

TWENTY YEARS OF RESEARCH ON THE QUALITY OF MEDICAL CARE

1964-1984

AVEDIS DONABEDIAN

*School of Public Health,
University of Michigan*

Since 1964 we have built on earlier work, with some refinements, particularly in the formulation of explicit criteria of process and outcome; the detailed standardization of case mix when outcomes signify quality; the prespecification of outcomes for follow-up, when adverse outcomes are only the occasion for later assessment of process; a greater emphasis on more subtle organizational characteristics in the study of structure; and the identification of the separate effects of structural attributes by multivariate analysis. We have also paid more systematic attention to questions of measurement, including the veracity and completeness of the record; the procedures of criteria formulation; and the reliability, validity, and screening efficiency of the criteria. A notable advance is the use of decision analysis to identify optimal strategies of care, including the introduction of patient preferences and monetary cost in the specification of such strategies, and the use of decisional algorithms to portray the criteria of quality.

AUTHOR'S NOTE: This article is based on a presentation for "Twenty Years of Research, Two Days of Celebration," Kaiser Permanente Health Services Research Center, Portland, Oregon, September 29-30, 1984. It is one product of work supported by the Commonwealth Fund, the W.K. Kellogg Foundation, and the National Center for Health Services Research. The opinions expressed in it are mine alone.

When I was first asked to consider this subject, my immediate reaction was that the last 20 years were a time of momentous advances in quality assessment. But a rapid scanning of several bibliographies led me to an additional discovery. I must now conclude that (1) by 1964 the foundations of almost all the major approaches to quality assessment had been laid down; (2) between 1964 and 1984, despite the astounding proliferation of quality studies, we achieved mainly refinements in what we already knew how to do; and (3) regrettably, some of the more fundamental questions pertinent to quality assessment are still not asked, or are misspecified, let alone systematically addressed.

DEVELOPMENTS PRIOR TO 1964

Let me give some examples of accomplishments prior to 1964, using date of publication as the criterion of dating.

STUDIES OF "OUTCOMES"

The stream of studies that use health status as a means of measuring quality can begin, most properly, with the work of Ernest Amory Codman, a great, yet tragic, figure. By 1916 Codman fully understood the place of what he called the "end result" of care in the evaluation of clinical practice, not only as a professional activity, but also as an organizational, administrative, and economic enterprise; and he had proposed and tested a method for monitoring "end results." His published work, though rather confused and flawed by excessive dogmatism, is arresting for its creativity and insightfulness, and also deeply moving for the clear evidence we see in it of both his courage and personal anguish. He is a man whose memory we should forever cherish.

In the same tradition was the later study on preventable maternal mortality conducted under the aegis of the New York Academy of Medicine. The monograph (published in

1933) in which Hooker described this study is so elegant in style, searching in its analysis, and uncompromising in its commitment to human welfare, that it must be recognized as one of the treasures of a literature not overly rich in such accomplishments.

Quite a bit later—1955, to be precise—we find a parallel piece of work sponsored by the same organization, but this time on preventable perinatal deaths (Kohl, 1955). We also see, at about this time, notable examples of comparative mortality in the work of Shapiro and his associates (1958, 1960) at the Health Insurance Plan of Greater New York (HIP), and later of Lipworth et al. (1963) in Britain, the former concerned with prematurity and perinatal mortality, and the latter with case fatality in teaching and nonteaching hospitals.

STUDIES OF “PROCESS”

Let us now visit the second major stream of quality studies, that which includes the assessments of the process of care.

Here again, we encounter an early, awe-inspiring landmark in the monumental work of Lee and Jones (1933) on the “Fundamentals of Good Medical Care.” This work offers a concept of quality rarely equaled in profundity or comprehensiveness, an uncompromising declaration of socially responsible professional norms as the standards of assessment, and an explicit enunciation of these standards. By a remarkable coincidence, this work was published in 1933, the same year the work on preventable maternal mortality appeared. Equally interesting, the next major advances occurred in the 1950s, but with some separation into several recognizably different offshoots, all similar in using professional norms as a foundation for assessment, but differing in the degree to which the criteria are specified in advance, the method of obtaining information on practice, the locus of care subject to assessment, and the extent of interest in the relationship between structural attributes and the process of care.

“Implicit” criteria. Let me first follow the offshoot made up of the studies that adopted implicit criteria, with varying degrees of superimposed guidance, to assess the quality of care. We find, in this group, the landmark study of general practice in rural North Carolina by Peterson et al. (1956): an assessment of office practice, based on direct observation of the practitioner-client transaction, by a trained visitor who used his own implicit criteria to arrive at a judgment of the quality of care.

At about the same time, a series of studies unfolded at the Health Insurance Plan of Greater New York, led at first by Makover and then by Morehead, assessing the quality of care, mainly based on information obtained by a review of records and conversations with physicians, and basing the judgments of quality primarily on the implicit criteria of expert clinicians (Makover, 1951; Daily and Morehead, 1956; Morehead, 1958, 1967). The apotheosis of the implicit criteria approach, however, is to be found in the assessment of the quality of hospital care received by members of the Teamsters Union and their families, as reported by Morehead and her associates (1964).

It is important to note that in all these studies there were explorations of the relationships between performance and structural characteristics that set the pattern for much that was to come. For example, in their study of hospital care, Morehead et al. (1964) paid much attention to the organizational characteristics of hospitals (such as ownership and medical school affiliation), the professional characteristics of physicians (such as specialty certification and hospital privileges), and the interaction of the two sets of characteristics. Peterson et al. (1956), who measured the quality of care only because they wished to find out what attributes and experiences were conducive to good general practice, included in their exhaustive inquiry factors such as the type of medical school, performance as a medical student, the nature and duration of postgraduate training, continuing education through refresher

courses and subscription to medical journals, medical society membership, hospital affiliation, practice alone or in a group, access to laboratory services, the presence of specific pieces of technical equipment in the doctor's office, seeing patients by appointment, and duration of the workday. Attributes such as education, training, hospital affiliation, and seeing patients by appointment were also included in the studies conducted at HIP, but there were added questions of special interest, such as length of association with the Plan, percentage of time devoted to HIP enrollees, and percentage of work done at a group-practice facility.

"Explicit" criteria. Let us now follow developments along another branch, that on which cluster the studies using explicit criteria. The premise that the norms of practice must be expressed in explicit criteria was the rock on which Lembcke (1956) built what he hoped to be a "science" of assessment. We find in his work the fundamental precepts that have guided the formulation of such criteria ever since.

A little later, quite independently, a group of workers at the University of Michigan—a group that included Fitzpatrick, Riedel, and Payne (1962)—adopted the use of explicit criteria to assess what they called "hospital effectiveness."

There are a few other offshoots that, although smaller, foreshadow later developments. We find, for example, in another work by Lembcke (1952), an epidemiological approach to quality assessment based on comparisons among hospital service areas. This is the prototype of, and a fully equal companion to, the spate of small-area studies that we have recently experienced. Similarly, the statistical indexes of outpatient care developed by Ciocco et al. (1950) and the indexes of hospital activities by Eisele et al. (1956) are the prototypes of what later came to be recognized as "professional activities studies" and "profiles." In fact, Lembcke (1967) has traced the origins of the statistical indexes of hospital performance even further back,

first to measures used by the staff of the Commonwealth Fund to assess the members of its rural hospital program (Southmayd and Smith, 1944) and then to Lembcke's own elaboration of these measures in assessing the performance of participants in the Council of Rochester Regional Hospitals, also under the aegis of the Commonwealth Fund (Lembcke, 1947: 27-29).

STUDIES OF "STRUCTURE"

I have one final major stream I must mention, even if briefly. It is the stream of studies that are concerned with the assessment of structural attributes.

Here, again, the foundations are rather old. For example, we find in the work of Goldmann and Graham (1954) a detailed model for assessing an ambulatory care program by means of a variety of structural characteristics. The work of Georgopoulos and Mann (1962) on "The Community General Hospital" also introduced a new set of more subtle organizational attributes that either appear to influence, or can, with good reason be expected to influence, the quality of care.

USE AND ABUSE OF THE PAST

I ask you to forgive me for this overly lengthy retrospective. Perhaps it is merely a failing of old age that prompts me to be so backward looking. But it seems to me that no field of intellectual endeavor can afford to remain as unmindful of its roots and indifferent to its indebtedness to its past as ours appears to be. Without a more widely shared knowledge of our antecedents, some of us remain condemned to reinventing what we already know, often in less successful forms; some less scrupulous others are tempted to plunder the past by failing to reveal their sources. We should, instead, do homage to our past by building upon it.

MORE RECENT DEVELOPMENTS

LINES OF CONTINUITY

Let me now survey the work published during 1964 and since. To begin with, I shall mention briefly some lines of continuity that one observes when old methods are used, with variable degrees of improvement, to obtain new information in new settings.

Perhaps the most notably derivative of all recent efforts was the near duplication of the Lee-Jones study (1933) by a group of investigators at Yale, led first by I. S. Falk (Falk et al., 1967) and then by Schonfeld (Schonfeld et al., 1975). In this reincarnation, the original work was in many ways improved upon, but in some ways the new failed to measure up to its majestic predecessor.

The comparisons of small geographic areas with respect to a large variety of attributes has continued—without striking improvements in method—in the work of many investigators, including that of Lewis (1969) in Kansas, Wennberg and Gittlesohn (1973) in New England, Griffith and his colleagues (1981) in Michigan, and Stockwell and Vayda (1979) in Ontario. A succession of international comparisons could also be placed in this category. (See, for example, Bunker, 1970; Vayda, 1973; McPherson et al., 1982).

During the period under examination, Morehead and her associates used the method developed originally at HIP, with relatively small refinements, to study the quality of care in neighborhood health centers and some other ambulatory care settings (Morehead, 1970; Morehead et al., 1971, Morehead and Donaldson, 1974). Similarly, Payne extended the scope of his activities by using the method originally devised to study hospital effectiveness in Michigan to assess the quality of care in Hawaii—an admittedly more pleasant locale! There were some refinements of method as well. These included the joint consideration of in-hospital and out-of-hospital care in each

episode of hospitalization, the differential weighting of the criteria, attention to the reliability of abstracting records, and the development of explicit criteria for ambulatory care (Payne et al., 1976).

ADVANCES IN STUDIES OF "PROCESS"

The extension of explicit criteria to assessing the process of ambulatory care appears to usher in a new stage in the assessment of process in general, a stage characterized by much greater interest in some basic questions of method. I shall mention only some of these.

The medical record. Because the study of process depends so heavily on what appears in the medical records, there has been a fair amount of attention to the completeness and accuracy of the record, and to the implications of this to quality assessment, as well as clinical management. The work of Starfield and her associates at Johns Hopkins is a good example (Zuckerman et al., 1975; Starfield et al., 1979). At the same time, there has been a great deal of interest, notably represented by the work of Weed (1969, 1971), in redesigning the medical record to make it a better tool in clinical management. By stretching our category a little, we might include here the introduction and testing of computerized record systems as well. (See, for example, McDonald, 1976; Barnett et al., 1978; Barnett, 1984). I believe that in these kinds of developments there is a growing realization that the nature of recording influences what is possible in quality assessment, and that the requirements of quality assessment can and should influence the design of the record system, although a thorough, systematic exploration of this interrelationship is yet to come.

Criteria formulation. A more important development in studying the assessment of process, and also of outcome, has

been the exploration of criteria formulation itself. One sees here several strands that cannot be—indeed should not be—separated, as they are so intimately intertwined.

One strand represents the interest in criteria formulation as a consensus-achieving process; another comprises work on the choice of criteria and the design of criteria lists so as to achieve greater correspondence between assessments of process and of outcome; still another strand consists of explorations of the relation between criteria design and the ability of the criteria to screen cases so as to achieve the most accurate separation between acceptable and unacceptable care at lowest cost in effort and money. Briefly, these are studies of consensus or reliability, causal validity, and screening efficiency.

Consensus. The study of criteria formulation as a social process leading to consensus is represented by the deliberate application of techniques such as the Delphi method or the Nominal Group Process (Delbecq and Van de Ven, 1971; Dalkey et al., 1972). Many investigators borrowed elements from these and other methods of arriving at consensus, and have described the consequences with variable degrees of attention to reliability and validity.

Notable among the many studies of process that have looked into these matters are, in rough chronological order, the early work of Brook (1973) in Baltimore, Hare and Barnoon (1973) for the American Society of Internal Medicine in several regions of the United States, Osborne and Thompson (1975) under the leadership of the American Academy of Pediatrics, Hulka et al. (1979) in North Carolina, Riedel and Riedel (1979) in New Haven and Hartford, and Palmer et al. (1984) in Boston.

Among other things, these studies have described the changes that occur in the inclusion or exclusion of criteria when these are selected according to degree of “importance” based on perceived contributions to making the right diagnosis or achieving the best outcomes, and according to the information

called for by the criteria being subject to recording. The degree of agreement on the criteria has been studied among members of a group of physicians, as well as among groups of physicians differentiated by specialty or other attributes. There are only fragments of information about the stability of such choices over time (Wagner et al., 1976, 1978). There is also some information, again not enough, about the relation between the perceived importance of the criteria and the likelihood of finding that they have been complied with in practice. We also do not know on what bases the criteria that have been chosen might be differentially weighted, and whether the weighting materially alters the judgments on quality. Nor do we know whether one method of arriving at consensus is better than another, and in what ways (Donabedian, 1982).

Screening efficiency and validity. The more elaborate methods of achieving and testing consensus on the explicit criteria of process are, of course, only a means of obtaining criteria that are efficient and valid. To test validity it is necessary to examine the correspondence among judgments based on explicit criteria of process, implicit criteria of process, and either implicit or explicit criteria of outcome. If one of any pair of methods can be taken as more valid, the efficiency of the less valid method is judged by a comparison of the proportion of "positives" that it detects and the proportion of "negatives" that it mistakenly identifies as "positive." McClain and Riedel (1973) and Greenfield et al. (1982) offer examples of such comparisons in the screening of cases for utilization review and for appropriateness of admission, respectively. Judgments resulting from the use of explicit and implicit criteria have been compared, among others, by Brook (1973), Novick et al. (1976), Hulka et al. (1979), and Kane et al. (1981). Comparisons of judgments based on explicit criteria of process with those based on either implicit or explicit criteria of outcome include those reported by Brook (1973), Hulka et al. (1979), Mushlin and Appel (1980), and Mates and Sidel (1981).

Though much more remains to be done, the findings of these comparisons suggest, on the whole, coherence and rationality rather than—as some have claimed—failure and disarray. We have, moreover, yet another refinement in the criteria of process—one that, by more directly linking criteria formulation to decision analysis, promises to yield progressively more valid conceptions and measures of quality.

Decision analysis and the criteria of care. As early as 1964, the year I arbitrarily chose to usher in the new age, Peterson and Barsamian described an approach to quality assessment based on the then burgeoning science of decision analysis as applied to clinical practice (Peterson and Barsamian 1964, Peterson et al., 1966). This approach was soon forgotten, but only to reemerge later, based on different antecedents, and in different hands, as the “criteria mapping” approach described by Greenfield et al. (1975, 1981).

In their original forms, the decision maps can be seen as more true-to-life representations of the sequences by which physicians arrive at clinical decisions. But they are similar to the more traditional lists of criteria in being merely current embodiments of the conventional wisdom. They do, however, offer opportunities for more formal testing, based on subjective probabilities of the relationship between successive branchings of the stepwise progression that they represent. Besides defining more precisely which of several possible alternative strategies is to be preferred, this testing can lead to an assigning of differential penalties to different kinds of failures to pursue the optimal path. At the same time, the empirical testing of such decisional algorithms through epidemiological observations, or by means of more rigorous clinical trials, can be directly translated to revised criteria maps. If one goes on to include monetary cost as well as patient preferences as important criteria for judging the success of alternative strategies of care, one arrives at a fundamental reformulation of the meaning of quality itself.

The various currents and crosscurrents of research that have brought about this last development are too diverse, widely dispersed, and complex to permit a ready review. But illustrative stages may be observed in the work of a single group of investigators led by McNeil (McNeil et al., 1976). We see (1) how the choice of strategies depends on the specificity and sensitivity of each step in a stepwise progression of diagnostic tests; (2) an example of how the choice of a therapeutic regimen depends not only on the benefits and risks associated with alternative regimens when these are implemented, but also, as one would expect, on the propensity of patients to adhere to each of the contending regimens (McNeil and Adelstein, 1975); (3) the importance of monetary cost as a factor in determining how far successively more exhaustive testing is to continue, and which regimen of therapy is preferable (McNeil et al., 1975); (4) how different patients make different choices when offered a treatment with higher immediate mortality but greater subsequent longevity for the survivors, as compared to a lower immediate mortality attained at the price of shorter subsequent longevity (McNeil et al., 1978); and finally (5) we are also shown analogous differences in preferences when there is a prospect of a shorter life with less disability contrasted with a longer life, but with greater disability (McNeil et al., 1981).

It is easy to see even from these sketchy examples why these developments promise to reform clinical practice by redefining the meaning of quality and its standards, and why I believe that they are the single most important advance pertinent to quality assessment in recent years. Needless to say, the advance I refer to rests on an expansion of the meaning of outcomes and a reexamination of the relationship between outcomes and process. Thus, we are brought back to a category of methods that use outcomes to assess the quality of care.

ADVANCES IN STUDIES OF OUTCOME

You will have noticed that the earlier studies of outcome that I described are divisible into two classes. In the first of these

classes are studies in which favorable adverse outcomes, such as mortality and case fatality, serve to indicate—or at least suggest—a judgment on quality, in and of themselves.

The stream of such studies continues, stimulated primarily by the observation of remarkable differences in postsurgical morbidity and fatality among hospitals, particularly in the National Halothane study (Bunker et al., 1969). The chief refinement of method in such studies has been a greater sophistication in making adjustments for differences in case mix that might influence outcome independently of the quality of care. The earliest adjustments, as in the work of Lipworth et al. (1963) to which I have referred, were simply for diagnosis, age, and sex. Later Roemer et al. (1968) proposed an ingenious adjustment for length of stay and hospital occupancy. More recent studies have supplemented information about age and sex by information gleaned from hospital discharge abstracts. Perhaps the most elaborate system of adjustments, using information on the relative risk of postsurgical fatality collected expressly for the purpose, is to be found in one part of the study of “institutional differences” conducted by a group of investigators at Stanford University (Staff of the Stanford Center for Health Care Research, 1976). The system devised by Gonnella for staging the progression of illness can also be considered to serve a similar purpose, among others, and to be a more recent, more highly elaborated form of an approach with rather ancient beginnings (Gonnella and Goran, 1975; Gonnella et al., 1976).

In a second class of outcome studies are those in which the identification of adverse outcomes is merely a trigger to the assessment of antecedent process. These studies derive their parentage, at least in part, from the “end result” system of Codman (1916), and the studies of preventable maternal and perinatal mortality that I mentioned earlier (Hooker, 1933; Kohl, 1955).

Most important among these more recent developments is the work of John Williamson (1971, 1978a) at Johns Hopkins University. In essence, the method consists of the specification of outcomes that one expects to achieve, the verification of

whether or not the expectations have been met, and an assessment of antecedent process, as well as of structure, if the expectations have not been met. With regard to method, the more notable features of this procedure include:

- (1) an order of priority for selecting conditions for study based on the principle of "maximum achievable benefit";
- (2) a carefully specified procedure, based on the Nominal Group Process, for selecting conditions and specifying goals;
- (3) an insistence that these conditions and these goals be specific to any given institution and to the population they serve, and that representatives of all members of the institution participate in their selection;
- (4) a recognition that outcomes may include a wider range of phenomena, beyond the conventional measures of morbidity and mortality;
- (5) a recognition that randomness introduces a specifiable degree of variability in the outcomes experienced by samples of patients, and that action is not indicated unless the observed frequency of such events is not easily explained by chance; and
- (6) the development of a standardized scale of functional states that can be used to measure outcomes in all conditions instead of, or in addition to, more discrete measures of outcomes more specific to each condition.

Besides contributing to methodological improvements of its own, Williamson's method has produced offspring, the most important of which is the "Problem Status Index" developed by Mushlin and Appel (1980) based on earlier work by Barr (1974)—all the work being based at Johns Hopkins. The method is adaptable to setting goals either for individual patients or categories of patients in groups, and it is designed to identify individual patients whose care is suspected of having been substandard. Mushlin and Appel have supplied a fair amount of information on the sensitivity and specificity of the method, and on the reasons for the failure to achieve expected outcomes.

As might have been expected, studies of outcome have shared in the more systematic attention to the process by which

criteria are selected and specified. The work of Williamson is particularly notable for its early adoption of a carefully specified procedure for selecting conditions that could be accorded high priority in the assessment of outcomes, and for checking the reliability and, to some extent, the validity of such choices (Williamson, 1978b; Williamson et al., 1978, 1979). Another scheme for selecting conditions for assessment is the "tracer method" proposed by Kessner et al. (1973) with a view to obtaining a representative or illustrative view of a health-care system.

As to the studies of how outcome criteria can be and perhaps should be formulated, the work of the group of investigators at the Rand Corporation stands as the supreme example (Brook et al., 1977). Notable in this work, in addition to careful description of the method by which the criteria were arrived at, is the emphasis on carefully specifying the categories of patients to be assessed, the method for measuring outcomes, the precise time at which outcome measurement should occur, and the sources of information concerning patient characteristics and outcomes. We still need to see, however, how the method performs in full-scale field tests.

Another line of development promises to contribute to quality assessment by providing measures of outcome that represent more accurately and comprehensively several aspects of the quantity and quality of life. Notable in this regard is the series of studies that began with the seminal work of Fanshel and Bush (Fanshel and Bush, 1970); the progressive refinements of the Sickness Impact Profile (Gilson et al., 1975; Bergner et al., 1981), and the measures of functional status developed at the Rand Corporation (Brook et al., 1979).

ADVANCES IN STUDIES OF "STRUCTURE"

As I move closer to my conclusions, allow me to review very briefly some developments in the assessment of structure as it influences either the process of care or its outcomes.

Perhaps the first significant improvement over the early studies of this relationship occurred when methods of multivariate analysis allowed an assessment of the independent effects of several coexisting variables, and of the extent of variation in quality that they are able to explain. An excellent example is the more thorough analysis of the findings of Payne et al. in Hawaii, conducted by Rhee, under the guidance of Darsky in our own department at the University of Michigan (Rhee, 1976).

A second development of major importance is a growing interest in the more intimate structures and processes of formal organizations and of their influence on quality. Continuing in a tradition already clearly established in the early work of Georgopoulos and Mann (1962), more recent investigators such as Scott et al. (1976, 1979) and Shortell et al. (1976) have studied the influence of variables such as "differentiation," "coordination," "power," "specification of work procedures," "visibility of consequences," and so on. The operational measurement of these variables in health-care organizations is, itself, a matter of great interest and considerable difficulty. Unfortunately, in many studies of such organizational characteristics, the measures of the dependent variable, which is quality, tend to be rather crude, being primarily measures of case fatality, postoperative complications, or reputations for good performance. Therefore, the findings, which are often rather unexpected as well, are difficult to interpret or accept. We urgently need to supplement the mere observation of adverse outcomes with investigations of how these outcomes were brought about.

Let me also point out that the studies of the relationship between structural attributes and quality are now and always have been primarily concerned with the attributes of the individual or institutional providers of care. Study of variations in the quality of care received by consumers with different characteristics is a much neglected field, though this is a matter of utmost importance to social policy (Wyszewianski and Donabedian, 1981). Perhaps we are unwilling to face the

possibility that even if near equality in access is achieved, subsequent differences in the quality of care might persist.

CONCLUSIONS

Although during the last twenty years of research on the quality of health care we have seen a consolidation of some previous gains, as well as a number of notable advances, we still face, as I have detailed elsewhere, a formidable array of challenging problems (Donabedian, 1978). Among these are two that I believe demand our most dedicated attention.

First, and at a most fundamental level, we need to look into the nature of quality itself, so that the conceptions we have of it are socially more relevant and scientifically more valid. This requires that all presumed relationships between process and outcome be rigorously tested while our technical concerns are supplemented by attention to the interpersonal, social, even moral dimensions of quality. It follows that individual and social valuations must enter the assessment of both the means and the results of health care. As for the means, we cannot continue as if monetary cost were still of no concern. We are obligated, it seems to me, to devise and test strategies of care that achieve the greatest improvements in health at lowest cost. Having such strategies at hand, we would be better able to resist the pressure to accept lower levels of quality because we cannot afford the cost. And if not, we would at least be able to show a truer picture of the losses and gains.

Second, we must pay much more attention than we have done in the past to the determinants of clinically relevant behaviors in the health-care system, and to the means of bringing about desired changes in behavior. The truest concepts of quality and the most elegant methods of assessment will mean little unless we are able to bring about the changes that make our realities correspond to our aspirations much more closely than they now do.

REFERENCES

- BARNETT, G. O. (1984) "The application of computer-based medical-record systems in ambulatory practice." *New England J. of Medicine* 310 (June 21): 1643-1650.
- R. WINICKOFF, J. L. DORSEY, M. MORGAN, and R. S. LURIE (1978) "Quality assurance through automated monitoring and feedback using a computer-based medical information system." *Medical Care* 16 (November): 962-970.
- BARR, D. M. (1974) "The patient's problem: concept, method, and use in health services research." Unpublished master's thesis, Johns Hopkins University.
- BERGNER, M. and R. A. BOBBITT with W. B. CARTER and B. S. GILSON (1981) "The Sickness Impact Profile: development and final revision of a health status measure." *Medical Care* 19 (August): 787-805.
- BROOK, R. H. (1973) *Quality of Care Assessment: A Comparison of Five Methods of Peer Review*. Washington, DC: Department of Health, Education, and Welfare, Bureau of Health Services Research.
- A. DAVIES-AVERY, S. GREENFIELD, L. J. HARRIS, T. LELAH, N. E. SOLOMON, and J. E. WARE, Jr. (1977) "Assessing the quality of medical care using outcome measures: an overview of the method." *Medical Care* 15 (September Supplement): 1-165.
- BROOK, R. K., J. E. WARE, A. DAVIES-AVERY, A. L. STEWART, C. A. DONALD, W. H. ROGERS, K. N. WILLIAMS, and S. A. JOHNSTON (1979) "Overview of adult health status measures fielded in Rand's Health Insurance Study." *Medical Care* 17 (July Supplement): 1-131.
- BUNKER, J. (1970) "Surgical manpower: a comparison of operations and surgeons in the United States and in England and Wales." *New England J. of Medicine* 282 (January 15): 135-144.
- BUNKER, J. P., W. H. FORREST, Jr., F. MOSTELLER, and L. D. VANDAM [eds.] (1969) *The National Halothane Study: A Study of the Possible Association between Halothane Anesthesia and Postoperative Hepatic Necrosis*. Bethesda, MD: National Institutes of Health.
- CIOCCO, A., G. H. HUNT, and I. ALTMAN (1950) "Statistics on clinical services to new patients in medical groups." *Public Health Reports* 65 (January): 99-115.
- CODMAN, E. A. (1916) *A Study in Hospital Efficiency*. Boston: Thomas Todd.
- DAILY, E. F. and M. A. MOREHEAD (1956) "A method of evaluating and improving the quality of medical care." *Amer. J. of Public Health* 46 (July): 848-854.
- DALKEY, N. C., D. L. ROURKE, R. LEWIS, and D. SNYDER (1972) *The Quality of Life: Delphi Decision-Making*, Lexington, MA: D. C. Heath.
- DELBECQ, A. L. and A. H. Van de VEN (1971) "A group process model for problem identification and program planning." *J. of Applied Behavioral Sci.* 7 (July-August): 466-492.
- DONABEDIAN, A. (1982) *Explorations in Quality Assessment and Monitoring, Volume 2. The Criteria and Standards of Quality*. Ann Arbor, MI: Health Administration Press. [See *Methods of Deriving the Criteria*, pp. 149-199; *Stringency of the Standards and Importance of the Criteria*, pp. 203-292; *Consensus, Replicability, Stability, and Relevance*, pp. 297-347.]

- (1978) *Needed Research in the Assessment and Monitoring of the Quality of Medical Care*. Hyattsville, MD: National Center for Health Services Research, Office of Scientific and Technical Information.
- EISELE, C. W., V. N. SLEE, and R. G. HOFFMAN (1956) "Can the practice of internal medicine be evaluated?" *Annals of Internal Medicine* 44 (January): 144-161.
- FANSHEL, S. and J. W. BUSH (1970) "A health status index and its application to health services outcomes." *Operations Research* 18 (November-December): 1021-1060.
- FALK, I. S., H. K. SCHONFELD, B. R. HARRIS, S. LANDAU, and S. S. MILLS (1967) "The development of standards for the audit and planning of medical care. I. concepts, research design and the content of primary care." *Amer. J. of Public Health* 57 (July): 1118-1136.
- FITZPATRICK, T. B., D. C. RIEDEL, and B. C. PAYNE (1962) "The effectiveness of hospital use," "Criteria and data collection," "Appropriateness of admission and length of stay," "The physician interview," "Summary and conclusions," and "Criteria of effectiveness of hospital use," chapters 24-28 and Appendix 2-B, pages 449-526 and 545-559 in W.J. McNerney and study staff, *Hospital Economics: Services, Costs, Methods of Payment, and Controls*, Volume 1. Chicago: Hospital Research and Educational Trust.
- GEORGOPOULOS, B. S. and F. C. MANN (1962) *The Community General Hospital*. New York: Macmillan.
- GILSON, B. S., J. S. GILSON, and M. BERGNER (1975) "The sickness impact profile: development of an outcome measure of health care." *Amer. J. of Public Health* 65 (December): 1304-1310.
- GOLDMAN, F. and E. A. GRAHAM (1954) *The Quality of Medical Care Provided at the Labor Health Institute, St. Louis, Missouri*. St. Louis: The Labor Health Institute.
- GONNELLA, J. S. and M. J. GORAN (1975) "Quality of patient care—a measurement of change: the staging concept." *Medical Care* 14 (June): 467-473.
- GONNELLA, J. S., D. Z. LOUIS, and J. J. McCORD (1976) "The staging concept—an approach to the assessment of outcome of ambulatory care." *Medical Care* 14 (January): 13-21.
- GREENFIELD, S., C. E. LEWIS, S. H. KAPLAN, and M. B. DAVIDSON (1975) "Peer review by criteria mapping: criteria for diabetes mellitus, the use of decision-making in chart audit." *Annals of Internal Medicine* 83 (December): 761-770.
- GREENFIELD, S., S. CRETIN, L. G. WORTHMAN, F. J. DOREY, N. E. SOLOMON, and G. A. GOLDBERG (1981) "Comparison of a criteria map to a criteria list in quality-of-care assessment for patients with chest pain: the relation each to outcome." *Medical Care* 19 (March): 255-272.
- GREENFIELD, S., S. CRETIN, L. G. WORTHMAN, and F. DOREY (1982) "The use of an ROC curve to express quality of care results." *Medical Decision Making* 2 (Spring): 23-31.
- GRIFFITH, J. R., J. D. RESTUCCIA, P. J. TEDESCHI, P. A. WILSON, and H. S. ZUCKERMAN (1981) "Measuring community hospital services in Michigan." *Health Services Research* 16 (Summer): 135-160.

- HARE, R. L. and S. BARNOON (1973) *Medical Care Appraisal and Quality Assurance in the Office Practice of Internal Medicine*. San Francisco: American Society of Internal Medicine.
- HOOVER, R. S. (1933) *Maternal Morality in New York City: A Study of All Puerperal Deaths 1930-1932*. New York: Oxford Univ. Press.
- HULKA, B. S., F. J. ROMM, G. R. PARKERSON, Jr., I. T. RUSSEL, N. E. CLAPP, and F. S. JOHNSON (1979) "Peer review in ambulatory care: use of explicit criteria and implicit judgments." *Medical Care* (March Supplement): 1-73.
- KANE, R. L., L. Z. RUBINSTEIN, R. H. BROOK, J. Van RYZIN, P. MASTHAY, E. SCHOENRICH, and B. HARRELL (1981) "Utilization review in nursing homes: making implicit level-of-care judgments explicit." *Medical Care* 19 (January): 3-13.
- KESSNER, D. M., G. E. KALK, and J. SINGER (1973) "Assessing health quality—the case for tracers." *New England J. of Medicine* 288 (January 25): 189-194.
- KOHL, S. G. (1955) *Perinatal Mortality in New York City: Responsible Factors*. Cambridge, MA: Harvard Univ. Press.
- LEE, R. I. and L. W. JONES (1933) *The Fundamentals of Good Medical Care*. Chicago: Univ. of Chicago Press.
- LEMBCKE, P. A. (1967) "Evolution of the medical audit." *J. of the Amer. Medical Assn.* 199 (February 20): 543-550.
- (1956) "Medical auditing by scientific methods: illustrated by major female pelvic surgery." *J. of the Amer. Medical Assn.* 162 (October 13): 646-655.
- (1952) "Measuring the quality of medical care through vital statistics based on hospital service areas: I. Comparative study of appendectomy rates." *Amer. J. of Public Health* (March): 276-286.
- [ed.] (1947) *The Regional Hospital Plan—The Second Year's Experience*. Rochester, NY: Council of Rochester Regional Hospitals, pp. 27-29.
- LEWIS, C. E. (1969) "Variations in the incidence of surgery." *New England J. of Medicine* 281 (October 16): 880-884.
- LIPWORTH, L., J. A. H. LEE, and J. N. MORRIS (1963) "Case fatality in teaching and non-teaching hospitals 1956-59." *Medical Care* 1 (April-June): 71-76.
- MAKOVER, H. B. (1951) "The quality of medical care." *Amer. J. of Public Health* 41 (July): 824-832.
- MATES, S. and V. W. SIDEL (1981) "Quality assessment by process and outcome methods: evaluation of emergency room care of asthmatic adults." *Amer. J. of Public Health* 71 (July): 687-693.
- McCLAIN, J. O. and D. C. RIEDEL (1973) "Screening for utilization review: on the use of explicit criteria and non-physicians in case selection." *Amer. J. of Public Health* 63 (March): 247-251.
- McDONALD, C. T. (1976) "Protocol-based computer reminders, the quality of care and the non-perfectability of man." *New England J. of Medicine* 295 (December 9): 1351-1355.

- McNEIL, B. J. and S. J. ADELSTEIN (1975) "Measures of clinical efficacy: the values of case finding in hypertensive renovascular disease." *New England J. of Medicine* 293 (July 31): 221-226.
- McNEIL, B. J., S. J. HESSEL, W. T. BRANCH, L. BJORK, and S. J. ADELSTEIN (1976) "Measures of clinical efficacy. III. The value of the lung scan in the evaluation of young patients with pleuritic chest pain." *J. of Nuclear Medicine* 17 (March): 163-169.
- McNEIL, B. J., P. D. VARADY, B. A. BURROWS, and S. J. ADELSTEIN (1975) "Measures of clinical efficacy: cost-effectiveness calculations in the diagnosis and treatment of hypertensive renovascular disease." *New England J. of Medicine* 293 (July 31): 216-221.
- McNEIL, B. J., R. WEICHSELBAUM, and S. G. PAUKER (1981) "Speech and Survival: tradeoffs between quality and quantity of life in laryngeal cancer." *New England J. of Medicine* 305 (October 22): 982-987.
- (1978) "Fallacy of the five year survival in lung cancer." *New England J. of Medicine* 299 (December 21): 1397-1401.
- McPHERSON, K., J. E. WENBERG, O. B. HOVIND, and P. CLIFFORD (1982) "Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway." *New England J. of Medicine* 307 (November 18): 1310-1314.
- MOREHEAD, M. A. (1970) "Evaluating the quality of medical care in the neighborhood center program of the Office of Economic Opportunity." *Medical Care* 8 (March-April): 118-1131.
- (1967) "The medical audit as an operational tool." *Amer. J. of Public Health* 57 (September): 1643-1656.
- (1958) *Quality of Medical Care Provided by Family Physicians as Related to Their Education, Training and Methods of Practice*. New York: Health Insurance Plan of Greater New York.
- and R. DONALDSON (1974) "Quality of clinical management of disease in comprehensive neighborhood health centers." *Medical Care* 12 (April): 301-315.
- and others (1964) *A Study of the Quality of Hospital Care Secured by a Sample of Teamster Family Members in New York City*. New York: School of Public Health and Administrative Medicine, Columbia University.
- MOREHEAD, M. A., R. S. DONALDSON, and M. R. SERAVELLI (1971) "Comparisons between DEO Neighborhood Health Centers and other health care providers of ratings of the quality of health care." *Amer. J. of Public Health* 61 (July): 1294-1306.
- MUSHLIN, A. I. and F. A. APPEL (1980) "Testing and outcome-based quality assurance strategy in primary care." *Medical Care* 18 (May Supplement): 1-100.
- NOVICK, L. F., K. DICKINSON, R. ASNES, S-P. M. LAN, and R. LOWENSTEIN (1976) "Assessment of ambulatory care: applications of the tracer methodology." *Medical Care* 19 (January): 1-12.
- OSBORNE, C. E. and H. C. THOMPSON (1975) "Criteria for evaluation of ambulatory child health care by chart audit: development and testing of a methodology." *Pediatrics* 56 (October, Part 2): 625-692.
- PALMER, R. H., R. STRAIN, J. V. W. MAURER, J. K. ROTHROCK, and M. S.

- THOMPSON (1984) "Quality assurance in eight adult medicine group practices." *Medical Care* 22 (July): 632-643.
- PAYNE, B. C., T. F. LYONS, L. DWARSHIUS, M. KOLTON, and W. MORRIS (1976) *The Quality of Medical Care: Evaluation and Improvement*. Chicago: Hospital Research and Educational Trust.
- PETERSON, O. L. and E. BARSAMIAN (1964) "Diagnostic performance," pp. 347-362 in J. A. Jacquez (ed.) *The Diagnostic Process*. Ann Arbor: University of Michigan School of Medicine.
- and E. MURRAY (1966) "A study of diagnostic performance: preliminary report." *J. of Medical Education* 41 (August): 797-803.
- PETERSON, O. L., L. P. ANDREWS, R. S. SPAIN, and B. G. GREENBERG (1956) "An analytical study of North Carolina general practice, 1953-54." *J. of Medical Education* 31 (December, part 2): 1-165.
- RHEE, S.-O. (1976) "Factors determining the quality of physician performance in patient care." *Medical Care* 14 (September): 733-750.
- RIEDEL, R. L. and D. C. RIEDEL (1979) *Practice and Performance: An Assessment of Ambulatory Care*. Ann Arbor, MI: Health Administration Press.
- ROEMER, MI., A. T. MOUSTAFA, and C. E. HOPKINS (1968) "A proposed hospital quality index: hospital death rates adjusted for case severity." *Health Services Research* 3 (Summer): 96-118.
- SCHONFELD, H. K., J. F. HENSTON, and I. S. FALK (1975) *Standards for Good Medical Care*, Vols. 1-4. Washington, DC: Government Printing Office.
- SCOTT, W. R., W. H. FORREST, Jr., and B. W. BROWN (1976) "Hospital structure and postoperative mortality and morbidity," pp. 72-89 in S. M. Shortell and M. Brown (eds.) *Organizational Research in Hospitals*. Chicago: Blue Cross Association.
- SCOTT, W. R., A. B. FLOOD, and W. EWY (1979) "Organizational determinants of services, quality and cost of care in hospitals." *Milbank Memorial Fund. Q./Health and Society* 57 (Spring): 234-264.
- SHAPIRO, S., L. WEINER, and P. M. DENSEN (1958) "Comparison of prematurity and perinatal mortality in a general population and in the population of a prepaid group practice, medical care plan." *Amer. J. of Public Health* 48 (February): 170-187.
- SHAPIRO, S., H. JACOBZINER, P. M. DENSEN, and L. WEINER (1960) "Further observations on prematurity and perinatal mortality in a general population and in the population of a prepaid group practice medical care plan." *Amer. J. of Public Health* 50 (September): 1304-1317.
- SHORTELL, S. M., S. W. BECKER, and D. NEUHAUSER (1976) "The effects of managerial practices on hospital efficiency and quality of care," pp. 90-107 in S. M. Shortell and M. Brown (eds.) *Organizational Research in Hospitals*: Blue Cross Association.
- SOUTHMAYD, H. J. and G. SMITH (1944) *Small Community Hospitals*. Cambridge, MA: Harvard Univ. Press.
- Staff of the Stanford Center for Health Care Research (1976) "Comparison of hospitals with regard to outcomes of surgery." *Health Services Research* (Summer): 112-127.

- STARFIELD, B., D. STEINWACHS, I. MORRIS, G. BAUSE, S. SIEBERT, and C. WESTIN (1979) "Concordance between medical records and observations regarding information on coordination of care." *Medical Care* 17 (July): 758-766.
- STOCKWELL, H. and E. VAYDA (1979) "Variations in surgery in Ontario." *Medical Care* 17 (April): 390-396.
- VAYDA, E. (1973) "A comparison of surgical rates in Canada and in England and Wales." *New England J. of Medicine* 289 (December 6): 1224-1229.
- WAGNER, E. H., R. A. GREENBERG, P. B. IMPREY, C. A. WILLIAMS, S. H. WOLF, and M. A. IBRAHIM (1976) "Influence of training and experience on selecting criteria to evaluate medical care." *New England J. of Medicine* 294 (April 16): 871-876.
- WAGNER, E. H., C. A. WILLIAMS, R. GREENBERG, D. KLEINBAUM, S. WOLF, and M. A. IBRAHIM (1978) "A method for selecting criteria to evaluate medical care." *Amer. J. of Public Health* 68 (May): 464-470.
- WEED, L. L. (1971) "Quality control and the medical record." *Archives of Internal Medicine* 127 (January): 101-105.
- WEED, L. L. (1969) *Medical Records, Medical Education, and Patient Care*. Chicago: Year Book Medical.
- WENBERG, J. and A. GITTELSON (1973) "Small area variations in health care delivery." *Science* 182 (December 14): 1102-1108.
- WILLIAMSON, J. W. (1978a) *Assessing and Improving Health Care Outcomes: The Health Accounting Approach to Quality Assurance*. Cambridge, MA: Ballinger.
- (1978b) "Formulating priorities for quality assurance: description of a method and its application." *J. of the Amer. Medical Assn.* 239 (February 13): 631-637.
- (1971) "Evaluating quality of patient care: a strategy relating outcome and process assessment." *J. of the Amer. Medical Assn.* 218 (October 25): 564-569.
- H. R. BRASWELL, and S. D. HORN (1979) "Validity of medical staff judgments in establishing quality assurance priorities." *Medical Care* 17 (April): 331-346.
- and S. LOHMEYER (1978) "Priority setting in quality assurance: reliability of staff judgments in medical institutions." *Medical Care* 16 (November): 931-940.
- WYSZEWIANSKI, L. and A. DONABEDIAN (1981) "Equity in the distribution of quality of care." *Medical Care* 19 (December Supplement): 28-56.
- ZUCKERMAN, A. E., B. STARFIELD, C. HOCHREITER, and B. KOVASZMAY (1975) "Validating the content of pediatric outpatient medical records by means of tape recording doctor-patient encounters." *Pediatrics* 56 (September): 407-411.