case-by-case prognostic analysis by the medical staff indicated as many as seven might conceivably be dead within a year. It was decided that nine (10%) would be maximum acceptable case fatality rate. *Measurement* revealed that on one-year follow-up, 18 (21%) were found to have died. *Process study* of Area D was definitely indicated. Preliminary analysis has revealed that of the 18 deaths, 11 may have been preventable; all but three of these 11 involved a cardiovascular death (stroke, coronary occlusion and/or heart failure). Data obtained revealed that only one of the 11 was receiving regular care from a physician and only two were taking antihypertensives and/or digitalis. There would seem to be serious administrative and educational implications for improving the follow-up of such patients. *Resulting action* taken at this institution will be described in the next example.

*Example 4:* Area B study in this instance focused on symptoms found among the 45 patients back to "work" one year following their admission to the Emergency Room. *Criteria* established by the staff would indicate that, if more than 50% of the patients returning to their major life activity had overt cardiovascular symptoms after one year, further inquiry would be required. *Measurement* revealed that 38 (84%) had such symptoms. *Process study* of Area D was indicated since, as with the case fatality rate, the symptom rate far exceeded staff criteria of maximum acceptable results. In this instance, investigation revealed that only 13 of the 38 symptomatic patients who had returned to "work" had seen a physician more than once in the preceding year and were taking needed antihypertensives and/or digitalis. Thirteen others were receiving care from

their physicians but were not taking medication for hypertension or heart failure. The remainder were neither under a physician's care nor taking needed medication. There seemed to be no question that a serious problem had been identified in the care of these patients. As with the preceding example, there appear to be administrative and educational implications for both patients and physicians, if not the entire present system of medical care, which, too often, does not respond to the patient's self-neglect of his own medical problems. *Resulting action*, since the time of this study, has included improved methods for evaluating and following diastolic hypertensive patients seen in the Emergency Room. These patients, as well as others with serious chronic problems, are now referred to a special clinic which has responsibility for the long-term follow-up of these individuals, in other words, if the patient is to be managed by an outside physician, this City Hospital clinic will still maintain responsibility for periodic monitoring of the patient's care and condition. By concentrating responsibility in a defined interest group and stressing follow-up evaluation of care, it is hoped that subsequent study will reveal improved patient outcomes.

#### FINAL CONSIDERATIONS

This approach to patient care assessment raises concern regarding the reliability and validity of the criteria used to determine the need for study of "care process." Since there is little outcome data in the literature, medical staff, using peer judgment methods, must usually develop their own criteria and standards. To test the reliability of these team criteria, the staff members in the heart failure studies were assembled three different times, at three month intervals, to obtain their estimates of "maximum acceptable" outcomes for the same group of patients. Although there was moderate variation of individual estimates, the maximum variation about the group mean was less than 3% comparing the three independent estimates of six different criteria. The staff members inferred that whether or not their estimates were valid, they at least seemed to be consistent. To test validity of these estimates, the team was assembled to provide individual prognosis for each of the 100 patients in the hypertensive cohort. Again, the individual variation was wide, but the group estimates proved specific and meaningful when compared to empirically measured follow-up findings. The comparisons indicated that group prognostic estimates were surprisingly valid.

The overall value of the ABCD Strategy appears to be supported by three factors: 1) It requires that the providers of care focus on prognostic judgments, probably the most critical element in clinical judgment subsuming both diagnosis and therapy. 2) It focuses attention on overall patient impairment and stimulates search for any of the multiple determinants (medical, social, cultural, economic, etc.) of such impairment that may be important. This approach is in contrast to the usual preoccupation with correcting only pathophysiologic causes of impairment. 3) Since this strategy focuses continuing educational resources on solving real problems in medical practice, it would seem to enhance educational effectiveness in two specific ways: a) it identifies learning needs, not only for the physician, but other health care personnel and patients depending upon the problem; and, b) it lends itself to educational assessment in terms of the objectives of the total care process—the improved health of those receiving care.

Finally, if we focus on the ultimate purpose of the evaluation of a health service system, namely, to facilitate improvement, the results of our experience would lead us to infer that the ABCD Strategy is definitely feasible and probably practical for this purpose. It offers an approach that may prove superior to the present haphazard method of planning patient care studies and

continuing education programs. It is hoped that subsequent use of this strategy may facilitate development of practical methods for and renewed interest in answering the critical question: WHO needs to learn WHAT to most improve the health status of the population receiving care?

### Bibliography

1. Williamson JW, Alexander M, Miller GE: Continuing education and patient care research—physician response to screening test results. *JAMA* 201:938-942, 1967
2. Williamson JW, Alexander M, Miller GE: Priorities in patient-care research and continuing medical education. *JAMA* 204:303-308, 1968
3. Storey PB, Williamson JW, Castle CH: Continuing Medical Education—A New Emphasis (monograph). Chicago, Division of Scientific Activities, AMA, 1968
4. Payne BC (ed): Hospital Utilization Review Manual. Ann Arbor, Univ. of Michigan, 1968
5. Gonella JS, Goran MJ, Williamson JW, et al: The evaluation of patient care—a new approach. *JAMA* (in press)
6. Williamson JW, Mitchell JH, Kreider SD: Outcomes of medical care: a study of heart failure management in emergency room patients. (Mimeographed)
7. Vital Statistics of the United States 1967, Vol. II: Mortality, Parts A and B, Washington, DC, US Dept. of HEW, Public Health Service, Health Services and Mental Health Administration, National Center for Health Statistics, 1969

# EVALUATION OF NEW CATEGORIES OF MANPOWER

## Participants

Harriet Kitzman  
Department of Pediatrics  
Strong Memorial Hospital  
University of Rochester

Sarah Mazelas  
Evaluation Director  
California Regional Medical Program--Area I

Elinor Walker  
Public Health Analyst, Health Services Manpower  
Special Research and Development Projects Division  
Nation Center for Health Services Research and Development

Veronica L. Conley, Ph.D.  
Head, Allied Health Section, Continuing Education and Training  
Branch  
Regional Medical Program Service

---

## Evaluation of New Categories of Manpower

### HARRIET KITZMAN

With the stress to prepare new health professionals to assist in meeting the growing health care needs comes the challenge to evaluate the new health professionals' contribution and their impact on the health care system. The program which I will discuss today is the Pediatric Nurse Practitioner Program at the University of Rochester. This program, formally begun in 1968, prepares registered nurses (primarily those who have not been employed during child rearing) to provide direct patient care in ambulatory health care settings. The settings most frequently have been the private pediatricians' offices.

During this session evaluation of the pediatric nurse practitioners' competencies at the end of their educational program will be discussed, as well as the effects of the care on the population being served.

With precise definition of expected behaviors in measurable terms, the task of evaluation is well underway. Expected behaviors of the graduating pediatric nurse practitioners were defined prior to the course. These behaviors were known to students as the course began. Periodic discussion assisted the students in self-evaluation during the course while curriculum flexibility enabled the students to find learning experiences that met their needs. Clinical preceptorship allowed for continuous feedback to students and faculty during the course. Video-taped patient visits allowed students and preceptors to critically review student-patient interaction. This was used both as a learning measure and an

evaluative tool. Objective criteria for rating the interaction has not as yet been developed.

At the completion of the course a written examination was given. This test examined the components of patient care process: problem definition, plan with intervention and evaluation. The test aimed at identifying process components and rating the components accordingly.

Another method of evaluating the competencies of the graduating pediatric nurse practitioners dealt with the graduates' perception of their abilities after reaching the work setting. Approximately one month after the nurses began practicing in their new roles, a research assistant interviewed each nurse to determine the nurse's judgment of her abilities. The questions developed for the interview were based on the expected behaviors as defined in the course objectives.

To evaluate the impact of pediatric nurse practitioners on the health care system two settings were used—a rural setting (a community with a population of 15,000) with no pediatrician and a Rochester suburb.

Two nurses who lived in the rural community were prepared as pediatric nurse practitioners and then established as well-child care providers in the community. Effects of the pediatric nurse practitioners on the level of preventive health care services to children in the community and the physicians' acceptance are presently under study. Baseline community data was obtained by interviewing mothers of all children who were born in the area six and seven months prior to the pediatric nurse practitioners' arrival to determine the level of well-child care the infants had received in their first months of life. The interview included a questionnaire which was pretested in a private pediatrician's

practice where records were available. One year after the nurses began practicing in the community (both in a well-child center and in general practitioners' offices) the population then six and seven months of age will be studied by the same method. By comparing the group studied prior to the availability of the pediatric nurse practitioners with the group studied after the pediatric nurse practitioners were established, the impact of the pediatric nurse practitioners on the total preventive health services to children will be determined. Physicians' acceptance of pediatric nurse practitioners is being studied by use of interview questionnaire both before pediatric nurse practitioners were established and one year after they began practicing in the community.

The study involving utilization of pediatric nurse practitioners in suburban Rochester private pediatricians' offices began in July, 1968. Four pediatric nurse practitioners were prepared and placed in four pediatricians' offices sharing the well-child care with the

pediatricians. The care given by pediatric nurse practitioner-pediatrician teams was studied to determine quality, quantity, cost and acceptance to consumers and professionals. A control group of patients cared for by pediatricians alone was used. Chart review and telephone call sampling showed the number and purpose of patient contacts and visits which provided information on the quality and quantity of care. Total cost of the care given was determined by an accountant. Patient questionnaires provided information about the acceptance of care given to the experimental and control groups. Individual interviews of the pediatricians and pediatric nurse practitioners involved provided information regarding the acceptance of the new role and relationships.

I have briefly described the methods which have been used to evaluate the pediatric nurse practitioner's contribution to the health care of children. Data will soon be available.

# TRAINING FOR EVALUATORS

## Participants

George E. Miller, M.D. – Moderator  
Director for Study of Medical Education  
University of Illinois

Donald Pochly, M.D.  
Chief of Training, Center for Study of Medical Education  
University of Illinois

Marion E. Leach, Ph.D.  
Head, Educational Sciences Section Continuing Education and  
Training Branch  
Regional Medical Programs Service

---

### Summary of Session

MARIAN E. LEACH, PH.D.

Dr. George Miller and Dr. Donald Pochly addressed themselves to issues relating to the functioning of evaluators vis-a-vis Regional Medical Programs. In providing a framework for discussion, Dr. Miller described briefly the Roles of the Evaluator and his need for training:

#### 1. The Evaluator as Trainer:

- a. to train the staff of a Regional Medical Program to understand and to use the evaluator:
- b. to generate the conviction in staff that evaluation is a process, therefore evaluation should be an integral part of planning and implementation; e.g., planning for evaluation begins with the specification of objectives;
- c. to train staff in the use of data (a prerequisite to the effective use of evaluation).

#### 2. The Training of the Evaluator to function within the Regional Medical Program:

- a. he needs to understand the Regional Medical Program objectives;
- b. he needs to understand health professionals;
- c. he needs to learn how to provide leadership in that setting.

In the discussion that followed the participants seemed to indicate that their perception was that they were supposed to recruit and hire "evaluators" without knowing what their functions were. In the absence of articulating their own problems in evaluation they, therefore, could not specify the kinds of competencies or resource needed. The group agreed that it would be more useful under the circumstances to deal with the concept of "evaluation" rather than a person "evaluator."

A number of inquiries were made by participants about training opportunities in evaluation. Some indicated that they had not been aware of the resource represented in the Center for Educational Development headed by Miller.

## ATTENDANCE BY ORGANIZATIONAL AFFILIATION

### REGIONAL MEDICAL PROGRAMS

#### *Alabama*

Joiner, Charles L., Ph.D.  
Director, Bureau of Research and  
Community Service  
Alabama Regional Medical Program

Packard, John M., M.D.  
Director, Alabama Regional Medical  
Program

Patterson, Douglas  
Evaluation Coordinator, Alabama  
Regional Medical Program

#### *Albany*

Brading, Paul L.  
Educational Psychologist  
Albany Medical College

Forer, Raymond, Ph.D.  
Assistant Coordinator of Evaluation  
Albany Regional Medical Program  
Albany Medical College

#### *Arizona*

Ewy, Gordon A., M.D.  
Cardiologist  
Arizona Regional Medical Program

Flynn, John  
Chief of Operations, Arizona  
Regional Medical Program

Humphrey, Alan, Ph.D.  
Assistant Coordinator for Evaluation  
Arizona Regional Medical Program

Ivey, William  
Associate Coordinator, Arizona  
Regional Medical Program

#### *Arkansas*

Rensch, Edward,  
Deputy Coordinator, Arkansas  
Regional Medical Program

Walter, Jacquelyn S., R.N.  
Nursing Coordinator, Arkansas  
Regional Medical Program

Warner, Roger J.  
Director of Planning & Evaluation  
Arkansas Regional Medical Program

#### *Bi-State*

Overman, Ralph, Ph.D.  
Planning Director  
Bi-State Regional Medical Program

Stoneman, William, III, Ph.D.  
Coordinator, Bi-State Regional  
Medical Program

Ulsomer, Dennis  
Project Manager, Bi-State  
Regional Medical Program

#### *California*

Andrew, Barbara, J., Ph.D. Area V  
Assistant Professor Medical Education  
Research  
Division of Reserach in Medical Education  
University of Southern California  
School of Medicine

Barnhill, Bruce M. Area V  
Research Associate  
Division of Research in Medical Education  
University of Southern California  
School of Medicine

Berkowitz, Sheldon Area I  
Associate Specialist Project Evaluator  
University of California

Butts, David, Ph.D. Area V  
University of Southern California  
School of Medicine

Combs, Robert C., M.D.  
Coordinator, Area VIII  
California Regional Medical Program

Elaimy, Wadie, Dr. P.H.  
Assistant Coordinator, Planning &  
Evaluation  
California Regional Medical Program -  
Area IV

Fielder, John, Ph.D.  
California Regional Medical Program

Horovitz, Lee  
Associate Coordinator, California  
Regional Medical Program - Area IV  
U.C.L.A. Rehabilitation Center

LeBrun, Jeanne  
Associate Coordinator, Area III  
California Regional Medical Program

Lloyd, John S., Ph.D.  
Assistant Coordinator for Evaluation  
California Regional Medical Program -  
Area V

Maples, Lila, R.N.  
California Regional Medical Program

Mazelis, Sarah  
Education-Evaluator - Area I  
California Regional Medical Program

Mykytew, Marion, M.D.  
Associate Director-Area VII  
California Regional Medical Program

Roush, Robert E. Ed.D.  
California Committee of Regional  
Medical Programs-Area V

Stroessler, John H.  
Assistant Coordinator, Area VIII  
California Regional Medical Program

Tesman, Bertram L., M.D.  
Assistant Coordinator, Stroke Program  
California Regional Medical Program -  
Area VIII

Thompson, Jack E.  
Coordinator of Evaluation  
California Regional Medical Program

#### *Central New York*

Curry, Walter  
Teaching Coordinator  
Central New York Regional Medical Program

Parresi, Anthony  
Teaching Coordinator  
Central New York Regional Medical Program

Schneider, Robert  
Instructional Communications Coordinator  
Central New York Regional Medical Program

Sovie, Margaret  
Nursing Coordinator  
Central New York Regional Medical Program

*Colorado/Wyoming*

Aldridge, M. Gene  
Continuing Education Division  
Colorado/Wyoming Regional Medical  
Program

Barrett, James, Ph.D.  
Continuing Education Division  
Colorado/Wyoming Regional Medical  
Program

Dyson, James E., Ph.D.  
Associate Director, Colorado/Wyoming  
Regional Medical Program  
Continuing Education Division

Gough, Anne, R.N.  
Chief of Nursing and Allied Health  
Colorado/Wyoming Regional Medical  
Program

Hastings, William O.  
Chief, Project Audit and Control  
Colorado/Wyoming Regional Medical  
Program

Stubblefield, Rex D.  
Executive Assistant  
Colorado/Wyoming Regional Medical  
Program

Syner, James C., M.D.  
Associate Director for Project  
Administration & Health Information  
Division  
Colorado/Wyoming Regional Medical  
Program

*Connecticut*

Clark, Henry T., Jr., M.D.  
Director and Coordinator, Connecticut  
Regional Medical Program

Glasgow, John, Ph.D.  
Associate Coordinator for Evaluation  
Connecticut Regional Medical Program

*D.C. Metropolitan*

*Florida*

Ausman, Robert K., M.D.  
Deputy Director, Florida Regional  
Medical Program

Engebretson, Gordon R., Ph.D.  
Associate Director, Florida  
Regional Medical Program

*Georgia*

Trantow, Donald  
Georgia Regional Medical Program

*Greater Delaware Valley*

Roberts, Dean W., M.D.  
Chief, Regional Medical Program Activities  
Hahnemann Medical College

*Hawaii*

Denney, Ruth  
Chief of Planning and Research Services  
Hawaii Regional Medical Program

Hasegawa, Masato, M.D.  
Program Coordinator and Director  
Hawaii Regional Medical Program

Tunks, Omar A.  
Chief of Operations  
Hawaii Regional Medical Program

*Illinois*

Auerbach, Harry  
Assistant Director for Research  
Illinois Regional Medical Program

Brown, Patricia R., R.N., Ph.D.  
Assistant Director  
Illinois Regional Medical Program

Creditor, Morton C., M.D.  
Executive Director  
Illinois Regional Medical Program

Lovelace, Bryan, Jr.  
Assistant Executive Director  
Illinois Regional Medical Program

*Indiana*

Svan, John, Ph.D.  
Director of Educational Services  
Indiana Regional Medical Program

*Intermountain*

Schorow, Mitchell  
Assistant Coordinator for Planning  
& Evaluation  
Intermountain Regional Medical Program

Smart, Charles M., M.D.  
Director, Tumor Registry  
Intermountain Regional Medical Program

*Iowa*

Caldwell, Charles W.  
Associate Coordinator  
Iowa Regional Medical Program

Latessa, Philip  
Director of Health Statistics  
Iowa Regional Medical Program

Tracy, Roger  
Director of Project Development  
and Evaluation  
Iowa Regional Medical Program

Weinberg, Harry B., M.D.  
Coordinator, Iowa Regional  
Medical Program

*Kansas*

Adair, Charles, Ph.D.  
Assistant Coordinator, Research and  
Evaluation Department  
Kansas Regional Medical Program

Brown, Robert W., M.D.  
Coordinator, Kansas Regional  
Medical Program

Petre, Richard, Ph.D.  
Assistant Evaluation Director  
Kansas Regional Medical Program

Taliaferro, J. Dale, Ph.D.  
Director of Social Systems Research  
Research and Evaluation Department  
Kansas Regional Medical Program

*Louisiana*

Baird, Beverly  
Planning Staff Associate  
Louisiana Regional Medical Program

Walker, Karen, M.P.H.  
Staff Evaluator  
Louisiana Regional Medical Program

*Maine*

Doran, Peter, Ph.D.  
Associate Coordinator  
Maine Regional Medical Program

Jones, Janet H.  
Director, Grant Program Policy  
Maine Regional Medical Program

*Maryland*

Herbert, Henry, M.D.  
Associate Coordinator for Continuing  
Education  
Maryland Regional Medical Program

McMurrin, Vern  
Associate Coordinator for Evaluation  
and Planning  
Maryland Regional Medical Program

*Memphis*

Amis, Lewis N., Ph.D.  
Chief of Planning, Research and  
Evaluation  
Memphis Regional Medical Program

Culbertson, James W., M.D.  
Coordinator, Regional Medical Program

Miller, Edward  
Systems Analyst, Memphis Regional  
Medical Program

*Michigan*

Ballinger, David  
Michigan Heart Association

Brickner, Abraham  
Executive Director  
Michigan Heart Association

Larocque, Gaetane M., Ph.D.  
Associate Program Coordinator  
Michigan Regional Medical Program

Lopushinsky, Theodore, Ph.D.  
Program Developer  
Michigan Regional Medical Program

Lyons, James  
Michigan State University

Welke, Wm. Graham, Ph.D.  
Detroit, Michigan

*Mississippi*

Bridgeforth, Edwin B.  
Program Evaluator, Department of  
Preventive Medicine  
Mississippi Regional Medical Program

Campbell, Guy D., M.D.  
Coordinator, Mississippi Regional  
Medical Program

Lampton, T.D., M.D.  
Coordinator for Heart Disease and Stroke  
Mississippi Regional Medical Program

*Missouri*

Donnell, Denny, M.D.  
Chief, Program Methodology Unit  
Missouri Regional Medical Program

Rikli, Arthur E., M.D.  
Coordinator, Missouri Regional  
Medical Program

Sights, Warren P., M.D.  
Director for Operations  
Missouri Regional Medical Program

*Mountain States*

Smith, C. Eddar, Ph.D.  
Coordinator, Planning and Evaluation  
Mountain States Regional Medical  
Program

*Nassau-Suffolk*

Beckman, Robert  
Director of Research  
Nassau-Suffolk Regional Medical Program

Hastings, Glen E., M.D.  
Program Coordinator  
Nassau-Suffolk Regional Medical Program

Owen, Harrison  
Assistant Coordinator for Project  
Development  
Nassau-Suffolk Regional Medical Program

Rothbell, Gladys  
Nassau-Suffolk Regional Medical Program

*Nebraska/South Dakota*

Marcy, Deane S., M.D.  
Project Administrator, Coronary Care  
Nebraska/South Dakota Regional Medical  
Program

Morgan, Harold S., M.D.  
Program Coordinator, Nebraska/South  
Dakota Regional Medical Program

Morris, George L., Jr.  
Project Administrator, Communications  
Facility  
Nebraska/South Dakota Regional Medical  
Program

*New Jersey*

Harkness, James P., Ph.D.  
Deputy Coordinator  
New Jersey Regional Medical Program

*New Mexico*

Fitz, Reginald H., M.D.  
Director, New Mexico Regional  
Medical Program

Griffith, Dudley  
Health Planning Technologist  
New Mexico Regional Medical Program

Key, Charles R., M.D.  
Assistant Director for Cancer  
New Mexico Regional Medical Program

*New York Metropolitan*

Aronson, Jesse, M.D.  
Deputy Director, New York Metropolitan  
Regional Medical Program

Eller, John  
Evaluation Specialist, New York  
Metropolitan Regional Medical Program

*North Carolina*

Hallman, Shannon P.  
Associate Director, Social Research  
Section  
University of North Carolina

Truscott, B. Lionel  
Director, Stroke Program  
North Carolina Regional Medical Program

Weaver, Ben F.  
Deputy Director, North Carolina  
Regional Medical Program

*North Dakota*

Dietz, Conrad  
Computer Service, North Dakota  
Regional Medical Program

Parker, Lorraine  
Assistant Director, Continuing  
Education for Allied Personnel  
North Dakota Regional Medical Program

Wright, Willard A., M.D.  
Program Director, North Dakota  
Regional Medical Program



*Northeastern Ohio*

Johnson, Bart  
Director of Education  
Northeast Ohio Regional Medical Program

Lifson, Arthur  
Assistant Director, Heart Disease  
Northeast Ohio Regional Medical Program

Meloy, Richard C.  
Director, Administration  
Northeast Ohio Regional Medical Program

*Northern New England*

Danielson, Donald  
Associate Director, Northern New  
England Regional Medical Program  
University of Vermont College of Medicine

Stewart, Caryl  
Assistant Director for Program  
Development  
Northern New England Regional  
Medical Program

Wennberg, John, M.D.  
Director, Northern New England  
Regional Medical Program

*Northlands*

Hill, Russell N., Ph. D.  
Evaluation Officer  
Northlands Regional Medical Program

Miller, Winston R., M.D.  
Program Director and Coordinator  
Northlands Regional Medical Program

*Northwestern Ohio*

Jenkins, Keith  
Northwestern Ohio Regional Medical  
Program

Tittle, C. Robert, Jr., M.D.  
Program Coordinator, Northwestern Ohio  
Regional Medical Program

*Ohio State*

D'Costa, Ayres, Ph.D.  
Director, Fellowship Program  
Ohio State Regional Medical Program

Pace, William G., III, M.D.  
Program Coordinator  
Ohio State Regional Medical Program

Ringe, Robert, Ph.D.  
Director of Communications  
Ohio State Regional Medical Program

*Ohio Valley*

Cook, Anne  
Research Associate  
Ohio Valley Regional Medical Program

McBeath, William H., M.D.  
Director, Ohio Valley Regional  
Medical Program

Moss, Judson, Ph.D.  
Educational Consultant  
Program Development  
Ohio Valley Regional Medical Program

*Oklahoma*

Bexfield, Frank W.  
Associate for Evaluation and Planning  
Oklahoma Regional Medical Program

Cameron, Charles M., M.D.  
Professor and Chairman  
Department of Health Administration

*Oregon*

Kole, Delbert M., M.D.  
Coordinator for Community Organization  
Oregon Regional Medical Program

Yagi, Kan, Ph.D.  
Consultant for Education and Evaluation  
Oregon Regional Medical Program

*Puerto Rico*

Mirando, Luis, M.D.  
Puerto Rico Regional Medical Program

Mullan, Bryan  
Puerto Rico Regional Medical Program

*Rochester*

Haynie, Gloria  
Statistical and Evaluation Unit Staff  
Rochester Regional Medical Program

Jacobs, Arthur, M.D., M.P.H.  
Director, Statistical and Evaluation Unit  
Rochester Regional Medical Program

*South Carolina*

Bowman, C.W.  
Associate Coordinator for Medical  
Districts Program Planning  
South Carolina Regional Medical Program

Coleman, J. Walker, III  
Director, Field Services, Planning  
and Evaluation  
South Carolina Regional Medical Program

*Susquehanna Valley*

Hoffman, John D.  
Chief of Internal Operations  
Susquehanna Valley Regional Medical  
Program

McKenzie, Richard B.  
Director, Susquehanna Valley Regional  
Medical Program

Taylor, David  
Coordinator of Research and Evaluation  
Susquehanna Valley Regional Medical  
Program

*Tennessee/Mid-South*

Metcalfe, Robert M., M.D.  
Associate Director, Tennessee/Mid-South  
Regional Medical Program

Zubkoff, Michael, Ph.D.  
Head, Medical Economics Section  
Tennessee/Mid-South Regional  
Medical Program

*Texas*

Ferguson, David K.  
Operations Officer  
Texas Regional Medical Program

Humble, Robert O.  
Chief of Planning and Evaluation  
Texas Regional Medical Program

McCall, Charles B., M.D.  
Coordinator, Texas Regional  
Medical Program

Reese, Hubert D.  
Data Acquisition Specialist  
Texas Regional Medical Program

*Tri-State*

Colt, Avery  
Tri-State Regional Medical Program

Farrisey, Claire  
Special Projects Coordinator  
Tri-State Regional Medical Program

Keairnes, Harold, M.D.  
Coordinator for Evaluation  
Tri-State Regional Medical Program

Mott, Ruth  
Research Associate for Data Collection  
Tri-State Regional Medical Program

Peterson, Osler L., M.D.  
Associate Director for Data Collection  
Tri-State Regional Medical Program  
Department of Preventive Medicine  
Harvard Medical School

Siebert, Dean J., M.D.  
Associate Coordinator for Dartmouth  
Medical School  
Tri-State Regional Medical Program

*Virginia*

Mason, Jack L., Ph.D.  
Education Sciences Officer  
Virginia Regional Medical Program

Proctor, Margaret W.  
Survey Officer  
Virginia Regional Medical Program

*Washington/Alaska*

Dohner, Charles W., Ph.D.  
Washington/Alaska Regional Medical School  
University of Washington School of  
Medicine

Duren, Gay  
Assistant Director for Evaluation  
Washington/Alaska Regional Medical  
Program

Johnson, William R., Ed.D.  
Washington/Alaska Regional Medical  
Program  
University of Washington School of  
Medicine

Larter, Mariella, M.S.  
Coordinator, Coronary Care Unit Training  
Washington/Alaska Regional Medical  
Program

*Western New York*

Ingall, John R. F., M.D.  
Program Director  
Western New York Regional Medical  
Program

Kaye, Leonard  
Western New York Regional Medical  
Program

Kellberg, Elsa R.  
Associate for Assessment and Research  
Western New York Regional Medical  
Program

*Western Pennsylvania*

Carpenter, Robert R., M.D.  
Director, Western Pennsylvania  
Regional Medical Program

Lapenas, CoraLee  
Research Assistant, Western Pennsylvania  
Regional Medical Program

Reed, David E., M.D.  
Associate Director for Evaluation  
Western Pennsylvania Regional Medical  
Program

Scheuer, Ruth  
Assistant to Director of Evaluation  
Western Pennsylvania Regional Medical  
Program

*West Virginia*

Costello, Joseph  
Biostatistician, West Virginia  
Regional Medical Program

Gallina, Peter P.  
Coordinator of Field Services  
West Virginia Regional Medical Program

Holland, Charles D.  
Director, West Virginia Regional  
Medical Program

*Wisconsin*

Hirschboeck, John S., M.D.  
Coordinator, Wisconsin Regional  
Medical Program

Kraegel, Janet, R.N.  
Nurse Utilization Project Director  
Milwaukee, Wisconsin

Lange, Norma, R.M.  
Nursing Coordinator  
Wisconsin Regional Medical Program

Lemke, Charles W.  
Director of Evaluation  
Wisconsin Regional Medical Program

Sheeley, William J.  
Assistant Coordinator  
Wisconsin Regional Medical Program

**NATIONAL  
RMP REVIEW COMMITTEE**

Besson, Gerald, M.D.  
Review Committee Member  
Los Angeles, California

Lemon, Henry M., M.D.  
Professor of Medicine  
Department of Internal Medicine  
The University of Nebraska Medical Center  
Omaha, Nebraska

Lewis, Edmund J., M.D.  
Chief, Naval Division  
Thorndike Memorial Laboratory  
Boston, Massachusetts

Schmidt, Alexander M., M.D.  
Dean, Abraham Lincoln School of Medicine  
University of Illinois  
Chicago, Illinois

Sister Ann Josephine, C.S.C., Ph.D.  
Administrator  
Holy Cross Hospital  
Salt Lake City, Utah

**NATIONAL  
RMP ADVISORY COUNCIL**

Cannon, Bland, M.D.  
Memphis, Tennessee

Wyckoff, Florence  
Watsonville, California

**REGIONAL  
MEDICAL PROGRAMS SERVICE  
ROCKVILLE, MARYLAND**

Abrams, Rhoda  
Assistant Branch Chief  
Evaluation Branch  
Office of Program Planning and  
Evaluation

Conley, Veronica L., Ph.D. Head, Allied Health Section Continuing Education and Training Branch	Leach, Marian E., Ph.D. Head, Educational Sciences Section Continuing Education and Training Branch	Schoen, Teresa Special Assistant Office of Program Planning and Evaluation
Conrath, Cecilia C. Chief, Continuing Education and Training Branch	Margulies, Harold, M.D. Acting Director	Witte, Lawrence M. Program Planner Office of Program Planning and Evaluation
Dana, Mary L. Secretary Office of Program Planning and Evaluation	Mark, Frank R., M.D. Chief, Operations Research and Systems Analysis Branch	<b>OTHER FEDERAL AGENCIES</b>
de la Puente, Joseph Chief, Program Studies Branch Kidney Disease Control Branch	McDonald, Carol B. Secretary Office of Program Planning and Evaluation	Askew, Cornelius D.P.H. Director, Health Services Evaluation Branch National Center for Health Services Research and Development, HSMHA
Dickenson, Emily Travel Clerk Division of Travel	Mullins, Patty S. Program Analyst Office of Program Planning and Evaluation	Blaxell, Martha Budget Examiner Officer of Management and Budget
Dunning, Herbert P. Program Management Officer Division of Professional and Technical Development	Munier, William B., M.D. Staff Assistant Continuing Education and Training Branch	Brandwein, Raymond A. Program Analyst National Center for Health Statistics, HSMHA
Ensor, Joan E. Program Analyst, Office of Program Planning and Evaluation	Nelson, Elsa J. Health Services Officer Continuing Education and Training Branch	Campana, Edward Program Analyst Comprehensive Health Planning, HSMHA
Frampton, Rita P. Public Information Specialist Office of Communications and Public Information	O'Flaherty, Harold F. Program Analyst Office of Program Planning and Evaluation	Featherstone, Frederick V., M.D. Special Assistant to the Deputy Administrator HSMHA—Office of the Administrator
Friedlander, Edward M. Assistant Director Office of Communications and Public Information	Peterson, Roland L. Assistant Director Office of Program Planning and Evaluation	Fox, Peter, Ph.D. Senior Economist Office of Management and Budget
Green, Dorothy E., Ph.D. Chief, Program Research Section National Clearinghouse for Smoking and Health	Price, Kathy Secretary Office of Administrative Management	Gozkiewicz, Marjorie Project Manager, Health Services Evaluation Branch National Center for Health Services Research and Development, HSMHA
Ichniowski, Francis C. Acting Director, Systems Management Office	Quave, Robert L. Administrative Officer Officer of Management	Logsdon, Donald N., M.D. National Center for Health Services Research and Development, HSMHA
Kaplan, Alan S., M.D. Deputy Chief, Continuing Education and Training Branch	Resnick, Leah Program Analyst Office of Program Planning and Evaluation	Walker, Elinor Public Health Analyst, Health Services Manpower National Center for Health Services Research and Development, HSMHS
Klieger, Philip A., M.D. Assistant Director—Stroke and Rehabilitation Office of Organizational Liaison	Robins, Morton Acting Chief Study Design and Analysis Staff	Yordy, Karl D. Assistant Administrator, Office of Program Planning and Evaluation—HSMHA
Larson, Carol M. Allied Health Specialist Continuing Education and Training Branch	Says, Luther J., Jr. Public Health Advisor Grants Review Branch	

**OTHER**

Bauer, Katherine G.  
Research Associate, Harvard Center  
for Community Health and Medical Care

Berg, Robert, M.D.  
Professor and chairman  
Department of Preventive Medicine  
and Community Health  
School of Medicine  
University of Rochester

Blamphin, John M.  
Assistant Bureau Chief—Washington Office  
*Medical World News*

Crawford, William R., Ed.D.  
Associate, Evaluation Studies Section  
Center for Educational Development  
College of Medicine  
University of Illinois

Deverin, Daniele  
Cybern Education, Inc.

Dunlap, James Ph.D.  
Arthur D. Little, Inc.

Engbretson, William, Ph.D.  
President, Governors State University

Fleisher, Daniel, M.D.  
Director of Health Professions  
Temple University

Forsyth, Jeannette, Ph.D.  
Project Administrator  
Information Support System

Gavett, J. William, Ph.D.  
Associate Professor, College of  
Management  
Department of Preventive Medicine  
and Community Health  
School of Medicine  
University of Rochester

Gustafson, David H., Ph.D.  
Assistant Professor  
Industrial Engineering Division  
University of Wisconsin

Hagedorn, Homer, Ph.D.  
Arthur D. Little, Inc.

Henderson, Maureen, M.D.  
Professor, Preventive Medicine  
Department of Preventive Medicine  
and Rehabilitation  
University of Maryland

Jeffers, James R., Ph.D.  
Director, Medical Economics  
Research Center  
Department of Economics  
University of Iowa

Kelman, Howard, Ph.D.  
Department of Preventive Medicine  
and Public Health  
New York Medical College

Kimmel, Wayne  
Department of Economics  
Public Service Laboratory  
Georgetown University Hospital

Krinneruch, Margery  
Cancer Program  
American College of Surgeons

Kitzman, Harriet  
University of Rochester

Lewis, Charles E., M.D.  
Professor and Head  
Health Administration Division  
School of Public Health  
University of California

Linden, George  
Chief, California Tumor Registry  
Department of Public Health

Maxwell, Stephen L., The Honorable  
Judge of the District Court  
St. Paul, Minnesota

Mayer, Andrew, M.D.  
Cancer Program  
American College of Surgeons

Metzner, Charles A., Ph.D.  
Professor, Department of Medical  
Care Organization  
School of Public Health  
University of Michigan

Miller, George E., M.D.  
Director, Center for Study of Medical  
Education  
Office of Research in Medical Education  
University of Illinois Medical Center

Miller, Irving, M.D.  
Evaluation Technologist  
Albuquerque, New Mexico

Peterson, Sara J.  
Biostatistician  
Fabius, New York

Pochyly, Donald, M.D.  
Chief of Training  
Center for Study of Medical Education  
Office of Research in Medical Education  
University of Illinois Medical Center

Rockwell, M.A., M.D.  
Director, Rand Health Program  
Rand Corporation

Schon, Donald A., Ph.D.  
President, Organization for Social and  
Technological Innovation

Seipp, Conrad, Ph.D.  
Associate Professor, Health  
Services Research Center  
University of North Carolina

Shapiro, Sam  
Director of Research and Statistics  
Health Insurance Plan of Greater New York

Shepherd, Rodger M., M.D.  
Assistant Program Director  
Continuing Education  
Presbyterian Medical Center

Tobis, Jerome, M.D.  
Chairman, Department of Physical  
Medicine and Rehabilitation  
University of California  
Irvine College of Medicine

Webber, Willard W.  
Cancer Program Advisor  
American College of Surgeons

Weckwerth, Vernon, Ph.D.  
St. Paul, Minnesota  
(Member, Northlands Regional Advisory  
Group)

White, Paul, Ph.D.  
Division of Behavioral Sciences  
Johns Hopkins University

Wilbur, Richard S., M.D.  
Assistant Executive Vice President  
American Medical Association

Williamson, John W., M.D.  
Professor, Department of Medical Care  
and Hospital School of Hygiene and  
Public Health  
Johns Hopkins University

Wylie, Charles M., M.D.  
Professor, Public Health Administration  
University of Michigan

Youngerman, Robert A.  
Interregional Representative—  
Georgia Regional Medical Program  
Medical Association of Georgia

