Measurement of Outcome: A Proposed Scheme

BARBARA STARFIELD

The need to demonstrate that health care has an influence on health status is increasingly pressing. Such demonstrations require tools of measurement which are unfortunately not available. Development of instruments has been hampered by a lack of consensus on appropriate frames of reference, and there appears to be little agreement on what should be measured and what relative importance should be ascribed to different dimensions of health status. An approach that does not require the assignment of numerical values or weights to various aspects of health status and is applicable to all age groups within the population and to the whole spectrum of health problems rather than to specific medical diagnoses would seem desirable. A scheme that is based upon the development of a "profile" rather than a single "index" for describing health status is proposed in this paper. The model is a conceptual framework whose usefulness will depend upon efforts of a large number of researchers from many disciplines to develop instruments which can be incorporated in it.

Although the problems in development of the scheme are complex, I hope that it will focus attention on the relevant dimensions and facilitate improved coordination of efforts to produce ways to demonstrate what health care contributes to health.

The ultimate test of the efficacy and effectiveness of medical care is its outcome on patients. Unfortunately, this is not easy to determine. Part of the problem lies in the difficulty of defining "outcome"; another obstacle is the unavailability of satisfactory tools to measure it. This paper presents an approach for its conceptualization and measurement.

Definition of Outcome

The dictionary (Webster, 1969) definition, "final consequence" seems quite explicit and suitable for use in its ordinary sense in medical care evaluations. Donabedian (1966) in his classic paper entitled "Evaluating the Quality of Medical Care" makes the definition operational. He includes as outcome "recovery, restoration of function, and survival." In a subsequent article which pro-

poses and categorizes a large variety of indicators of the quality of care, 11 categories of health outcome and two categories of satisfaction (which he considers another type of outcome) are presented (Donabedian, 1968). Shapiro (1967) refers to outcome as "some measurable aspect of health status which is influenced by a particular element or array of those elements of medical care." Elinson (as quoted by others) has suggested as accepted targets for evaluation: "death, disease, disability, discomfort, dissatisfaction" (White, 1967). A sixth, "social disruption," was subsequently added (Sanazaro and Williamson, 1968).

Confusion arises, however, when the term is employed in other than its usual sense. If outcome means final consequence, is it meaningful to have an "intermediate outcome" (Barro, 1973), that is, an "intermediate final consequence"? Phenomena which are processes of medical care should not be confused with outcome. Thus, although hospitalization and compliance are occasionally used to indicate outcome, they are quite clearly not subsumed under the original definition. Hospitalization is not always an undesirable occurrence, and failure to hospitalize may be a manifestation of inadequate medical practice; an evaluation system which has as a goal a large reduction in hospitalization rates will have difficulty in controlling for the need for hospitalization in its study design. Similarly, compliance, although a highly desirable objective, is not always appropriate behavior, and therefore cannot be considered an invariably desirable end point. Health knowledge or health attitudes (which are also processes directed toward the attainment of desired end results), or phenomena which are more properly conceived of as inputs or structural characteristics of the system (such as costs, availability, or accessibility of care) are also not part of the concept of outcome. Although these are all legitimate and important areas of concern for the planning and evaluation of health care delivery mechanisms, they cannot be considered even proxy measures of outcome until such time as they can be shown to have consistent influences on health status.

Criteria for a Scheme to Measure Outcome

An appropriate scheme for measuring outcome, apart from being concerned with elements of health status, must bear some rela-

tionship to the process of medical care (at least theoretically). It must be suited to the type of condition for which medical care is provided and to the age of the patients under care and not be susceptible to varying interpretation due solely to the passage of time. Because the values imparted by society to different aspects of health status vary with time and place, the usefulness of a scheme of measurement will be enhanced if its components are not weighted according to their relative value (Grogono, 1973). The scheme would ideally allow for measurement of both the maintenance of good health and the relief of illness.

Multiple diagnoses and certainly multiple health problems are not uncommon in the general population; health status is therefore a reflection of interacting levels of health and ill health. As a consequence, a scheme for evaluating outcome should not be restricted to specific medical diagnoses. For the same reason, it should be independent of specific medical therapies, although suited to evaluation of the effect of the wide range of medical interventions. The activities of nonphysician health personnel as well as those of physicians should be amenable to evaluation by the scheme. And, last, the scheme should not require artificial distinctions between mental and physical ill health and should allow for recognition of their interrelationship.

A Proposed Model

This section will briefly describe an approach suited to describing the health status of a population as well as to evaluating the effectiveness of medical care for individual patients or groups of patients. At present, the model is a theoretical construct in which outcome is described as a profile. Seven categories of outcome are proposed; all are vectors (they have magnitude and direction). These seven categories are arrayed in parallel: longevity, activity, comfort, satisfaction, disease, achievement, and resilience (Fig. 1).

Longevity is a prognostic category involving an estimate of life expectancy as well as a reflection of the current state of affairs (alive or dead). It is a measure of the actual or expected *duration* of life which can, of course, be quantitated; the results may be expressed either directly in numbers or indirectly as a percentage of normal life expectancy.

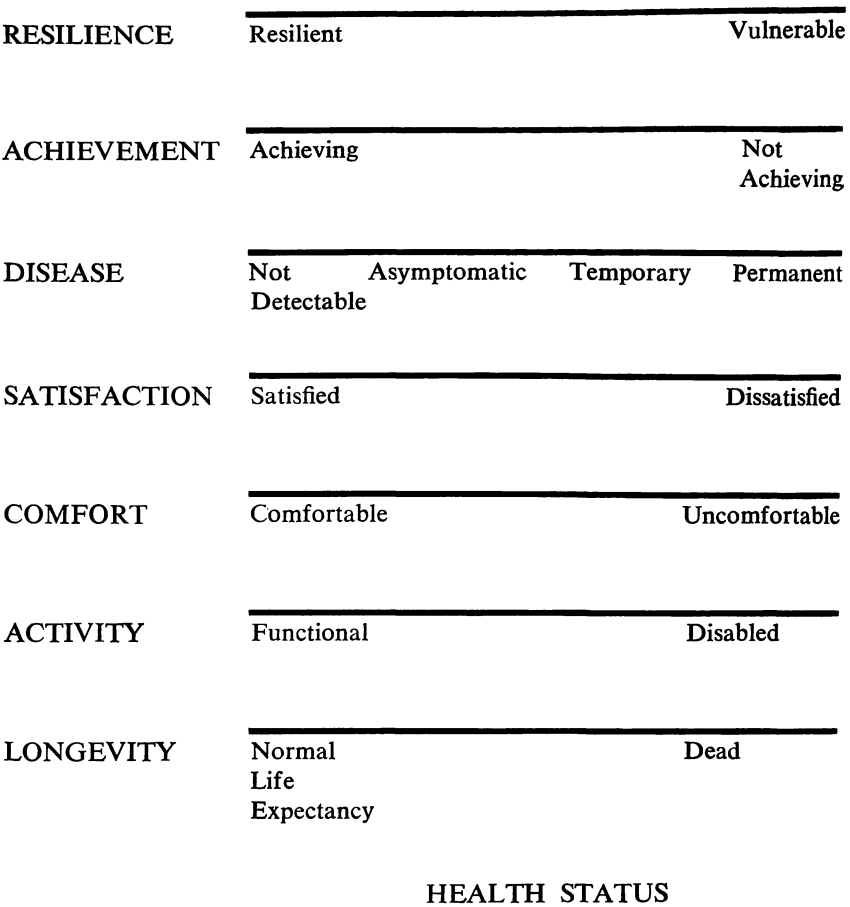

FIG. 1. Categories of outcome.

Activity is the category of outcome which has been the focus of considerable attention. Virtually all schemes for evaluating outcome rely heavily on descriptions of *functional capacity* of the individual. One example is the Activities of Daily Living used by the Commission on Chronic Illness in the United States (1957) and more recently by Katz et al. (1972) in adults with chronic illness. Another approach is taken by the National Center for Health Statistics (Sullivan, 1971). In its National Health Survey, two types of activity limitation are used: restriction to bed, and restriction of "usual activities" (which is defined differently for different age

groups). While attendance at school serves as the "usual activity" for children over five years of age, the concept has been difficult to make operational for preschoolers and especially for infants. A recent study (Schach and Starfield, 1973) presents evidence for the usefulness of difficulty in eating and sleeping and unusual irritability, particularly for younger children, but this refinement is mainly suited to description of the effects of acute illness.

Comfort and satisfaction have been accepted as proper outcome measures (White, 1967) but have not as yet been widely applied. Rosser and Watts (1972) devised and tested an instrument to measure distress that takes into account pain, mental disturbance, and reaction to disability. Satisfaction (in the sense of outcome) refers to satisfaction with one's state of health rather than to satisfaction with one's encounters with the health care system (which is more properly considered a measure of the acceptability of the process of medical care) (Starfield, 1973).

Disease is the classic *morbidity* category. In the proposed approach, the disease category incorporates both occult illness and manifest illness as well as an estimate of prognosis. "Disease" ranges from "not detectable" (not present by any available method of diagnosis) to "permanent disease," but intervening points along the continuum include "asymptomatic" (but detectable) disease, and temporary illness. This latter category includes conditions generally considered to be self-limited or curable with medical intervention.

The last two categories incorporate positive aspects of health (Sigerist, 1941). Achievement signifies the *level of development* or *accomplishment* and the potential for further development measured in comparison with a reference group. A variety of achievement tests (for example, see Escalona and Heider, 1959) can be used to ascertain attainment in cognitive fields; development in the motor and perceptual area might be measured by tests such as the Denver Developmental Screening Test for young children (Frankenburg et al., 1971).

Resilience is a category which deals with the *ability to cope* with adversity (Lazarus, 1966) and therefore has prognostic and preventive connotations. It is the category which measures the potential for resisting a range of possible threats to health (Jenkins, 1972). Ability to respond to stress may be measured by psychological techniques (Korsch et al., 1973), by physiological techniques

(Selye, 1973), or by evidence that certain defenses known to increase resistance are present or have been provided. Immunization is an example of a preventive procedure which influences vulnerability to illness. The measurement of outcome of immunizations is simplified by the knowledge that most immunization procedures have been well validated by clinical trials. Therefore, measurement of the process of immunization may suffice as a proxy measure of outcome of the immunization. Preventive procedures such as "anticipatory guidance" and "health education" have not been well validated, and new outcome measures which are thought to reflect their effectiveness will have to be developed.

Screening procedures which have been well validated could serve as appropriate measurement tools for appropriate categories. The results of such procedures could place individuals (or populations) at appropriate points on at least two of the categories: the *disease* continuum for actual morbidity which is discovered, and the *resilience* continuum for risk factors which are identified as a result of the screening.

Development of the Outcome Profile

Most proposals to describe health status involve the reduction of the several dimensions of health to a single scalar measure (International Journal of Epidemiology, 1972, 1973). In contrast, the proposed model involves the development of a profile similar in type to the Minnesota Multiphasic Personality Inventory (Dahlstrom and Welsh, 1960). The first step would be to develop indicators for various positions along each outcome continuum. Existing inventories (for example, Katz et al., 1972; Schach and Starfield, 1973; International Journal of Epidemiology, 1972, 1973; Barnoon and Wolfe, 1972; Williamson, 1970; Bickner, 1970; Perrin et al., 1972; Sultz et al., 1972; Patrick et al., 1973) could be incorporated or modified as necessary, but others would undoubtedly have to be developed, particularly for some of the categories. In the MMPI, each indicator statement is dichotomized (true–false); a similar approach could be employed for the proposed scheme. The development of a valid system of scoring for these scales would probably require a collaborative effort. "Normal" as well as "abnormal" populations would have to be tested. Profiles which provide pictorial representation of performance for

each of the various categories can be drawn using the calculated scores for each category; the profile can also be coded for processing of the data. Codes are assigned to each of the scales, and are arranged according to the scores on each of the scales. Symbols are employed to facilitate rapid comparison of a profile with standard profiles or other comparison profiles. This procedure reduces the large amount of data to a workable size while retaining enough information to accomplish the goal of describing health status.

Changes in the profile over time would indicate changes in health status; with appropriate study design, specific health care interventions could be evaluated. These interventions might involve the introduction of specific changes in ongoing health care programs (which might involve alterations of the structural aspects such as personnel, organization and financing, aspects of medical care practice such as screening, diagnosis, management, or reassessment, or phenomena related to patient behavior such as utilization, acceptance, understanding, and compliance). Alternatively, the intervention might be the establishment of a new program of health care for which effectiveness needs to be substantiated. It will be increasingly necessary to demonstrate effectiveness of existing and proposed modes of health care services, and these will undoubtedly involve consideration of their influence on health status. Distinction should be made, however, between studies which merely describe the influence of these changes on health status, and those whose design permits exploration of the reasons for the changes. For this reason, it is important that study designs incorporate examination of the way in which structural changes influence health status. Donabedian (1969:4) stressed the advantage to be gained when structure, process, and outcomes are examined simultaneously. Structural changes influence health care only by changing the way in which individual practitioners deal with health problems, and by altering the way in which individuals recognize their own needs and take appropriate steps to deal with them (Starfield, 1973). In the long run, therefore, the greatest understanding will be gained when health care interventions are evaluated not only for what they accomplish, but for the reasons for accomplishment as well (Starfield, 1973).

If the model were to be adopted, the next step would be to identify all existing methods of measuring these various types of

outcome. For this, an interdisciplinary team would be required. For some of the categories, the combination of existing inventories would suffice to identify a series of positions along the continuum. For other categories, the development of new measures would have to be initiated.

Mapping of points for all of the categories would undoubtedly take several years, but application of the model could proceed using those categories which have been completed. New ones would be added as they became available.

Realization of the need for appropriate measures of outcome is gaining momentum. Within the last few years, many investigators have been involved in efforts to develop both conceptual schemes and the necessary tools for gathering the data. The lack of agreement on relevant and appropriate dimensions for considering health status is evident and the field is clearly in an early stage of development. Although the attempt to coordinate current efforts will be difficult, and the time required to accomplish it will be great, it is believed that an approach such as the one suggested in this paper would focus attention and facilitate a consensus on the various areas which need to be included. At the same time, it would allow individual investigators from a variety of disciplines to concentrate on those areas which are of particular interest to them. This is of critical importance, as professionals from diverse backgrounds must be involved in the development of methods to evaluate the outcome of health care.

Barbara Starfield, M.D.
The Johns Hopkins Hospital
601 N. Broadway
Baltimore, Maryland 21205

The author is indebted to Dr. Barbara M. Korsch for suggesting and stressing the importance of some of the dimensions, and to Dr. I. Barry Pless for providing a forum in which the ideas could be presented and discussed.

This paper was prepared for The Workshop on Measurement of Outcomes, The University of Rochester School of Medicine and Dentistry, November 26–27, 1973, and was supported by a research grant (R01 HS 00110) from the National Center for Health Services Research and Development, U.S. Department of Health, Education, and Welfare. Dr. Starfield is the recipient of a Career Scientist Development Award (K02 HS 46225) from the Center.

References

Barnoon, S., and A. Wolfe
1972 Measuring the Effectiveness of Medical Decisions: An Operations Research Approach. Springfield: Charles C. Thomas.

Barro, A.
1973 "Survey and evaluation of approaches to physician performance measurement." Journal of Medical Education 48: 1051–1055.

Bickner, R.
1970 "Measurement and indices of health." Pp. 75–101 in Outcomes Conference I–II: Methodology of Identifying and Evaluating Outcomes of Health Service Programs, Systems, and Subsystems. Rockville, Maryland: U.S. Department of Health, Education, and Welfare, National Center for Health Services Research and Development.

Commission on Chronic Illness in the United States
1957 Chronic Illness in a Large City. Cambridge: Harvard University Press.

Dahlstrom, W., and G. Welsh
1960 An MMPI Handbook. Minneapolis: University of Minnesota Press.

Donabedian, A.
1966 "Evaluating the quality of medical care." Milbank Memorial Fund Quarterly 44 (3, Part 2): 166–206.
1968 "Promoting quality through evaluating the process of patient care." Medical Care 6: 181–202.
1969 "A guide to medical care administration vol. II: medical care appraisal—quality and utilization." The American Public Health Association.

Escalona, S., and G. Heider
1959 Prediction and Outcome. A Study in Child Development. New York: Basic Books, Inc.

Frankenburg, W., A. Goldstein, and B. Camp
1971 "The revised Denver Developmental screening test: its accuracy as a screening instrument." Journal of Pediatrics 79:988.

Grogono, A.
1973 "Measurement of ill health: a comment." International Journal of Epidemiology 2: 5–6.

International Journal of Epidemiology
1972 "Symposium on the measurement of ill health."
Hetzel, B.S. "The implications of health indicators: a comment." 1: 315–318.
Fanshel, S. "Meaningful measure of health for epidemiology." 1: 319–337.
Maddox, G. L. "Interventions and outcomes: notes on designing and implementing an experiment in health care." 1:339–345.
Breslow, L. "A quantitative approach to the World Health Organization definition of health: physical, mental and social well-being." 1: 347–355.
Lamm, G. "Problems in the definition of ill health." 1: 357–359.

1973 "Symposium on the measurement of ill health."
Chiang, C. L., and R. D. Cohen. "How to measure health: a stochastic model for an index of health." 2: 7–13.
Himatsingani, C. "Approaches to health and personal social services planning in the national health service and the place of health indices." 2: 15–21.
Pole, J. D. "The use of outcome measures in health service planning." 2: 23–30.
Ashford, J. R., K. L. G. Read, and V. C. Riley. "An analysis of variations in perinatal mortality amongst local authorities in England and Wales." 2: 31–46.

Jenkins, C. D.
1972 "Psychology in epidemiology: the growing edge." Chapter 4 in G. Stewart, Trends in Epidemiology. Springfield: Charles C. Thomas.

Katz, S., A. Ford, J. Downs, M. Adams, and D. Rusby
1972 "Effects of Continued Care: A Study of Chronic Illness in the Home." U.S. Department of Health, Education, and Welfare Publication No. (HSM) 73–3010 (December).

Korsch, B., V. Negrete, J. Gardner, C. Weinstock, A. Mercer, C. Grushkin, and R. Fine
1973 "Kidney transplantation in children: psychosocial follow-up study on child and family." Journal of Pediatrics 83: 399.

Lazarus, R. S.
1966 Psychological Stress and the Coping Process. New York: McGraw-Hill.

Patrick, D. L., J. W. Bush, and M. M. Chen
1973 "Toward an operational definition of health." Journal of Health and Social Behavior 14: 6–23.

Perrin, J., E. Rusch, J. Pray, G. Wright, and G. Bartlett
1972 "Evaluation of a ten-year experience in a comprehensive care program for handicapped children." Pediatrics 50: 793–800.

Rosser, R., and V. Watts
1972 "The measurement of hospital output." International Journal of Epidemiology 1: 361–368.

Sanazaro, P., and J. Williamson
1968 "End-results of patient care: a provisional classification based on reports by internists." Medical Care 6: 123–130.

Schach, E., and B. Starfield
1973 "Acute disability in childhood: examination of agreement between various measures." Medical Care 11: 297–309.

Shapiro, S.
1967 "End-result measurements of quality of medical care." Milbank Memorial Fund Quarterly 45 (2, part 1): 7–30.

Sigerist, H.
1941 Medicine and Human Welfare. New Haven: Yale University Press.

Starfield, B.
1973 "Health services research: a working model." New England Journal of Medicine 289:132.

Sullivan, D.
1971 Disability Components for an Index of Health. National Center for Health Statistics, U.S. Department of Health, Education, and Welfare, Public Health Service, 1000 Series 2, No. 42. Washington, D.C.: U.S. Government Printing Office.

Sultz, H., E. Schlesinger, E. Mosher, and J. Feldman
1972 Long-Term Childhood Illness. Pittsburgh: University of Pittsburgh Press, Chapter 5.

Webster
1969 Webster's Seventh New Collegiate Dictionary. Springfield, Massachusetts: G & C Merriam Company.

White, K. L.
1967 "Improved medical care statistics and the health services system." Public Health Reports 82: 847–854.

Williamson, J.
1970 "Outcomes of health care: key to health improvement." Pp. 133–149 in Outcomes Conference I–II: Methodology of Identifying and Evaluating Outcomes of Health Service Programs, Systems, and Subsystems. Rockville, Maryland: U.S. Department of Health, Education, and Welfare, National Center for Health Services Research and Development.