

**Integration of Psychology and Economics:
Initiative on Measuring Well-being and Burden of Illness**

Prepared by Daniel Berch, Ph.D.

with assistance from Rose Li, Jared Jobe, and Merry Ward

Three related workshops have been sponsored (in whole or in part) by NIA/BSR since June, 1999: the NIH Meeting on Burden of Illness (June, 28, 1999); Psychology and Economics (August 25, 1999); and Old and New Measures of Human Well-Being, and Their Possible Significance for Policy (January 27-28, 2000). All three meetings (organized by Rose Li with assistance from Jared Jobe) were concerned in part with issues that are of interest to the fields of psychology, economics and medicine, and the various hybrid disciplines that have emerged at their interface, such as behavioral medicine and behavioral economics. The latter field is particularly germane to the issues treated at the Psychology and Economics and the Well-Being workshops. This field emphasizes the validation of modeling assumptions, integration of micro-level data on decisions (including experimental evidence), and adoption of lessons learned from psychology.

It is also apparent that the construct of burden of illness (BOI) is fundamental to the assortment of measures, methods, and models discussed at the three workshops. More specifically, the components of BOI most relevant to the fields of psychology and economics are the behavioral dimensions comprising the quality of life: psychological and social functioning; impairments or pain; and perceptions of well-being. Simultaneous consideration of the economic, social, psychological and environmental influences on aging-related decision-making is crucial to arriving at a comprehensive understanding of these processes. Accordingly, an individual's valuations of subjective well-being not only constitute a primary dimension of the burden of illness, but also contribute substantially to aging-related decisions regarding living conditions, work and retirement, and health-related issues. Consequently, there is a vital need to improve our understanding of how subjective well-being is measured or mis-measured in the major summary surveys, as well as to find a comparatively economical and unbiased survey measure of well-being that could be validated against the cumbersome and expensive physiological techniques and molecular self-report methods.

Finally, it should be noted that the development of initiatives concerned with psychology and economics in general and burden of illness and well-being in particular would admirably complement a recent BSR Program Announcement: Higher-Order Cognitive Functioning and Aging. This announcement solicits research for determining how older adults process complex information, the cognitive processes that subserve how judgments and decisions are made, and the types of information that are most salient to older adults when they make decisions about retirement, finances, health and life insurance, and medical treatment.

A summary of each meeting along with recommendations follows.

NIH Meeting on Burden of Illness (BOI)

The Burden of Illness (BOI) workshop was organized in response to recommendations in a report from the Institute of Medicine (IOM) that NIH reexamine how it uses BOI metrics when

setting research priorities. Economists and medical scientists were brought together to consider how several such metrics, as well as available epidemiologic data, can help in setting research priorities to meet public health needs. The participants explored the possible applications and limitations of measures of BOI information. Topics discussed included: the definition and measurement of BOI; whether the NIH investment in improving these measurements could enhance research priority setting, research productivity, and efforts to reduce disease burden; and possible steps to improve the measurement and use of BOI data.

Measuring disease-specific BOI can inform policy. However, burden has many facets: morbidity, mortality, length of life, and quality of life. Mortality measures are important, but provide incomplete and insensitive information about overall population health. So-called summary measures capture more than one dimension of health, and they incorporate information on preferences for different health states. Therefore, summary measures also recognize the impact of physical and psychological conditions that cause individual suffering and disability and limit social and economic opportunities. Furthermore, these measures are relevant to both public health and medical decision makers. Actual and proposed uses include: describing differences and trends in the health of populations; informing decisions about alternate uses of public health care dollars; and assessing the cost-effectiveness of alternative medical interventions and consumer information programs. Finally, summary measures may also facilitate the assessment of the contribution of research to health.

There are currently five or six leading examples of summary measures, which include: the Quality of Well-being Scale (QWB); the Health Utilities Index (HUI); the EQ-5-D; the U.S. National Center for Health Statistics Years of Healthy Life; and the Disability Adjusted Life Years (DALY). These are distinguished by how the health status of each respondent is assessed and how a preference weight for each health state is derived. All of these measures have been found useful, and each has a constituency, including different countries, organizations, and large teams of creative people who have invested in their own measure. It is also known that each measure is flawed because it fails to address all dimensions of health or because it fails to detect small but important changes in health. Therefore, it is premature to select a single summary measure as the best and designate it as the instrument of choice for all future assessments. Instead, researchers should try to test the hypothesis that all of these measures will result in essentially the same implications for policy, despite the fact that they may get there by different conceptual routes.

Although summary measures are useful for assessing particular improvements in medical treatment, they do not explain how those improvements occurred. However, models that provide frameworks for synthesizing information can help explain the contributions of research to the improvement of medical treatments. Furthermore, these kinds of models can be used to determine whether a modestly effective treatment is worth pursuing. The important implication of this approach is that clinical studies not employing such models may sometimes be underpowered. That is, they may fail to detect clinically small, but important differences for public health. However, while models are technically possible to use, there are currently very few of them, they require a great deal of work to construct, and they are expensive to develop and manage.

Another possible next step is to gather information on multiple summary measures. This would allow investigators to test the hypothesis that each of the measures results in essentially the same

conclusions about health priorities. For example, information on two or more summary measures might be gathered during the next round of the AHCP Medical Expenditure Panel Survey (MEPS), a national data set that provides information on household use of health care and related medical economics data. The National Health Interview Survey (NHIS) operated by NCHS is also a potential mechanism for estimating summary measures. Comparison of multiple measures on the same data set could advance understanding of the cross-sectional and longitudinal behavior of summary measures, and facilitate development of a crosswalk between measures. The experience with multiple measure data sets might also encourage consensus around a new measure that combines the best aspects of the available measures. Amassing sufficient data to support crosswalks and comparisons of summary measures would require several years of research.

Recommendations

The participants in the June 28 meeting on BOI recommended a variety of follow-up actions based on discussions at the meeting. These can be subsumed under two categories: 1) conduct further research to improve measurement of BOI, particularly summary measures; and 2) explore the feasibility and value of developing models for specific diseases and conditions. With respect to the first recommendation, it was suggested that despite the numerous advantages of using existing summary measures of burden of illness, there are opportunities to refine and improve these measures. In carrying further research toward this end, several questions need to be addressed: What are the implications for constructing summary values and for tracking trends in BOI of alternative approaches for assessing health states and for assessing the preferences, utilities, or values assigned to different health states? How sensitive are values of the summary measure to changes in subcomponents such as mortality, incidence, or specific dimensions of functioning and quality of life? How do summary measures respond to changes in disease or health problems over time? How sensitive are these measures to advances resulting from the NIH portfolio? How do preferences or relative values attributed to each health state change over time?

A recommended next step toward addressing these research questions is that a temporary trans-agency working group should be established at NIH. Because there is shared interest and responsibility in this area, the working group could facilitate cooperation and determine where responsibilities lie among agencies for sharpening measures. It was also recommended that the ICs collaborate and share access to information generated on surveys and related long-term studies. Likewise, much of the health burden data collected throughout the Federal Government should be linked across agencies.

Finally, it was concluded that any strategy to develop summary measures of health must address the following seven challenges (all are candidates for further research):

1. Attribution of reduced quality of life and premature mortality to disease or condition in the face of comorbidities.
2. Non-independence of health state weights and time in state. People adapt to a lower than optimal health state over time and increase the value they place on the health state. The general public, for example, values prevention of blindness more than the blind value a cure. The

process of adaptation must be understood and the empirical, moral, and ethical implications must be given due consideration.

3. Consensus bound on maximum life span for measuring premature death. Before years of life lost to premature mortality can be measured, there must be agreement on maximum life expectancy.
4. Imputation of summary measures where data are sparse. Sparse or incomplete data are often a problem for subpopulation and measurement over short time frames.
5. Cohort effects and future projection of mortality. Future burden as well as present burden must be considered in order to assess priorities and progress. Right now, projections of the values of burden measures are based on ad hoc judgments.
6. Health condition of one may affect others (caregivers, family). Assessment of one family member's health may depend on that person's perception of the health of other members. If so, health of the family unit is not necessarily a simple summation of the independent assessment of each family member's health.
7. Who speaks for "societal" decision maker? How do we assess social or societal preferences for different health states as opposed to personal preferences?

Psychology and Economics Workshop

In order to explore the nexus of psychology, economics, and medicine, a workshop was organized where the focus was on issues related to behavioral decision making. Presentations included the measurement of welfare outcomes of economic policies, decision-making, cognition, and aging; learning, bounded rationality, and experimental methods; planning and mis-planning for the future; and research opportunities at the interface of psychology, medicine, and economics. Specific applications included: saving and investing over the life cycle; asset allocation of retirement savings; spending behavior of older adults; and patient and provider decision-making in the health care system.

The ways in which people make decisions are influenced by a combination of psychological, economic, social and environmental factors. Achieving an understanding of these converging influences is necessary for explaining more comprehensively the decisions people make as they age, as well as the investments they make to prepare for later life. Attitudes about risk, the social environment, and community interactions are examples of the psychosocial influences that can have significant effects on behavior. These influences may be expressed in a wide range of aging-related decisions, as in the areas of work and retirement, saving and consumption, fitness and nutrition, health and long-term care, compliance with medical treatment regimens, and living arrangements.

This area of research can be illustrated by questions concerning how people make decisions about saving and investment allocation. For example, to what extent does the workplace environment and workplace education influence the amount that people save in employer-sponsored saving plans? To what extent do future expectations about retirement or health or longevity affect saving? To what extent do attitudes about risk, or personality measures such as

“conscientiousness” relate to saving? By simultaneously considering the economic, social, psychological, and environmental influences, a much more comprehensive understanding of aging-related decisions is likely to emerge.

Recommendations

Several recommendations for further research emerged from this workshop. First, with respect to the use of experimental methods for integrating research in economics and psychology, it was noted that standard experimental measures have been developed for evaluating attitudes toward risk, uncertainty, delay of gratification, and reciprocal altruism. These measurement tools also provide ways to compare survey results with more controlled (albeit less realistic) experiments. Although many studies have investigated how measured attitudes vary with gender, culture, SES, intellectual background (e.g., college major) and so forth, there are few studies that compare older and younger adults. This should be an important focus of new research.

Another recommendation concerns planning and mis-planning for retirement. Recent research documents lifecycle variation in attitudes toward spending during retirement; however, new research is needed to measure and explain lifecycle variation in consumption tastes and to link this variation to normative theories of saving. It was also recommended that policy tools should be evaluated using the tools of behavioral economics (i.e., theories of bounded rationality, self-control, peer group effects, and learning). Finally, it was proposed that future research efforts be directed toward exploring the psychological foundations of economic evaluation, because summary measures such as quality-adjusted life years do not adequately reflect the values of medical patients, and thus are sorely in need of further validation and refinement.

Workshop on Old and New Measures of Human Well-Being, and Their Possible Significance for Policy

Theoretical models of well-being involve some integration of the following: hedonic experiences, global retrospective evaluations, mood, personality dispositions, emotional state and other on-line measures of experience, social and economic resources, social comparisons, and life-span considerations, historic and expectations. The central purpose of this meeting was to identify the major obstacles preventing the efficient and accurate measurement of well-being. In order to better understand the nature of the various recommendations for research that emerged from the workshop, several issues and topics will first be summarized, including: 1) basic measurement issues; 2) treadmill effects; 3) affective experience: biological correlates and potential ambulatory measures; 4) life history: cumulative advantage and adversity, allostatic load, and mortality; and 5) well-being as a measure of real value and its potential to affect public policy.

Basic measurement issues. Currently, there are two main types of measures of well-being: 1) retrospective self-report measures by which participants respond to questionnaires, basing their responses upon their memory and integration of experience; and 2) moment-based measures that probe the present and recent past using experience sampling methods to collect psychological and physiological data. Theoretically, aggregates of moment-based indicators of affect and arousal should predict subjective and global measures of well-being. However, this is not always the case. Despite the tendency of researchers to consider the results of moment-based measures as "the real thing," there are advantages and disadvantages to each approach. Memory-

based measures are important because they measure what people think they experienced; these include assessments of the practical and diverse aspects of past experience; measuring the memory and the meaning of past experience (which is all we retain from past experience); and these measures are based upon people's intuition about well-being. The disadvantages of memory-based measures are: affective memory is poor, responses are biased by recency effects and duration neglect, integration is particular, and responses are susceptible to scaling.

Moment-based measures probe the present and recent past, and are based upon more objective instrumentation that is less susceptible to scaling effects. These measures can serve as criteria to improve retrospections required for memory-based measures, are useful for group comparisons, and are important to the evaluation of different settings. However, they are cumbersome, expensive, and rely upon the researcher to infer an individual's overall sense of well-being. Ecological Momentary Assessment (EMA) measures take into account the value of situatedness for understanding well-being. The advantages of using this approach are that phenomena are assessed as they occur, with repeated observations frequently employed and measurements made in the environs typically inhabited by the research participants. Moreover, this approach does not require memory or biased summation. The challenges of EMA include participant burden, compliance, and reactivity, and include increased costs, large volumes of data to manage, and questions of construct validity.

The workshop revealed that those who view well-being as a summative construct believe it is best measured through the accumulation of emotional, physiological, and subjective momentary evaluations. Nevertheless, they are reluctant to dismiss personal evaluations of general well-being, even as they recognize that self-reported measures of well-being may not accord completely with more objective evaluations.

Treadmill effects. Despite large cultural differences in well-being, research indicates that the effects of life circumstances on well-being are quite small. Among the explanations is the hedonic treadmill, which suggests that increased resources do not provide more pleasure; rather, they increase the requirements necessary for pleasure. In other words, people seem to adapt to improving circumstances to the point of affective neutrality; thus the improvements yield no real benefits. An alternative construct proposed by Daniel Kahneman is the satisfaction treadmill, which consists of a change in the aspiration level for pleasures. That is, some people (e.g., wealthy individuals) could perhaps experience more actual pleasure than others, but if they demand more pleasure than others do, they will not be *subjectively* more satisfied or happier than others. From this perspective, one's sense of well-being is established with respect to recent, current, and potential experience. Kahneman also suggests the following clever test of the satisfaction treadmill hypothesis: Compare two groups who have pronounced differences in physical resources by examining the statistical interaction of self-rated well-being and the ratio of pleasure to pain. That is, differences in resources are reflected in differences in the amount of pleasure required for satisfaction.

Affective experience: Biological correlates and potential ambulatory measures. The valence (pleasant-neutral-unpleasant), duration, and intensity of affective experience is considered fundamental to well-being. Scientific knowledge and understanding of affective experience is accumulating by building upon an integrative research paradigm that examines brain anatomy and activation patterns, physiology, individual response patterns, and self-report of experience. For example, affective style, prefrontal asymmetry, and prefrontal activation have

been linked to individual differences in cortisol and immune function. This kind of knowledge provides the opportunity to develop ambulatory techniques for measuring anatomical and physiological correlates of affective experience and affective style, including measures of brain asymmetry and prefrontal activation, startle response, salivary cortisol, and autonomic function. Although these techniques are not highly consistent and reliable, they can provide important complements to self-report measures and measures of cumulative adversity.

Basic research that examines the relationships among these proposed ambulatory techniques and their relationship to measures of cumulative adversity could enable researchers to understand more precisely the relationship between memory-based measures and biological markers. Information gained from basic research could then be used to examine the impact and/or utility of interventions designed to improve human welfare (e.g., to improve recovery from illness; to improve quality of life of those who live in assisted living facilities; and to decrease illness frequency and severity).

Life history: Cumulative advantage and adversity, allostatic Load, and mortality. Allostasis is an adaptive, momentary state achieved in response to ongoing challenges to physiological systems. Systemic processes involved in achieving allostasis include the regulation of hormones, blood pressure, heart rate, blood sugar levels, and immunological functions. The concept of allostatic load is based upon this theoretical model: The effects of adverse stresses and events accumulate across the lifespan and eventually break down the systems involved in maintaining allostasis. In other words, the systems show signs of wear and tear. Allostatic load is considered an indirect measure of this wear and tear. Interestingly, socioeconomic hierarchies fail to predict later-life measures of allostatic load, which are also measures of morbidity and mortality. However, it appears that the quality of early life and/or adult relationships may play an intervening role that deters the effects of early life economic hardship. Measuring allostatic load is grounded in the notion that life history is reflected in an individual's current physiology and physiological responsiveness.

Well-being as a measure of real value and its potential to affect public policy. Aggregate measures of well-being may be inadequate for placing value on some goods and services. Estimates of market value are also inappropriate for some goods and services that service overall well-being. Economists tend to think that market value is a good indicator of people's preferences. However, people sometimes have preferences for things that either have no market value or are given the wrong market value. Those preferences may not be considered when calculating the gross domestic product (GDP) or, when included with inaccurate measurement, can skew the outcome. The error in excluding or under/over estimating important preferences leads to bad public policy. Including factors which are elusive to measurement and which may be essential to well-being is important to the development of micro-government policy related to public goods (e.g., deciding to allocate funds for tree-lined streets or highways that shorten daily commutes); to how the government chooses to serve the public (e.g., shorter lines, electronic tax filing); and to the development of regulations for such things as occupational safety and pollution control.

Moment-based measures of well-being can provide crucial data when attempting to determine the relative importance of "non-market goods and services"; however, these aggregate measures of well-being may exclude factors important to well-being. For example, aggregate measures may indicate the value of work without considering the relative value of spending time with

children. Moreover, they may not indicate the value of living in a pollution-free environment, which may be underestimated by considering health costs alone. It is clear that non-market goods and services should be considered for the development of public policy and financial resources allocation. Furthermore, the value of these items may be determined with judicious use of both moment- and memory-based measures of well-being. Two key questions are: 1) How do economists incorporate the importance of non-market goods and services into models that influence public policy? 2) What measurement techniques are the best indicators of the real value of particular goods and services?

Recommendations

Several families of tasks were specified in advance of the workshop, and the following summarizes some of the ideas and suggestions that emerged from the meeting:

1. Investigating the conceptual structure underlying well-being research (i.e., levels at which this can be studied) and mapping the connections between the measures employed at these levels (e.g., mutual validation of physiological and self-report measures of affect, including tests of hypotheses about unique variance; the relationship between willing-to-pay and hedonic outcomes; and correlations between life circumstances and health outcomes). In addition, it was suggested that these measures be embedded at various levels into large-scale panel studies.
2. Understanding the nature of adaptation to circumstances is of critical importance for several reasons: a) it may identify a crucial bias in standard measures of subjective well-being, or it may reveal that the bias is minor; b) if the facts about adaptation differ from lay theories, as research suggests, an educational effort at making people hedonically-literate may be useful (medical decisions and decisions about retirement are obvious areas of application). Examples include: understanding different aspects of allostasis (i.e., physiological and psychological); and examining hedonic adaptation and the “satisfaction treadmill” (i.e., a change in the aspiration level for pleasures).
3. The measurement of experienced utility and well-being as a policy tool: a) develop a large-scale national panel study of various aspects of well-being, to be used as an index of the state of society to supplement GDP; b) add measures of well-being to all large scale experimental programs, in which people are randomly assigned to different treatments (e.g., school vouchers, non-ghetto housing, health insurance); c) measure mood and stress in different settings—a tractable problem in principle, even though the techniques to do so are yet to be developed.

In addition, several major research needs were identified: relate memory-based measures to moment-based measures; determine what should be used as a base for construct validity (e.g., mortality); conduct large panel studies, but the greater immediate need is laboratory studies; employ approaches from several disciplines (e.g., personality, social interaction, physiology, anatomy, health and medicine, economics); improve technology.

Research is also needed for developing measurement tools that can serve multiple purposes (e.g., from clinical research to public policy) and that can interface the developed tools with the various research needs and situations for which they are designed. However, the trans-

disciplinary, multi-institutional effort required for making progress toward these goals will naturally be expensive, and thus may best be served through the program project mechanism. Additional mechanisms and funding options under consideration include: a program announcement with a set-aside, a collaborative RFA, interagency efforts, infrastructure development, and graduate training programs.

* * *

The meeting agenda and list of meeting participants follows.

OLD AND NEW MEASURES OF HUMAN WELL-BEING, AND THEIR POSSIBLE SIGNIFICANCE FOR POLICY

National Institute on Aging, National Institutes of Health

January 27-28, 2000

Agenda

January 27, 2000

8:00 p.m. **Introductory Remarks**
Richard Suzman, Associate Director, BSR, NIA

**Alternative Measures of Well-being and the
Ambiguity of Adaptation**
Daniel Kahneman, Chair, Princeton University

General Discussion

9:15 p.m. Adjourn

January 28, 2000

8:30 a.m. **Brain Measures of Affect**
Richard Davidson, University of Wisconsin

9:05 a.m. **Long-term Measures of Stress Exposure**
Burton Singer, Princeton University – presented by
Daniel Kahneman

10:00 a.m. **Experience Sampling Methodology**
Arthur Stone, SUNY, Stony Brook

10:35 a.m. **Using Measures of Welfare in Policy Analysis**
Bill Dickens, The Brookings Institution

11:30 a.m. **General discussion**

1:00 p.m. **Discussion of specific issues in small groups**

1. Research designs for the study of adaptation to life circumstances
Schwarz (Chair), Juster, Loewenstein, Schkade, Stewart
2. Research designs for the evaluation of public goods and public policy
Kahneman (Chair), Dickens, Frank, Garber, Laibson, Oswald, Wise
3. Research on the development of measurement tools
Diener (Chair), Cacioppo, Davidson, Fryback, Kessler, Stone

1:45 p.m. **General discussion**

2:45 p.m. **Discussion of basic research agenda in small groups** (A 5-10 years research agenda, assuming unlimited resources)

3:15 p.m. **General discussion of a potential research agenda**

Old and New Measures of Human Well-Being, and Their Possible
Significance for Policy
January 27-28, 2000

LIST OF PARTICIPANTS

John T. Cacioppo
Department of Psychology
University of Chicago

David I. Laibson
Department of Economics
Harvard University

Richard J. Davidson
Wisconsin Center for Effective
Science
University of Wisconsin-Madison

George Loewenstein
Department of Social & Decision
Sciences
Carnegie Mellon University

Ed Diener
Department of Psychology
University of Illinois

Andrew Oswald
Department of Economics
Warwick University

Robert Frank
Department of Economics
Cornell University

David Schkade
Business School
University of Texas

Dennis G. Fryback
Departments of Preventive Medicine
And Industrial Engineering
University of Wisconsin-Madison

Norbert Schwarz
Institute for Social Research
University of Michigan

Alan Garber
College of Medicine and
Department of Economics
Stanford University

Anita L. Stewart
Institute for Health & Aging
University of California San
Francisco

Thomas Juster
Institute for Social Research
University of Michigan

Arthur Stone
Department of Psychiatry
SUNY, Stony Brook

Daniel Kahneman
Department of Psychology
Princeton University

David Wise
JFK School of Government
Harvard University, and
Natl. Bureau of Econ. Research

Ronald Kessler
Medical School
Harvard University

Alan Krueger
Department of Economics and WWS
Princeton University