

---

## MEASURING AND EVALUATING HOSPITAL AND MEDICAL CARE\*

AVEDIS DONABEDIAN, M.D.

Professor of Medical Care Organization  
School of Public Health  
The University of Michigan  
Ann Arbor, Mich.

I have been asked to speak to you briefly about what I consider “the most serious problem in evaluating hospital and medical care, primarily focusing on inpatient treatment.” This means that I must take a negative view, emphasizing how far short we are of our goal rather than extolling the considerable progress we may have made. Further, since the entire proceedings of this conference are within the context of professional responsibility for the quality of health care, I shall view the subject primarily from the standpoint of policy and action rather than of research.

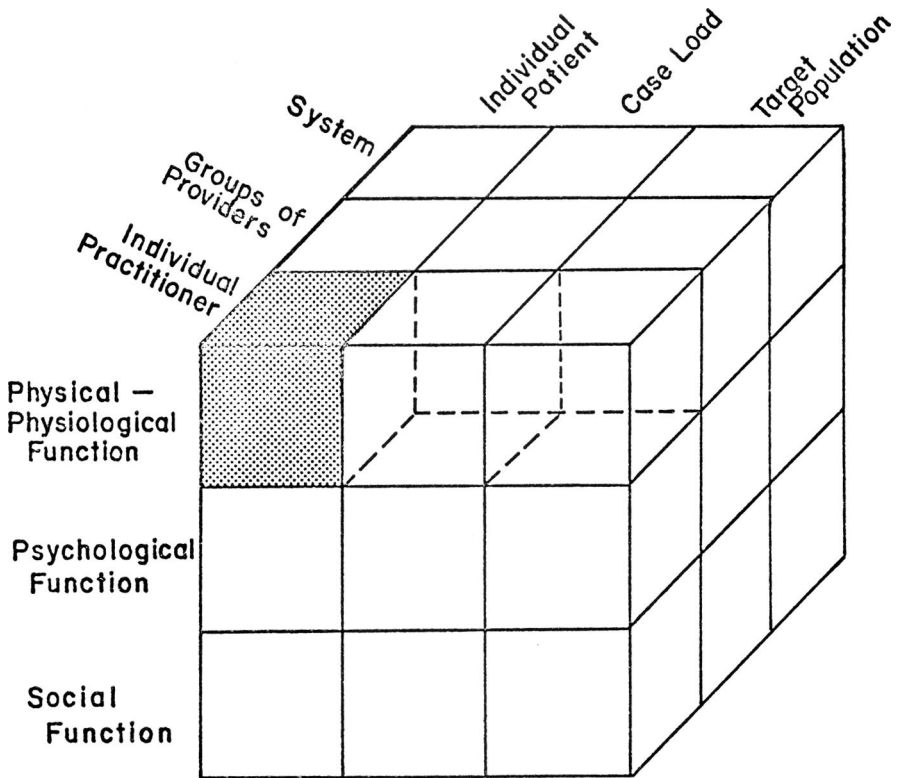
At the outset, please note that my assignment is to speak about “the most serious *problem*”—in the singular, not the plural. “Surely,” I thought when I first read this curious formulation, “this must be an error in typing. The problems are legion, and all are important; how could I single out one as the most serious?” And yet, strangely enough, there is one such problem that is the very root of all of our other difficulties, and it is this: we attempt to pass judgment on what we do not understand, or understand very imperfectly, at best. Everything I shall say from here on will be a variation on this central theme.

Ask an individual to talk about quality health care and you are likely to get a catalogue of platitudes. Ask two people and you will probably get an argument. Ask three, and you will probably have chaos. The reason is that each, like the proverbial blind men, has some different portion of the elephant in his hand.

Let me illustrate the problem by first showing you a visual approximation of the nature of the beast (see accompanying figure). It does not

---

\*Presented in a panel, A Discussion of Methods, as part of the 1975 Annual Health Conference of the New York Academy of Medicine, *The Professional Responsibility for the Quality of Care*, held April 24 and 25, 1975.



Schematic of frame of reference in simplified form.

show great understanding that I have reduced the “elephant” to a cube. At least it is a solid; it is a beginning!

Note that one side of the cube shows several definitions of health, or extensions of a single definition; namely, physical-physiological, psychological, and social function. These are merely illustrative and are meant to show that our concept of quality—and, consequently, our criteria, our standards, and our measures—depend on how we define health and our responsibility for it.

On another side of the cube are illustrative steps indicating larger and more complex aggregations of the instrumentalities that provide care: the individual practitioner, groups of providers, and the system of care. This suggests that beyond the most elementary level it becomes necessary to introduce into the criteria and measures of quality

the joint contributions of more than one profession, as well as the attributes of continuity and coordination.

On the third side of the cube are steps in the aggregation of the client: individual patient, case load, and target population. The figure implies two separate notions: the distinction between patient and person and the distinction between an individual and an aggregate. Why should we make these distinctions? Because if our concern is with people as well as patients the concept of quality must include the attribute of access to care. Similarly, if our concern is for the aggregate as well as for the individual, our concept of quality must be broadened to include considerations of optimum allocation, given our necessarily limited resources. The allocation of resources is not simply a problem for those who make policy at the community level; it confronts every practitioner every moment of every day. Each practitioner is responsible for a case load of patients or a population of clients, and he must decide how to use his most precious resources—the practitioner's own time, attention, and concern—so as to serve best not only each patient singly, but all those who look to the practitioner for care.

Some additional concepts are illustrated by this cubic figure. First, it is likely that these three aspects do not exhaust the dimensions of quality. The reality is much more complex. Second, it becomes clear that almost all of our concern for quality has, so far, focused on only a small part of the figure: that representing the management of physical-physiological pathology in individual patients by individual practitioners, usually physicians. The immediate reaction to this observation is that we must expand the scope of our concern. But this cannot be done unilaterally or capriciously; some fundamental questions of professional domain and professional responsibility are at issue. In the most general terms, we cannot hold health practitioners responsible for that over which they have no control and we may not wish to expand their domain of control simply to achieve more conceptually satisfying definitions or estimates of the quality of care. However, it is not unreasonable to expect a practitioner to concern himself with the social and psychological concomitants of illness or to judge him by the manner in which he manages his entire case load of patients. Similarly, it is not unreasonable to expect a hospital to add to its concerns that of the aggregate effects of all the services that it offers. Finally, when a program, institution, or individual practitioner is responsible for an

enrolled population, the care of that entire population rather than of the subset of patients it generates must become the basis for judgment.

All of these extensions in the scope of our concern for quality are not unreasonable; yet, consider what horrendous difficulties they must create. We who are unable to decide how best to assess the quality of the technical management of physical-physiological illness shall be required, first, to determine criteria and standards for social and psychological management and, second, to resolve the ethical dilemmas of reconciling the interests of the individual patient with those of the collectivity—the heart of resource allocation. But we have no choice. We must venture forth.

From this prospect, which some will consider a challenge and others an unwarranted imposition, I shall turn to the more familiar realm of the technical management of physical-physiological illness.

The Health Insurance Plan of Greater New York pioneered in the study of the quality of medical care. In a paper which describes some of its work,<sup>1</sup> Henry Makover summarizes its principles by quoting from an editorial which appeared in the *Lancet*.<sup>2</sup>

The other day an experienced physician was asked what criteria he would apply in judging the efficiency of a hospital; what relative importance he would attach to the qualifications of the staff, the ratio of beds to nurses, the adequacy of special departments, the catering, the facilities for reablement, and the various other items on which inspecting authorities commonly make notes. He replied, "I should not inquire into any of these things. I should simply go into the wards, select six patients, and find out precisely what had been done for them, and the care that they had received, since the day of their admission." This wise answer has implications beyond even the hospital services, for it embodies the truth that any kind of machinery, however ingenious, is but a means to an end.

Subsequent work at the Columbia University School of Public Health and Administrative Medicine under the leadership of Dr. Mildred Morehead (with whom I am honored to share this podium) was in this same tradition, except that almost exclusive reliance had to be placed on the medical record with all its recognized limitations as a source of information.<sup>3, 4</sup> This deficiency—which could be rectified only in part by interviewing the attending physicians—was counter-

balanced by the soundness of the basic idea of this study: that the judgment of quality in this context must depend on a complete assessment of the total care of a patient who presents as a problem for diagnosis and management, and that this assessment must reflect the best judgment of a mature and skilled physician charged with reviewing each case. Unfortunately, in hands less skillful than Dr. Morehead's the reliability of such judgments was found to be so low that the entire procedure has fallen into disrepute.

An alternative method that has been proposed originated in the pioneering work of Paul Lembcke. He devised what he called a "scientific method," which purported to remove the subjective element in assessments of quality by explicitly specifying the criteria by which judgments were to be made.<sup>5</sup> Subsequently, a similar approach was re-discovered independently by Beverly Payne,<sup>6</sup> it was tested in its most developed form in a study of hospital and ambulatory care in Hawaii.<sup>7</sup> The "criteria approach," as this method has come to be called, has since received such wide recognition that it has earned a central position in the projected nationwide system of Professional Standards Review Organizations. The sets of criteria that this system will generate presumably will stand as concrete embodiments, if not of quality care, at least of acceptable care. It is this concreteness and specificity that render the criteria approach so attractive. It adapts readily to the computer and the management of mass data. The judgments which it yields are likely to be highly replicable. What is at issue is their validity.

In assessing the validity of this approach we recognize that we are dealing with a generality. No doubt, the lists of criteria do vary in structure and content and may be put to different uses. For example, the method originally described by Lembcke was applied only to specified surgical operations and used only to judge whether the operation was to be considered justified or criticized, based on the evidence at hand. Even then, the judgment was further tempered by accepting the performance of teaching hospitals as the standard. By analogy, one might expect that the criteria would yield reasonably valid judgments as to the necessity for admission and readiness for discharge, although this has not been proved. But do the criteria, as currently designed, adequately represent the quality of care as more richly conceived? I shall argue that they do not and most probably cannot do so. Some of the shortcomings I shall mention can be remedied by a better de-

signed program, but others are so basic that they call for a totally different approach.

All methods of assessment that begin with a sample of hospital records selected according to the primary discharge diagnosis or operative procedure share a number of defects. First, the sample excludes all cases that should have been so diagnosed or operated on in that manner, but were not. Thus a whole slice of performance remains totally in the dark. Second, there is considerable doubt that the performance of a physician or institution is so homogenous that a selection of diagnostic categories which is not a probability sample can stand for the total case load of that physician or institution.<sup>8</sup> Recent evidence based on a performance index using the criteria approach has shown generally low correlations in the performance of the same physicians across diagnostic categories.<sup>9</sup> Finally, in all instances the validity of these judgments is contingent upon the completeness and accuracy—if not the truthfulness—of the material in the record. While it is true that hospital practice and recording are correlated, the level of correlation appears to be low.<sup>10, 11</sup> Even if this correlation were high, the hospital record ordinarily would have little or no information about two important segments of care: those prior and subsequent to hospitalization.

Some additional deficiencies are particularly relevant to the criteria approach as ordinarily designed and implemented. To begin with, no account is taken of the presence of diagnoses additional to the primary one which may influence management—unless we assume that in large samples the presence and influence of such additional diagnoses tends to become comparable. Similarly, no account is taken of redundancy and wastefulness in management, whether in diagnostic investigation or treatment. The emphasis on justification of admission and readiness for discharge is an exception to this general observation. However, even here some important deficiencies may intrude. For example, if the criterion is readiness for discharge without attention to average length of stay, one could lose sight of how long it takes to become fit to leave the hospital and why. In considering the appropriateness of admission, it is unusual to ask some questions that should be fundamental to an assessment of quality: Could the disease have been prevented in the first place? Has hospitalization become necessary in part because of prior mismanagement?<sup>12</sup>

Closer to the assessment of hospital care itself, it is very unusual for

those using the criteria approach to pass judgment on the accuracy of the diagnosis for discharge, the need for surgical intervention, or the choice of a particular operation. These are generally taken as given and attention is focused on whether the record of prior management includes the minimum set of activities that corresponds to these diagnoses or interventions—no matter what else has been done, how circuitous the path, or how proper the final destination. Strictly speaking, this is a test of internal consistency, an attribute which is at least one step removed from the quality of performance.

Assuming the diagnosis for discharge to be accurate and the minimum set of criteria to be relevant, further difficulties arise in constructing an over-all measure of performance which requires a summation of component subparts. It is usual to assume equal weights for all the items. Even when differential weights are assigned, as in the most recent work of Payne, it is not clear on what basis this is done.<sup>7</sup> It is seldom, if ever, recognized that the process of management consists of highly interactive parts and that when certain key elements are lacking the care as a whole must be judged as poor no matter how many of the other criteria elements have been performed and recorded.

This is the core of the problem. Clinical management is a complex process of making diagnostic and therapeutic decisions. It is based on factual knowledge, but requires something additional which we recognize as clinical judgment—which is at the heart of technical quality. Clinical judgment involves deciding what alternatives to consider in diagnosis and therapy, how far to go in seeking what degree of certainty, what means to use, what risks to incur relative to what probability of success in seeking how large a benefit, when to act, and when to watchfully wait. What a far cry from the stereotyped behavior embodied in the lists of criteria—which so easily can become distorted into an indiscriminate assemblage of the elements of care until they are almost a caricature of clinical judgment rather than its true representation. No wonder that physicians are almost intuitively repelled by the very thought of being judged by these.

What, then, is to be done? Shall we, because clinical judgment is so complex and elusive to measure, give up all thought of assessing the quality of care? This certainly is not my conclusion. We need to further develop and refine the criteria so that they become a realistic model of the clinical process rather than a distortion. But we cannot do this

unless we gain a greater understanding of the clinical process itself through more rigorous conceptual and empirical work. I repeat, we cannot evaluate in a defensible manner what we do not clearly understand, and we now only dimly perceive the attributes of clinical judgment. We urgently need basic research on the clinical process itself, so that we can delineate the different pathways it can take and the costs and benefits of each—both monetary and nonmonetary—to the individual and to society. In the pursuit of this end we may learn from what is already known about problem-solving and trouble-shooting behavior in general<sup>13</sup> and about clinical decision-making in particular.<sup>14</sup>

But what are we to do as we await the advent of the more perfect instrument? First, it is salutary to remember that many deficiencies in care are so gross that no refinement is necessary to detect them: they fairly leap at you, provided you care to look. Second, the criteria approach can be modified to meet many of the criticisms that we have leveled against it. In particular, the use of these criteria must be judicious and prudent, with full cognizance of their many limitations as arbiters of quality. It would be tragic to yield to the temptation to use them primarily as devices to police and punish. Third, any system of quality assessment and assurance must include information about all the important aspects of care obtained from a variety of sources. Among the most important of these, as the Committee on the Standardization of Hospitals recommended more than 60 years ago,<sup>15</sup> is information about proximate and remote end results of care. Finally, we must remember that until a more objective method is devised the ultimate arbiter of the quality of clinical management must remain the judgment of our wisest and most competent colleagues when they are given all the facts of each case. We shall be unable to improve upon this until the bases for judgment are made explicit and measurable and are subjected to wider professional and social validation.

#### NOTES AND REFERENCES

1. Makover, H. B.: The quality of medical care: Methodology of survey of medical groups associated with the Health Insurance Plan of Greater New York. *Amer. J. Public Health* 41:824-32, 1951 (see esp. pp. 825 and 826).
2. Mechanism and purpose. Editorial. *Lancet* 1:27-28, 1950.
3. Ehrlich, J., Morehead, M. A., and Trussell, R. E.: *The Quantity, Quality and Costs of Medical and Hospital Care Secured by a Sample of Teamster Families in the New York Area*. Monograph. New York, Columbia University School of Public Health and Administrative Medicine, 1962.



4. Morehead, M. A., Donaldson, R. S., Sanderson, S., and Burt, F.: *A Study of the Quality of Hospital Care Secured by a Sample of Teamster Family Members in New York City*. Monograph. New York, Columbia University School of Public Health and Administrative Medicine, 1964.
5. Lembecke, P. A.: Medical auditing by scientific methods. *J.A.M.A.* 162:646-55, 1956.
6. Payne, B. C.: Continued evolution of a system of medical care appraisal. *J.A.M.A.* 20:536-40, 1967.
7. Payne, B. C. and Lyons, T. F.: *Method of Evaluating and Improving Personal Medical Care Quality: Episode of Illness Study*. Ann Arbor, Mich., The University of Michigan School of Medicine, 1972.
8. Donabedian, A.: *A Guide to Medical Care Administration, Volume II: Medical Care Appraisal—Quality and Utilization*. New York, Amer. Public Health Ass. Inc., 1969, pp. 53-56.
9. Lyons, T. F. and Payne, B. C.: Inter-diagnosis relationships of physician performance measures. *Med. Care* 12:369-74, 1974.
10. Rosenfeld, L. S.: Quality of medical care in hospitals. *Amer. J. Public Health* 47:856-65, 1957.
11. Lyons, T. F. and Payne, B. C.: The relationship of physicians' medical recording performance to their medical care performance. *Med. Care* 12:714-20, 1974.
12. These aspects of performance were pointed out to me by Paul M. Gertman, Director, Health Services Research and Development Program, Boston University Medical Center, Boston, Mass.
13. Donabedian, A.: Evaluating the Quality of Medical Care. *Milbank Mem. Fund Quart.* 44:166-203, 1966, Part 2 (see esp. pp. 194-195 and references 63-70).
14. Barnoon, S. and Wolfe, H.: *Measuring the Effectiveness of Medical Decisions: An Operations Research Approach*. Springfield, Ill., Thomas, 1972.
15. Committee on Standardization of Hospitals: End result record system suggested by the Committee on Standardization of Hospitals. *Surg. Gynec. Obstet.* (Suppl.) 18:9-12, 1914.