Commentary

Choosing Measures of Health Status for Individuals in General Populations

JOHN E. WARE, JR., PHD, ROBERT H. BROOK, MD, SCD,
ALLYSON R. DAVIES, MPH, AND KATHLEEN N. LOHR, PHD

Abstract: This paper offers suggestions to aid the selection of appropriate instruments and data gathering methods for studies that require measures of personal health status applicable in general populations. Before selecting measures, the reason for studying health status must be identified. Next, definitional issues arise when attempting to specify the components of health that are to be studied. Evidence supports restriction of the definition of personal health status to its physical and mental components, rather than including social circumstances as well. In evaluating the suitability of available measures, three features must be considered: 1) practicality in terms of administration, respondent burden, and analysis; 2) reliability in terms of the study design and group or individual comparisons; 3) validity, in terms of providing information about the particular health components of interest to the study. Evaluating validity will be difficult for most available measures; careful attention to item content will be helpful in choosing appropriate measures. Despite problems in development and interpretation, overall health status indicators will prove useful to many studies and should be considered, as should both subjective and objective measures of health status. Given that the reasons to measure health have been identified, the aspects of health to be measured specified, and attention paid to their suitability, appropriate measures may often be found among those now available. (*Am J Public Health* 1981; 71:620–625.)

Health status is a broad concept, and many issues complicate its definition and measurement. Advances in the methods used to measure health status have taken place during the past decade, although more may be needed. This paper offers suggestions that may prove helpful in selecting appropriate instruments and data gathering methods for studies that require health status measures. What we propose is something of a "shopper's guide": not a catalog of specific measures or brand-name recommendations, but a list of things to think about while looking for health status measures.

Address reprint requests to Dr. John E. Ware, Jr., Senior Scientist, Rand Corporation, 1700 Main Street, Santa Monica, CA 90406. Dr. Ware is also Lecturer in Public Health and Staff Psychologist, Department of Medicine, University of California at Los Angeles. Dr. Brook is Senior Staff Health Services Researcher at Rand Corp., and Associate Professor of Medicine and Public Health, UCLA. Ms. Davies is Health Services Researcher, Rand Corp., Santa Monica; Dr. Lohr is Associate Social Scientist, Rand Corp., Washington, DC. This paper, submitted to the Journal September 15, 1980, was revised and accepted for publication December 8, 1980.

We cannot cover all the issues involved in gathering and interpreting health status data, or the varied contributions to health status measurement that have appeared during the past ten years. We will note some issues that we consider important, and some we think have been overemphasized in the literature.*

Why Measure Health Status?

When searching for measures of health status, one first needs a clear understanding of the reasons for studying

*We draw our comments on the field of health status measurement largely from our experiences during the development of health status measures used in Rand's Health Insurance Study (HIS). These experiences and extensive reviews of the literature are presented in detail in a series of Rand publications (see Rererences/ including their extensive bibliographies). We do not mean to imply that the HIS measures necessarily fulfill all the criteria that should be considered, or that they will be useful for all studies requiring health status measures.

health status. These reasons seem to fall into five broad categories:

1. *Measuring the efficiency or effectiveness of medical interventions*. Health and medical interventions should be designed either to improve health status without placing untenable strains on health care budgets or to contain or lower costs without impairing health. Thus, health status must be considered in any equation defining benefits or effectiveness of interventions.

2. *Assessing quality of care*. Health status measures are important to evaluations of medical care in terms of patient outcomes (in contrast to measuring characteristics of providers or processes of care).

3. *Estimating the needs of a population*. Health status measures are useful tools for describing population health levels. Such information can be used for areawide health planning activities that must anticipate need for specific services or facilities. It can also serve as a benchmark by which to judge the results of health planning efforts. In addition, information about the health of populations can be used in making decisions about how to allocate health resources among programs, areas, or regions.

4. *Improving clinical decisions*. Standardized health status measures can be an adjunct to patient-specific information collected by providers of care. For example, a comprehensive health status summary could be incorporated into the more usual history and workup of a new patient.

5. *Understanding the causes and consequences of differences in health*. Health status measures can be used to study changes in health over time and associations between health and other variables (such as attitudes toward seeking care or medical care consumption) when one is developing and testing theories about ways to improve health status in general populations.

A full discussion of the implications of these five reasons for the choice of health status measures goes beyond the scope of this paper, although we can offer some general principles:

- When studying general populations, consider using positively defined health measures. Only some 15 per cent of general population samples will have chronic physical limitations, and some 10 to 20 per cent will have a substantial psychiatric impairment.[1, 2] Relying on these negative definitions of health tells little or nothing about the health of the remaining 70 to 80 per cent of general populations.
- By contrast, when studying severely ill populations, the best strategy may be to emphasize measures of the negative end of the health status continuum.
- To evaluate outcomes of a specific intervention, use measures that reflect the most likely effects of that intervention. Little can be gained by measuring something that the intervention cannot possibly change, or cannot change within the time frame of the study.

What Aspects of Health Are of Interest?

A second requirement when selecting health status measures is a clear statement of the aspect of health being studied: specifically, what question about health status is to be addressed by a particular study? This requirement addresses both the scope and definition of health status measures. Issues relating to definitions of the health of families, communities, or nations, and to definitions that rely on secondary data sources such as mortality statistics, go beyond this overview. Instead, we address issues of defining health at the individual level: what is personal health status?

A good dictionary yields some clues: "health" connotes "*completeness*", nothing is missing from the person; it connotes "*proper function*", all is working efficiently; it suggests "*well-being*", more than just freedom from disease (e.g., feeling vigorous). Finally, whereas dictionary definitions clearly emphasize the soundness of the *body*, they also introduce us to the concepts of *physical* and *mental* health; in other words, health status has at least two major components.[3]

With one exception, this definition sounds very much like the often-quoted definition of health that the World Health Organization (WHO) offered three decades ago. In its constitution, the WHO described health as a "state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity."[4] The major difference between the WHO and dictionary definitions is that WHO includes a social component. This may reflect WHO's interest in addressing all of society and not limiting its perspective to the individual.

Physical health refers to the physiologic and physical status of the body. An indicator of the former is blood pressure; of the latter, walking. Mental health refers to the state of mind, including basic intellectual functions such as memory and feelings. An indicator of the former is short-term recall of a list of numbers; of the latter, ratings of mood or affect. Questions about feelings toward your body—whether it hurts, whether you are happy about it—address the interface between physical and mental health. Although physical and mental health are distinct concepts, they are also substantially interrelated; the state of the one often affects the state of the other.

Our research and literature reviews have identified several reasons for restricting the definition of personal health to its physical and mental components.[5] In a model of health status at an individual level, physical and mental variables are similar in that they "end at the skin." They do not directly involve other people or factors outside the individual. By contrast, social functioning extends the concept of health beyond the individual to include the quantity and quality of social contacts and social resources. In a model of health status containing social variables, a change in social support (such as death of a loved one) by definition indicates a change in personal health status. Likewise, one of two persons enjoying the same level of physical and mental health would be considered less healthy if that person resided in a strife-torn community or was separated from family members.

A model of health status that defines social factors (along with others such as life events) as external but related to an individual's health status explains empirical results better than one that includes social factors as an integral com-

ponent of individual health. Such a model acknowledges that social circumstances may directly affect health status and that they may lessen the effects of such factors as stressful life events on health status.[6] It does not, however, define personal health status in terms of social circumstances.

How Suitable Are the Measures?

Another important issue to keep in mind when choosing health status measures is their relationship to the underlying expressions of health or disease that you really want to measure. Health status cannot be observed directly. One can only make inferences about health from fallible indicators. Having defined the aspect of health you want to know about, the next step is to judge the suitability of the available measures. Thus, the next few points are critical for judging how practical, reliable, and meaningful are the measures chosen, keeping in mind your study design and what you want to know about health.

Practicality

Taking a good look at the total measurement resources available and deciding how much can be devoted to health status are helpful first steps. The sensible next step is to establish priorities for allocating those resources to various health status concepts. Practical considerations will determine whether any given health status indicator can be considered. For example, if health is one of many things that need to be measured, the amount of time and money devoted to health status is obviously limited. Relevant questions are whether you can afford interviews in person or over the phone and whether self-administration of standardized instruments would work just as well. The measurement of health status by means of self-administered questionnaires has advanced considerably in recent years, and such questionnaires can be fielded less expensively than personal interviews.

An important aspect of practicality is respondent burden, indicators of which include refusal rates, rates of missing responses, and administration time. At the extreme, potential respondent burden can also entail risk of loss of life, e.g., the risk of administering a cardiovascular stress test to persons with heart disease in order to help plan future activities; psychological threats and risks are often associated with questions about sensitive and embarrassing topics.

Finally, those developing and using health status measures should work with the least complicated instruments and methods possible. At every step along the way—the task presented to respondents, the difficulties faced at the time of scoring, the complexity of interpretation—the simplest approach should be adopted. The introduction of complexity, such as items that require reversals in coding, or that must be standardized or weighted before they are summed, makes it more difficult and sometimes impossible for others to use and understand the same instruments, methods, and results.

Reliability

Health status scores are made up of several elements. Part of the score is nothing more than "noise" or random error. A reliability coefficient indicates the proportion of information, rather than random error, that a score contains. Thus, a reliability coefficient of 0.80 means that the score contains 20 per cent noise.

The degree of reliability needed depends on the purpose of the study. Generally, a more reliable measure is required to assess health status on a person-by-person basis (such as in clinical decision-making) than to compare two groups of people (such as those receiving different treatment interventions): a health status measure with a reliability of 0.50 may be acceptable for comparing two groups that are likely to show substantial differences in health status, whereas a reliability of 0.90 is required to be confident in the score assigned to one person.

Unfortunately, most publications about health status measures report little or nothing about reliability.[7] Knowing something about reliability is critical, however, because along with practical issues such as respondent burden, the measures must achieve the minimum standard of reliability necessary for a study's design and purpose. Predicting health and illness behavior is difficult in part because many measures of behavior have proven unreliable.

Some rules of thumb may help. Usually, poorer reliability can be expected from short scales—a single-item measure rather than a multi-item scale. This typically holds for both objective (behavioral) definitions of health and more subjective ratings. Reliability also tends to be lower for disadvantaged persons (those with less education or lower incomes). It is best to have the "worst case" in mind and select health status measures on the basis of the expected reliability for groups that will provide the least reliable scores.

Typically, higher reliability coefficients cost more than lower ones because they require more information—more items or more observers. If the analyses will compare groups only, striving for very high reliability coefficients may not be the best use of scarce measurement resources. Settling for lower reliability and using the remaining resources to assess another important variable may be a better strategy.**

Validity

Validity focuses on the meaning of information contained in the score on a health status measure. A valid score contains information about health status, not some other variable. More specifically, it contains information about the particular aspect of health status needed for the study and the analyses planned. Unfortunately, the field is just beginning to evaluate the validity of health status measures. Be prepared for considerable difficulty when attempting to determine whether a given health status measure will be valid when used in a particular study.

Validity can be studied in several different ways. Some are empirical, such as concurrent, construct, and predictive

**Veit CT and Ware JE: Differences between single- and multi-item measures of health and health care preceptions. Paper presented at the American Public Health Association's Annual Meeting, New York City, 1979.

validity. Nonempirical approaches include face and content validity.

Face validity refers to what an item appears to measure based on its manifest content. Content validity refers to how well a measurement battery covers important parts of the health component(s) to be measured. Although both can be very useful in selecting among many possible health status measures, there seems to be general prejudice against using evidence of face and content validity. This prejudice is unfortunate for at least two reasons.

First, analyses of face and content validity are relatively easy to do. All that is needed is a copy of the instrument and a good idea of what you want to know. A look at the content of specific tests in a medical screening examination or of items in an interview schedule can tell much about what the tests measure or the meaning of responses to the items.

To illustrate, some general-population mental health measures have failed the crucial test of discriminant validity: they fail to correlate more highly with other mental health measures than with measures of physical health. This is because these general-population measures include a great many questions about physical health and health habits (somatic complaints, physical limitations, etc.) in addition to questions that explore mental health directly.[8] This state of affairs becomes obvious by simply noting the basic content of the items in these various measures. In brief, it is not a good idea to select measures solely by the labels assigned to them or by the names of health status batteries.

Examining the content of all health status measures considered will help avoid problems that arise because different measures are often given the same label, and the same variable is often labeled differently. To reject the value of assessing face and content validity is to lose a first line of defense against selecting the wrong health status measure and to fall into the mire of confounded definitions of health status and other variables, a major problem in the field.

Another very practical reason for examining face and content validity is that there is rarely more than this available to use in judging the validity of most health status measures. Without exception, available empirical information about validity (concurrent, construct, and predictive) falls short of what is needed.[9]

Most of what is known about "validated" health status measures pertains to how much information they provide about health rather than about other variables, such as attitudes toward medical care or satisfaction with care. To know about health status may seem enough, but it is not. Knowing what component of health the measure reflects is also important; for instance, whether a physical health measure correlates very highly with another physical health measure and not very highly with measures of mental health. Few studies examine validity this thoroughly. In addition, very little is known about the extent to which variables other than health status (such as various behavioral propensities) influence scores.

The literature includes hundreds of studies on health status measurement, and the number of standardized health status measures is increasing. Although this situation can be viewed as a healthy one, the increasing availability of standardized health measures poses a real danger. A particular measure may be tempting because it has been "validated." A "validated" measure may not be valid for the purpose of a given study, however. The best measure of "X" may be of no value if "Y" is the concept to be measured. For this reason, the amount of research that has been done to develop and validate a measure should not dictate choice. Before selection, critically review available measurement research findings to determine whether the measure is valid for purposes of a particular study.

Many health status measures are put forth as an unbreakable package: the measure is fragile and something terrible will happen if items are left out, or so the argument runs. One may inherit a lot of "excess baggage" by adopting this view. Consider an example from the field of intelligence testing. Intelligence has multiple components and numerous indicators for each component. The practice of extracting one indicator from an IQ battery for a specific purpose has been employed successfully. Likewise, using a subset of health status indicators carefully selected from a comprehensive battery may be better, if those indicators pertain most closely to what you want to know. On the other hand, investigators often find themselves in the position of evaluating interventions without knowing their most likely effects on health. In this situation, a very comprehensive battery is a good strategy.

In summary, one should assure oneself that the content of health status measures to be used encompasses both the component of health (physical, mental) and the specific aspects of each component (such as physical abilities and emotional stability) to be measured, and that the measures are not excessively confounded with other variables that will simply confuse, if not bias, the results.

Objectivity vs Subjectivity

Health status concepts and measures also differ in terms of their objectivity and subjectivity. Typically this distinction is based on the extent to which the measured variable is observable vs the extent to which inference must be used to interpret it. In the physical health area, for example, objective variables include whether a person can dress without assistance, walk, or run. More subjective assessments of physical health would be personal ratings of overall physical shape or condition.

The field has generally favored more objective health status definitions. This preference was originally based on the argument that subjective ratings were not reliable. This argument is no longer true: Subjective health measures have been constructed that more than satisfy the reliability standards mentioned earlier.[10]

Currently, the preference for objective measures seems to be based on validity arguments—for example, that personal ratings of health do not agree completely with ratings by trained professionals. This argument is also less than convincing. First, professionals do not agree completely among themselves. Second, patients and providers rely on different information in assessing health, and their ratings should not agree completely. Providers have information that they often do not share with their patients. Further, providers may

rarely ask their patients how they feel about their health; therefore, they do not benefit from their patients' views.

Whether the objective or subjective approach is more valid depends, again, on your purpose. Suppose you want to estimate demand for medical care. Asymptomatic conditions obviously do not create much patient-initiated demand for medical care (at least not until they have been diagnosed). Hence, objective measures of such illnesses would be understandably poor predictors of patient-initiated visits. Subjective measures of what people think about their health, regardless of whether they are right or wrong, have proven to be valid for such purposes.[11] Another argument in favor of the more subjective measures is that they are proving to be more precise than the so-called objective measures; they permit finer discriminations among people throughout the full range of the health status continuum.[12]

In short, what is needed is a better understanding of the associations between objective and subjective measures of health. Until studies of these associations are done, we suggest using both kinds of health status measures.

Should Global Indexes Be Included?

Considerable effort has been devoted to the search for an overall health status index. The tradeoff seems to be between the simplicity of a single indicator of health and the loss of information that results from the aggregation of very different health status variables and the erroneous inferences to which this might lead.

To illustrate, consider the various economic indicators often quoted on the evening news: the Consumer Price Index (CPI), the unemployment rate, the direction of trading on the stock market. These indicators do not always agree, and sometimes behave counter to economic theory. Further, a change in any one indicator may not accurately reflect all sectors of the economy; for example, the CPI may remain constant while prices go up for one commodity and down for another.

Health status measures present much the same problem. Composite or global indicators are an imperfect way to summarize the state of a person's health. At best, within both physical and mental health, we are ready to reduce the number of indicators needed to a meaningful few. With the possible exception of functional status, no indicator of physical morbidity yields scores that can be interpreted at both the extremes and in between.[13, 14]

One problem, then, with using a global indicator is that scores may be crude, difficult to interpret, and misleading. Put another way, health status is like fruit in a bowl. What is the average fruit? How can we add and subtract apples and oranges? Determining answers to such questions is beyond the current state of the art of health status measurement.

Despite this statement, pessimism about developing global health status indicators that meet the suitability criteria we noted above is not entirely justified. The effort required to solve the many problems involved in developing a reliable and valid overall index of health status will be substantial, but we believe it will be rewarding. Such an index, for example, would allow direct comparisons among programs designed to achieve entirely different health status outcomes and might provide a basis for making difficult decisions regarding the best allocation of scarce resources among competing programs.

Areas for Future Research

Future research should address several high priority issues. The use of health status data during the health care planning process has generated much interest. Unfortunately, the validity and precision of data from most convenient secondary sources can be questioned; both need to be studied further before secondary sources are used to make critical decisions. We need to know more about the distinction between performance and capacity measures of health status, how to incorporate the probability of transition from one health level to another over time, and about the values people place on different health states. Resolution of such issues should provide a variety of reliable and valid measures of the health components as well as overall indicators of health status.

Discussion

A number of health status measures have already been developed. None is perfect, and selections must be made carefully, according to the particular needs and resources of the planned study and the guidelines we have suggested. Except in special instances, new measures need not be developed completely from scratch. Given that the reasons for measuring health status have been identified, the aspects of health to be measured specified, and attention paid to the suitability issues noted above, we believe there is a good chance that appropriate measures and data gathering methods can be found among those now available.

A good place to begin identifying health status measures is the Health Status Index Clearinghouse at the National Center for Health Statistics (NCHS).*** The Clearinghouse provides a number of services that can be very useful to those looking for health status measures. Their quarterly bulletins present annotated bibliographies of research published both in the United States and foreign literatures as well as information about research in progress. Mailing addresses for authors and investigators are included in these bulletins so that interested persons can seek further information. The quarterly bulletins are also summarized yearly and literature searches can be performed by the Clearinghouse.

REFERENCES

1. Stewart AL, Ware JE, Brook RH, *et al:* Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. II, Physical Health in Terms of Functioning. R-

***For further information or to place your name on the Clearinghouse mailing list, contact Pennifer Erickson, NCHS, 3700 East-West Highway, Hyattsville, MD 20782

1987/2-HEW. Santa Monica, CA: The Rand Corporation, July 1978.
2. Ware JE, Johnston SA, Davies-Avery A, *et al:* Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. III, Mental Health. R-1987/3-HEW. Santa Monica, CA: The Rand Corporation, December 1979.
3. Oxford English Dictionary. Oxford: The Clarendon Press, 1961.
4. World Health Organization: Constitution of the World Health Organization, IN: Basic Documents. Geneva: WHO, 1948.
5. Donald CA, Ware JE, Brook RH, *et al:* Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. IV, Social Health. R-1987/4-HEW. Santa Monica, CA: The Rand Corporation, August 1978.
6. Ware JE, Davies-Avery A, Brook RH: Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. VI, Analysis of Relationships Among Health Status Measures. R-1987/6-HEW. Santa Monica, CA: The Rand Corporation, 1980.
7. Brook RH, Ware JE, Davies-Avery A, *et al:* Overview of adult health status measures fielded in Rand's health insurance study. Med Care (Supplement) 1979; 17:1-131.
8. Ware JE, Johnston SA, Davies-Avery A, *et al:* Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. III, Mental Health. R-1987/3-HEW. Santa Monica, CA: The Rand Corporation, July 1979.
9. Brook RH, Ware JE, Davies-Avery A, *et al:* Overview of adult health status measures fielded in Rand's health insurance study. Med Care (Supplement) 1979; 17:1-131.
10. Ware JE, Davies-Avery A, Donald CA: Conceptualization and Measurement of Health Status for Adults in the Health Insurance Study: Vol. V, General Health Perceptions. R-1987/5-HEW. Santa Monica, CA: The Rand Corporation, September 1978.
11. Manning WG, Newhouse JP, Ware JE: The status of health in demand estimation: Beyond excellent, good, fair, and poor. IN: VR Fuchs (ed): Economic Aspects of Health. Chicago: University of Chicago Press, in press 1981.
12. Ware JE, Davies-Avery A, Donald CA: Conceptualization and Measurement of Health Status for Adults in the Health Insurance Study: Vol. V, General Health Perceptions. R-1987/5-HEW. Santa Monica, CA: The Rand Corporation, September 1978.
13. Kaplan RM, Bush JW, Berry CC: Health status: Types of validity for an index of well-being. Health Services Research 1976; 11:478-507.
14. Stewart AL, Ware JE, Brook RH: Advances in the measurement of functional status: Construction of aggregate indexes. Med Care (in press) 1981.

Postgraduate Course on Child Abuse and Family Violence

Harvard Medical School's Department of Continuing Education has announced a new postgraduate course on "Child Abuse and Family Violence" to be held June 4–5, 1981 at the Children's Hospital Medical Center in Boston. The program will be under the direction of Eli H. Newberger, MD, Richard Bourne, PhD, JD, and associates. Guest speakers include Drs. Ann Burgess, Ann Harris Cohn, David Finklehor, Richard Gelles, and others.

This two-day course is structured to provide a comprehensive review of the field. Focused particularly for physicians, nurses, and legal and social work personnel who care for children and adolescents, it presents the current knowledge base with emphasis given to the goals, ethics, and methods of sound clinical practice. This is accomplished by counterposing research and clinical presentations on three principal themes: understanding physical family violence, the sexual misuse of children, and the basis and methods of interdisciplinary practice. The course is designed to be taken as a discrete, two-day exercise or as part of a five-day continuum, in coordination with the course "Assessment of Newborn and Older Infants."

Registration is limited. Fee: $180. This course has Category I accreditation on an hour by hour basis toward the AMA Physicians' Recognition Award. Credit also to be arranged for nurses and social workers.

Information and application forms for the course should be requested directly from Harvard Medical School, Department of Continuing Education, 25 Shattuck Street, Boston, MA 02115, telephone 617/732-1525.