

## Commentary on some studies of the quality of care

by Avedis Donabedian

*This essay is a personal commentary on general aspects of quality assessment that are either illustrated or challenged by the articles presented in this volume. It is not a critique of the individual papers in the collection; papers are mentioned in order to illustrate or clarify a point, without implication of either merit or demerit. This article addresses the definitions of quality, the relationship of cost to quality, approaches*

*to assessing quality, the use of case fatality as an indicator of quality, measures of post-hospital morbidity, case-mix adjustment, sources of information, selection of referents, criteria and standards, quality assurance, the role of consumers, the prevalence of quality, the effects of prospective payment, and a research agenda.*

### Introduction

This essay is not meant to be a critique of the individual articles and symposium contributions in this collection. Rather, it is a commentary on the more general aspects of quality assessment that are either illustrated or challenged. Accordingly, the organization of this article is thematic. Only occasionally are particular contributors mentioned, and then mainly to illustrate or clarify a point, without necessarily implying either merit or demerit.

### Definition of quality

Every attempt to assess the quality of care must begin with a conceptual formulation that defines quality in general and also with reference to the particular context in which the assessment is to take place (Donabedian, 1980). When quality is most narrowly defined, the object of assessment is the technical performance of individual health care practitioners. As the definition of quality is broadened, it may include, successively:

- The manner in which a practitioner manages the personal interaction with the patient.
- The patient's own contribution to care.
- The amenities of the settings in which care is provided.
- Facility in access to care.
- The social distribution of access.
- The social distribution of the health improvements attributable to care.

At the same time, the definition of "health," which is the product of care, broadens beyond physical and physiological function to include, ultimately, something very much like the "quality of life."

In the articles presented in this special issue, we see almost the full gamut of these several definitions. At one extreme, no doubt constrained by the availability of information, quality merely means not dying, the most basic necessity of all. By contrast, Berwick and Knapp offer a richly multidimensional view of

quality, one that includes not only technical care but also the interpersonal process, the ambience of the settings in which care is given, and ease of access. A further extension occurs when Bates and Connors propose to assess access in response to symptoms and to determine whether the aged, as a group, are less able to enter and use more highly structured forms of organized health care. Whether or not, after gaining access, the aged or any other subgroup of persons are not treated as well as other patients is a question that we hesitate to ask but must eventually confront. Ultimately, it is the health of those for whom health care providers are responsible, whether they come to us as patients or not, that should be the measure of our success.

It is more than a coincidence that the broader definitions of quality are formulated in the context of health maintenance organizations. Dependence on attracting and keeping enrollees, as well as responsibility for the health of this defined population group, should compel attention to aspects of care and measures of accomplishment that are less salient or less relevant in other settings (Donabedian, 1983).

Concern for assessing entire episodes of care, rather than disjointed segments, is more likely to characterize the assessment of care in health maintenance organizations. However, assessing episodes of care that include hospitalization is an option that is more broadly available (Payne et al., 1976). In fact, the concern, uniformly encountered in these articles, for assessing events subsequent to discharge from the hospital is only partly an attempt to capture the delayed effects of poor quality in hospital care. In part, whether it is obtained for the purpose or not, this information reflects the nature of care for an episode that may begin with a hospital stay but does not end there.

### Cost and quality

The relationship of cost to quality continues to be a source of difficulty, both when quality is to be defined and when the consequences of quality monitoring and cost containment are to be assessed (Donabedian, Wheeler, and Wyszewianski, 1982; Donabedian, 1980).

Reprint requests: Avedis Donabedian, University of Michigan School of Public Health, Department of Health Services Management and Policy, 109 Observatory Street, Ann Arbor, Michigan 48109.

The simplest formulation of technical quality is to identify it with effectiveness, by which I mean the degree to which improvements in health currently attainable by the science and technology of health care are, in fact, obtained (or can be expected to be obtained, given information on what the practitioner has done). Under this formulation, the cost of obtaining any given improvement in health care would reflect efficiency, not quality.

If we accept this formulation, we can go on to identify several varieties of inefficiency, only some of which are the result of clinical decisions. The managers of health care institutions are responsible for running them efficiently so that goods and services of acceptable quality are made available to clinicians in planning and implementing care. Although this form of efficiency can be expected to have a profound effect on the cost of care, it is not a component of "quality," as that term is usually understood.

A second form of inefficiency occurs when clinicians include in their care elements that are unnecessary but virtually harmless. I have argued that this form of inefficiency should also be recognized as poor quality, because it is the result of poor judgment or carelessness. Besides, there is precious little in health care that does not pose some hazard, even if small.

Another form of inefficiency that can result from clinical decisions comes from adding elements of useful care even though the corresponding increases in health improvement are too small to justify the added expense. This formulation depends on a prior assumption of diminishing returns, which means that added health improvements become smaller and smaller when care becomes progressively more elaborate and complete, even though everything done can be expected, on the average, to be appropriate and useful. This formulation has the added assumption that we have an accurate and legitimate method for placing a money value on improvements (or deteriorations) in health.

Under these circumstances, we have an option of specifying the best quality as either the care that can be expected to produce the greatest improvement in health or the care that can be expected to produce the optimum improvement in health when costs and benefits are compared. Under the first option, inefficiency is separate from quality; under the second option, inefficiency constitutes less than the best quality.

By offering these options in passing judgments on clinical performance, I do not mean to imply indifference to the choice. On the contrary, the choice, particularly with regard to the distinctions between maximum and optimum improvements in health, have far-reaching consequences, both practical and ethical (Donabedian, 1986).

Obviously, the empirical determination of the relationship between cost and quality will be influenced by the definitional considerations just detailed. It will also be influenced by the inclusivity with which costs and benefits are measured. For

example, the costs of an entire episode should be measured so that the cost of all care is encompassed. Costs borne by a health care financing program are not complete; costs borne by the patient and the family, directly or indirectly, are not included. In a definitive analysis, one may also wish to consider the long-term effects of quality, or the lack of it, in gains or losses to productivity and in reductions of or additions to the cost of caring for people over a lifetime. In light of this brief exposition, it is no surprise that the findings reported in this issue about the relationship between cost and quality are so ambiguous and inconclusive.

It is fashionable these days to adopt a rather nihilistic stance in speculating about the value of health care, or at least of additions to it. This is a healthy skepticism necessary to the investigator. At the same time, we ought to realize that speculations not based on conclusive evidence can be seized on as fact. Some health planners are avidly searching for a pretext to cut back on the investment in health care, particularly in the public sector. We must be certain that what we tell them is truth, not merely conjecture.

## Approach to assessing quality

I have identified three approaches to assessing quality: structure, process, and outcome (Donabedian, 1966). Happily, the terminology has been widely adopted and often properly used. Unhappily, it is too frequently misunderstood and abused.

The importance of structure as an influence on clinical performance is amply evident in these articles, for example, in the exploration by Gaumer, Poggio, and Sennett of the effects of hospital characteristics on postadmission case fatality and in the proposal by Bates and Connors to compare the performance of individual practice associations with other forms of prepaid practice, both differentiated further by rate of growth. Less often studied are the more subtle features of organization: differentiation, coordination, power, specification of work procedure, visibility of consequences, and so on (Georgopoulos and Mann, 1962; Scott, Forrest, and Brown, 1976; Scott, Flood, and Ewy, 1979; Shortell, Becker, and Neuhauser, 1976). This more detailed study is needed so we can tell by what mechanisms the more obvious features of organizational structure exert their influence.

Apparently more tangled, and certainly more subject to differences of opinion, is the relationship of process to outcome. As described by Berwick and Knapp, some hold outcome to be a surrogate for process, whereas others maintain the reverse. The uncertainties of the process-outcome relationship impel some to seek refuge in the superior face validity of outcomes, whereas others choose to shake off what Berwick and Knapp call the "tyranny" of outcomes, asserting that what health care delivers is not outcomes "but rather process, itself."

In my opinion, the truth is in neither camp but somewhere in between. As I have demonstrated in detail elsewhere, there is a nearly perfect symmetry between process and outcome in assessments of quality (Donabedian, 1980). The validity of either depends on the validity of the assumed causal linkage between the two. If that is valid, either can be used to assess quality; if that is invalid, neither can be used. Process and outcome are, therefore, complements to each other in the assessment of quality, not alternatives. A preference for one over the other must be based, not on causal validity, but on something else—the availability of information, for example, ease of accurate measurement, or timeliness relative to the uses to which the information is to be put.

That is not to say that measures of process and outcome, each taken as a whole, do not have distinctive characteristics. Process measures, in general, are more timely, sensitive, and specific. Outcomes, by their nature, are delayed, less sensitive, and less specific. Outcomes, however, have the advantages of being more comprehensible to consumers and of reflecting all antecedent care, including clinical judgment and skill as well as the contribution of patients to their own care. However, by enjoying this degree of inclusivity, researchers using outcome as a measure suffer in not being able to tell us precisely what may have gone wrong, in whose hands.

There are, of course, exceptions to all this. Changes in physiological status by which clinicians guide the management of patients are micro-outcomes that are timely, sensitive, and often specific (Brewster et al., 1985). Similarly, although patients often are in no position to assess the technical quality of the process of care, they are exquisitely sensitive to the finest nuances of the interpersonal relationship (Donabedian, 1980). Perhaps no aspect of process is so subtle that it has no counterpart in an outcome, be it a change in health, in knowledge, in behavior, or in satisfaction.

A distinction in the use of outcomes to assess quality is relevant to these articles and also recognized by several of their authors. As Lohr and also Iezzoni and Moskowitz clearly recognize (symposium), it is one thing to use an outcome (for example, case fatality) as a measure of quality without additional verification; it is quite another thing to use an outcome only to indicate that more detailed assessment of process is needed. As Lohr points out, a modest amount of adjustment for differences in case mix, although it may not suffice to make an outcome a sufficiently precise measure of quality, can still contribute immensely to making that outcome a much more efficient screening device for directing attention to the subset of cases that demand careful review.

On the whole, there is not only efficiency but also safety in the joint assessment of outcome and process. Any discrepancy between the two is a signal that one or the other of the two approaches is misspecified or is deficient in some other way. Sometimes, as Knaus

(symposium) beautifully illustrates, the comparison leads to an unexpected discovery: in this case, a discovery concerning the efficacy of a particular method for treating neonates with acute respiratory problems.

Such serendipitous observations should not, however, lead us to believe that quality monitoring, besides being an administrative tool, is also a form of clinical research. The information used in monitoring is seldom precise enough or gathered under sufficiently controlled conditions to permit confident conclusions about the relative efficacy of varieties of care. The verification of efficacy is the responsibility of the clinical research establishment. Quality assessment contributes merely by calling attention to unexpected conjunctions between process and outcome.

### **Case fatality as an indicator**

No one would argue that case fatality should be excluded as a measure of quality. At the same time, everyone would agree with Eggers that death is “a worst case scenario.” Other outcomes are needed, including measures of functional performance, of psychosocial adaptation, of knowledge gained, of behavior changed, and of satisfaction with the outcomes of care and with the manner in which it is received. The virtual absence of such measures from assessment of the impact of prospective payment is mostly attributable to the unavailability of information. However, I wonder whether the lack of attention to the accessibility and acceptability of care is not also the result of a rather constricted view of public responsibility for health.

The aspect of case fatality that has attracted the greatest attention is its lack of specificity, a defect attributable to the presence of so many factors other than the quality of care that influence fatality. Accordingly, adjustment for case mix, the device calculated to neutralize the effect of extraneous factors, occupies center stage in many of the articles, as in other literature on the subject. I agree with Lohr (symposium), however, in believing that no amount of adjustment for case mix, whether in using case fatality or using any other outcome, can remedy other defects, for example, in the accuracy and completeness of information or, more fundamentally, in the linkage between process and outcome.

Partly because it is influenced by so many extraneous factors and partly because it represents such massive failure in care, case fatality is not a sensitive measure of variations in quality that could be reflected in other outcomes or, with even greater sensitivity, detected through examining the antecedent process of care. For many conditions, the risk of dying as a result of poor care is so low that the avoidance of fatality is hardly an objective of care. Perhaps this is why Gaumer and colleagues find that prospective payment may have adversely affected fatality from transurethral prostatectomy without having had a noticeable effect on fatality from repair

of inguinal hernia. Hemorrhoidectomy, another procedure included in the study, may have been another inappropriate choice.

An outcome can become more or less important as a measure of quality, depending on the valuation placed on it by clients as they weigh the desirability or undesirability of the consequences of alternative forms of care that are available options. As McNeil has so elegantly demonstrated, even the prospect of death can be more or less abhorrent, depending on what the alternative consequences are (McNeil, Weichselbaum, and Parker, 1978 and 1981).

Fatalities associated with specific diagnostic entities are incomplete in still another way, that of not including the consequences to cases not diagnosed or improperly diagnosed. Unless a discriminating search for the most likely misdiagnoses is made, only the fatality of an entire caseload can reflect this kind of incompleteness. For an even more complete accounting, one that includes cases that have not been hospitalized at all, the mortality experience of an entire population needs to be examined at the expense of great attenuation in both sensitivity and specificity.

Still another form of incompleteness, one fully recognized in the articles, results from the persistence of the effects of in-hospital care subsequent to discharge. The remedy, of course, is to extend the period of observation. The importance of doing so is amply documented. Roos, Roos, and Sharp find, for example, that of all deaths occurring within 90 days of surgery, the proportion that occurred after discharge from the hospital was 39 percent when the procedure was a cholecystectomy, 54 percent for prostatectomy, and 69 percent for hysterectomy. Unfortunately, neither Roos and colleagues nor any of our other investigators tell us the degree of correlation between in-hospital deaths and postdischarge deaths. Perhaps this is because that correlation can be altered manipulatively by shortening or lengthening hospital stays, particularly for patients who might be expected to die in any case. However, if so, the ratio of postdischarge to pre-discharge deaths could serve as an indicator of such manipulation.

Because there is some discretion in lengthening or shortening hospital stays, there is general agreement that the point of departure for counting deaths is not the day of discharge but the day of admission or of the surgical intervention. How long the period of observation lasts beyond these points of departure is highly variable. Besides the period of 90 days after surgery, used by Roos and colleagues, one encounters postadmission periods of 6 weeks (Eggers); 15, 30, 45, 90, 180, and 360 days (Gaumer, Poggio, and Sennett); and 30, 90, 180, 270, and 360 days (Krakauer, 1987). In a more extended followup, Krakauer (1987) adds observations at 540 and 720 days. Eggers points out that these observations, whether one or more, are abridged life tables that can be extended to examine the entire distribution of time until death in any cohort of admissions.

The patterns of followup described reflect, no doubt, the still exploratory nature of our studies. The object, apparently, is to capture the consequences of in-hospital care to the fullest extent possible with the least contamination by the effects of subsequent care. However, there is a sense in which posthospital care, if it is part of the episode that included hospitalization, is also relevant to assessing the quality of care for Medicare beneficiaries. Furthermore, because the patients' condition at admission also may reflect prehospital care, as well as access to the hospital, something could be learned from posthospital mortality per 1,000 enrollees, unadjusted for case mix. Eggers proposes this rate, in conjunction with total mortality per 1,000 enrollees, as a useful method of assessing the performance of the Medicare program, taking into account persons who have experienced one hospitalization or more as well as those who have not. The attempt, through case-mix standardization and other means, to separate the hospital's responsibility from that of other sources of care is relevant only if the effects of prospective payment are the specific object of investigation.

## Measures of posthospital morbidity

Driven by the inadequacies of case fatality, the search goes on for measures of morbidity and disability that can be easily derived from available information. Payments by Medicare for ambulatory care, skilled nursing care, and home health services have been proposed as possible measures. We need to recognize, however, the incompleteness of the information on expenditures for these purposes, as well as the highly inferential nature of the interpretation to be placed on the findings.

With readmissions to the hospital, we are on somewhat firmer ground, although readmissions are recognized to be only partly the result of inadequacies in hospital care; partly they are caused by subsequent unrelated illnesses and the care these receive. Readmission is most frequent during the first 30 or 90 days after discharge, which is also the period during which the readmission is most likely to be related to the initial admission. I judge that Roos and his coworkers have made a notable contribution to our methods by developing algorithms that establish more convincingly the relatedness of readmissions to the care received earlier (Roos et al., 1985).

## Case-mix adjustment

Ever since Moses and Mosteller (1968) demonstrated that astoundingly large differences in postsurgical fatality among some teaching centers could be drastically reduced by diagnostic categorization, with additional corrections for other patient characteristics, a fever of case-mix adjustment has seized the land. The purpose of adjustment for differences in case mix is to reveal the true effects of differences in quality by reducing, as much as

possible, the confounding effects of other factors. As Knaus (symposium) argues so persuasively, we want to characterize patients as they present for care, using attributes that indicate, in part, the already predetermined tendencies toward progression or remission and, in part, degrees of challenge to diagnostic acumen, therapeutic judgment, and skill in implementation of care. Therefore, the assessment of the patients' status must be based strictly on biological, physiological, and pathological grounds. In particular, the assessment should not include, as an index of progression or severity, any element dependent on the nature or intensity of the care given subsequently, a specification that renders several widely used classifications (the diagnosis-related groups, for example) less than fully suitable for the purpose.

There are considerations, however, that may lead us to relax these strict requirements. First, I am not sure to what degree the initial state of patients is a sufficient indication of the likelihood that unpreventable subsequent developments may occur. If there are doubts about this, it would be legitimate to include in the classification of patients subsequent changes in status to which the care given has made no contribution either for better or for worse.

Taking another step toward a less rigid stance, we could argue that even if subsequent adverse events are contributed to by prior care, their inclusion could be justified because the ability to deal decisively with the consequences of error is a mark of the good doctor. It is best not to have made the error; but having made it, there is saving grace in having minimized the harm that it may have done. With these considerations in mind, I would not take the position that a retrospective assessment of the patients' status during the hospital stay is uniformly unsuitable for characterizing case mix. As a general rule, the classification of case mix must be suitable to its purposes, and these include uses other than quality assessment.

Whatever the uses to which case-mix adjustment might be put, the accuracy of the data on which it is based is a matter of the greatest concern. Information that, if available, would in itself indicate quality could introduce curious logical distortions in the analysis. For example, care of higher quality might lead to the discovery and recording of comorbidities that move a case upward on the scale of severity and downward in the expectation of good outcome. As another example, inattentive providers of care might miss the occurrence of complications that good care providers would identify and record, to their own disadvantage. As we know, diagnostic categorization is itself often faulty and sometimes subject to intentional manipulation. We are faced, therefore, with an ambiguity in our classification that only a direct independent assessment of patients can fully dispel.

In addition to the inaccuracies in our data, very often we lack the kind of information most critical to a determination of severity and prognosis. We make do, therefore, by using more remote indicators: age,

sex, race, socioeconomic status, whether or not surgery has been done, the number of reported diagnoses, the reported presence of certain comorbidities, and so on. Attributes such as race or socioeconomic status also introduce peculiar biases into the assessment. This is because race and socioeconomic status signify, not inherent biological predispositions, but, to a large extent, obstacles to implementing the most effective care. To treat characteristics of the underprivileged as inherently adverse is to excuse those who care for them from making the extra effort they are called upon to expend. Not to include such characteristics is to ignore some cruel realities in our health care system.

Because the information that we use to adjust for case mix is neither complete nor fully accurate, we continue in doubt as to our success in revealing the differences in quality that we seek. Rather than continuing to argue about the matter, it is time, as Brook (symposium) proposes, to find out by a direct assessment of process what differences in quality we have found and what other differences in quality we have missed.

## Other aspects of assessment

Several of the articles in this issue, particularly those by Berwick and Knapp and by Bates and Connors, offer an occasion for at least some brief remarks on quality assessment and monitoring in general. As I said at the very beginning of these comments, the methods that we employ are conditioned largely by the definition of quality that we adopt. It is no surprise, therefore, that when the definition is multidimensional, the methods are varied.

## Sources of information

Differentiation begins with the sources of information that we seek to tap. Unfortunately, medical records give mainly a picture of technical care, and an often incomplete one at that. The record is particularly incomplete for ambulatory care, especially care provided in the offices of individual practitioners. This is a problem likely to plague at least part of the study proposed by Bates and Connors, despite their decision to confine their attention to information that should appear in the record. Even when assessments are based on much more complete hospital records, we cannot fully neutralize the accusation that the recording of care, rather than care itself, is being judged. One happy consequence of quality monitoring should be to bring about appropriate adaptations in the design and technology of recording so that medical records become more useful tools for clinical management as well as for quality monitoring. Computerized recording systems are particularly promising in this respect (McDonald, 1970; Barnett, 1984).

If we want more definitive information about clinical performance or the capacity for it, or any

information about other dimensions of quality, including accessibility and acceptability, we must use a variety of other sources, including direct observation of clinical practice, assessment of performance under test situations, interviews (whether direct or by phone), mailed questionnaires, and so on. Certain client behaviors (such as complaining, "shopping," disenrolling, or using services outside a prepaid plan) can also signal some possible deficiencies in a system of care.

Berwick and Knapp describe a system of monitoring that is admirable in using a multiplicity of means in a coordinated manner to learn about several components of quality. I balk only at their proclaimed indebtedness to industry for their model. If I had time, I believe I could demonstrate how richly endowed our own tradition of quality assessment is in these respects. We need not go elsewhere for teachers.

### **Selection of referents**

If we wish to obtain a truly representative picture of care, we need, of course, a statistically valid probability sample of cases. This happens infrequently. Often we seek what could be called an illustrative sample or even one that is purposively biased to include the worst examples of care. Accordingly, our plans for assessing or monitoring quality often include an early selection of "referents," a term I have used to mean the conditions or occurrences to be assessed (Donabedian, 1982).

I have already commented on the suitability of certain indicator conditions for detecting variations in quality sufficiently likely to be reflected in different fatality rates. Bates and Connors illustrate still other concerns when they take into account how frequent the referent is, how serious it is, and how important good care is to obtaining favorable results. In this, they seem to follow the rule of "maximum achievable benefit" enunciated by Williamson (Williamson and Miller, 1968).

Bates and Connors make a useful distinction between a condition and a diagnosis, a distinction also encountered in earlier work, for example, that of Brook (Brook et al., 1977). If one begins with a diagnosis, one can only judge if the diagnosis is justified; one cannot tell about the management or fate of cases not diagnosed. "Conditions," by contrast, present as "problems" (for example, headache, abdominal pain, or urinary distress). The conditions selected by Bates and Connors (diabetes and hypertension) are fairly specific. They are selected, however, for the same purpose that the more nonspecific conditions serve. Because their presence is indicated in ways independent of a formal diagnosis, they permit study of attentiveness and diagnostic skill (both initially and at subsequent stages in the development of disease) as well as the ability to appropriately manage each disease in its several manifestations.

### **Criteria and standards**

All assessment requires criteria and standards, because these embody in more concrete form, amenable to measurement, the more general concepts that define the meaning of "quality." It has become customary, since I first used the terms, to distinguish two methods of assessment, one using implicit and the other explicit criteria (Donabedian, 1969). When an expert clinician is asked to assess a medical record without any external guidance, the judgment is based on criteria and standards that are implicit in the sense that they are personal and undeclared. By contrast, explicit criteria are specified in advance, usually by panels of experts, and apply, therefore, only to carefully prescribed referents.

Logically speaking, there are no exclusive relationships among any of the following: variety of criteria (explicit, implicit); an aspect of quality (technical care, management of the interpersonal process); an approach to assessment (structure, process, outcome); or a source of information (medical records, questionnaires, observation of practice, etc.). At least in theory, all combinations are possible.

When Brook (1973) asked physicians to say whether or not an observed outcome could have been better, he expected the respondents to use implicit criteria. By contrast, explicit criteria were enunciated when Mushlin, Appel, and Barr (1978), declared that, within a month of having experienced an upper respiratory infection, no one should continue to have symptoms, disability, or anxiety. Similarly, a physician watching another work could make a judgment based on personal opinion or be given a detailed list of what to look for and how to rate each occurrence observed.

To take still another example, accessibility could be judged using structure (the factors that facilitate or impede access); process (the actual experience of seeking and obtaining care, for example, mostly in response to symptoms); or outcome (the improvement in health that may result), as mentioned by Aday and Andersen (1974). Information about each of these could be obtained in a variety of ways, and each of the findings could be judged without benefit of prior criteria (as when a patient expresses satisfaction with the length of waiting in the clinic) or according to preformulated criteria (as when the objective is that no patient should wait more than 15 minutes before being ushered in to see the doctor).

The reader must look elsewhere for a detailed discussion of the merits and demerits of the two forms of criteria (Donabedian, 1982). I only wish to say that, in my opinion, the notion that the two forms yield highly discrepant judgments of quality is not sustained by the evidence. On the contrary, there are reasonably strong correlations.

As in other cases in which one has two methods with contrasting strengths and weaknesses, a combination of the two could be the best strategy. A rather simple set of explicit criteria could be an

excellent screen; the subset of cases not passing the screen could then be assessed in greater detail using expert judgment, perhaps guided by a more detailed set of explicit criteria.

## Quality assurance

In my opinion, quality assurance has two components. The first is system design, which includes all the measures that a particular organization, and also society at large, use to safeguard and promote the quality of health care. The second component is monitoring, which is the process by which performance is periodically or continuously reviewed and, when found to be deficient, first modified and then monitored once again. System design and monitoring should be an inseparable, mutually supportive pair. Design brings about rough adjustments in performance; monitoring is responsible for fine tuning.

Whether or not we have much to learn from industry with regard to either of these components of quality assurance, as Berwick and Knapp contend, is still moot. I find almost all of what they describe admirable but also utterly familiar.

I do agree, however, that we have not been able to make as rigorous use of statistical control methods as industry has. Perhaps this is because our material is less tractable. Our work is done by autonomous professionals under conditions of great uncertainty that do not permit easy routinization. We work with people, not things; our objectives are multiple, our products diverse. These reasons, however, could be no more than pretexts for inaction. We still should try.

We can perhaps obtain some sense of direction from beginnings already present in our literature. A little more than 30 years ago, Harvey Wolfe offered a method for estimating acceptable variations in length of stay using five variables (reduced from an original set of 355) to estimate a regression line with upper and lower control limits (at 86.64 percent) on either side. He believed that, using this method, "an entire month's cases could be screened by computer in just a few minutes" (Wolfe, 1967).

One of Williamson's earlier descriptions of the method he later called "health accounting" contains a proposal (to my knowledge, not subsequently repeated) that a review be instituted if performance in groups of cases is worse than a preset standard by a distance exceeding a specified confidence interval (Williamson, 1971). It seems to me that deviations in a favorable direction might also call for an inquiry, because they could indicate inappropriate standards, faulty diagnosis, or treatment when it is not required.

In contrast to Williamson's method of what might be called "control by batches," the method proposed by Mushlin, Appel, and Barr (1978) allows case-by-case control because every patient who does not meet the standard of wellness expected to be achieved by a certain time is, so to speak, "recalled." Needless to say, Williamson's method is capable of modification so that findings for successive small batches are

pooled, prompting intervention as soon as the control limits are transgressed.

The Commission on Professional and Hospital Activities has many years of experience in using abstracts of hospital records to generate periodic tabulations that allow comparison of performance patterns with preformulated standards or with statistical norms of a hospital's own performance, as well as the performance of comparable hospitals (Slee, 1974). One can also find in the Commission's publications numerous examples of readily comprehensible visual displays suitable for conveying information to clinicians and administrators.

Perhaps the most interesting prototype of all is the "Alerter" system, introduced by Bundesen in 1954 to continuously obtain and visually display the cumulative number of newborn deaths in each of Chicago's hospitals in a manner that permitted early identification of unacceptable deviations. When these occurred, there was a prompt investigation by visitors from the health department, apparently leading to reforms that caused rapid, remarkable improvement in hospital performance (Bundenen, 1955).

In the area of quality assurance in the health profession, I seem to be pleading, yet once more, for a rediscovery and revivification of our own past.

## Role of consumers

Consumers play a variety of roles in quality assessment and monitoring. Most fundamentally, by expressing their preferences, they supply the valuations needed to choose among alternative strategies of care. Thus, they help define the meaning of quality in the technical sense. Moreover, their preferences (of course, subject to social legitimacy) are the paramount consideration in defining the quality of the interpersonal process and of the amenities of care.

The fact that consumer participation is, in most cases, necessary for the success of health care leads to two kinds of judgment on the quality of care. One is concerned with judging practitioners based on what they control or influence. The other, more inclusive, judgment takes into account the client's participation as well. It is this latter that determines success or failure and, therefore, is more closely reflected by the outcomes of care.

Consumers are also valuable, even indispensable, sources of information in judging the quality of care. Some data (mainly, but not exclusively, about the nontechnical aspects of care) are most easily obtained from consumers. Consumers can also verify, or fail to confirm, the practitioner's reports or perceptions of care. Most importantly, consumers can and do, through expressing satisfaction or dissatisfaction, pass a judgment about many aspects of the process of care and its outcomes.

Finally, consumers, if appropriately informed, could help regulate the quality of care by means of their choices. This is obviously true for the nontechnical aspects of care. Many of the simpler

attributes of technical care and of its purveyors are also comprehensible to consumers. Consumers could certainly understand and use a system for certifying and grading providers of care if this were available.

Recently, there has been much controversy about releasing information on case fatality rates for individual hospitals, a subject briefly debated in these articles as well. In my opinion, there are now two obstacles to the general release of this information. The first is that of insufficient particularization, especially to individual physicians, a degree of detail needed by consumers who are attached to physicians rather than to hospitals. The second, more important obstacle is the insufficient accuracy of the information as an indicator of quality. It seems to me that, at present, information on case fatality is best used by the peer review organizations (PRO's) and, as Brook (symposium) suggests, the hospitals themselves. If, however, subsequent investigation shows the judgment to be valid, I would favor release of the information (accompanied by careful monitoring of the consequences) unless the hospital has made the necessary reforms.

As Codman contended so many years ago, the public is entitled to know what kind of care it can expect to receive for the money it spends. However, when Codman announced "a hundred dollar hospital with a hundred dollar surgeon," the product he offered for sale was not a record of mortality or morbidity but considered judgments on end results after meticulous analysis of the care itself (Codman, 1917).

## **Prevalence of levels of quality**

Despite the large number of studies in which the quality of health care has been assessed over so many years, we are unable to construct a representative picture of quality nationwide or to speak confidently of secular trends. This is because our studies of quality have been partial, highly localized, of short duration, and noncomparable in their methods. The work of Payne et al. (1976) in Hawaii perhaps comes closest to giving us at least a partial view of the quality of care received by a total population in its natural setting. The study proposed by Bates and Connors, by yielding information about care for Medicare beneficiaries in a variety of settings and in several locations across the United States, would be a welcome addition, even though the persons under study do not necessarily represent all ages.

Conclusions about secular trends in quality would be difficult even if the data were available. Judgments on the process of care are bedeviled by the constant evolution of criteria and standards. We could, of course, decide that performance in each period is to be judged by the criteria and standards of its time. Quality would then be defined as concordance of behavior to professional expectations, which is the definition of quality most often used. We could, alternatively, use today's criteria and standards to

judge past performance—say, through a review of old medical records. This would invariably show an improvement, but the significance of the improvement could be assessed only by a comparison of outcomes.

Judging by outcomes, seemingly a more valid approach, has its own problems. Two things are happening simultaneously with the passage of time. One is a change in the science and technology of health care, and the other is a possible change in the extent to which the best knowledge and methods are being applied. It is reasonable to combine the two, but we ought to realize that the resulting definition of quality is quite different from the one that is concerned merely with conformity to the best available knowledge.

The benefits of improvements in technology and its application can be concealed by certain paradoxical effects. For example, a meticulous restriction of certain interventions to only those who need them may increase case fatality through the elimination of unnecessary interventions that are less likely to be fatal. To avoid this trap, Lembcke (1956) has proposed a case fatality rate that would, in effect, omit unjustified surgical interventions from the denominator. Measurement of deaths per 1,000 beneficiaries (a rate similar to one proposed by Eggers) may achieve the same purpose, that of adjusting for the incidence of true need for the intervention.

Improvements in technology may permit us to intervene on behalf of much sicker persons who, in the past, would not have been suitable for such treatment. Yet, this new category of patients may experience an unusually high fatality rate, a rate that is tolerable only because nonintervention is even more likely to be fatal. Another possibility is that a higher risk of fatality in the short term is being incurred in order to attain more than offsetting future gains in survival, in the quality of life, or in both. Not to fall victim to these paradoxes requires study of those who are subjected to a particular intervention and those who, although having the same condition, are not. It could also require lengthier followup, with attention to both survival and the quality of life.

I imagine I see, in the fascinating findings reported by Roos and his coworkers, some hints of the interplay of the phenomena I have described. Changes have occurred in technology (abdominal versus vaginal hysterectomy, transurethral versus "open" prostatectomy, exploration versus nonexploration of the common bile duct); the age structure of the patients has shifted, generally toward an older group; and seemingly surgery is more often in more skilled, more experienced hands. I do not know how these several factors interact, but some patterns can be observed. For both hysterectomies and prostatectomies, there is less frequent recourse to a more hazardous procedure; for both, fatalities are reduced for all but the oldest patients. The results for cholecystectomy are aberrant, perhaps partly because of the more frequent resort to bile duct exploration, a distinctly more hazardous procedure. Whether or not

subsequent benefits compensate for the added hazard, I cannot say.

Although Roos and his coworkers are distinctly more skeptical than I in interpreting their findings, the data for case fatality, coupled with the reduction in readmissions related to the initial hospital stay, are consistent with an improvement in quality. I have already commented on the difficulty of saying whether or not the improvements are bought at too high a price. I should like, however, to make an additional comment.

In assessing trends in the prevalence of high quality, it would be important also to examine its distribution among subgroups of the population. If improving quality for certain subgroups is unusually costly, the gains in equity should be added to total gains in quality when a comparison is made with costs.

## Effects of prospective payment

Perhaps one should begin by noting that no system of reimbursement, no matter how elegantly designed, can precisely match use of services to need. It will err in favoring either overuse or underuse. Assuming responsibility for a defined population under prepayment could be a possible exception, because the temptation to skimp on services might be counteracted by a realization that any consequent deterioration in health would later place the organization at greater risk. I doubt, however, that this is an effective safeguard. It is neutralized partly by the rapid turnover in enrollment that we know occurs and also by the rather short perspective that guides most of our actions. If enrollment were stable and the perspective farsighted, the nightmare of continuing responsibility for a progressively aging population would be enough to drive an administrator mad unless there were assurance of correspondingly rising capitations.

To fine tune use of service to need, system design (including the method of reimbursement as a feature of that design) has to be accompanied by effective monitoring. That is what we have now, assuming the PRO's are effective.

To study the effect of quality of prospective payment, with or without the added intervention of the PRO program and other inhospital monitoring activities, it is necessary to begin with a conceptual model from which the likely consequences can be systematically derived. Only then can we direct our attention to the full range of phenomena that need to be studied.

Unfortunately, of the articles in this issue, only that by Gaumer, Poggio, and Sennett offers a theoretical model of hospital behavior. I applaud its presence, although, not being an economist, I am unable to judge it, nor does my time allow me to offer an alternative formulation. I would like, however, to draw attention to two admirable monographs that can serve as guides (Lohr et al., 1985; Hammons, Brook, and Newhouse, 1986). The first of these, in particular, demonstrates how widely one must throw one's net in order to capture not only immediate direct effects but

also those dispersed consequences that would become apparent in the more distant future as the system of health care undergoes adaptive changes. I am concerned, especially, that hospitals, by dropping useful programs that do not produce surplus revenue or by refusing to embark on such ventures, will adversely affect the health of vulnerable population groups. This means not only that hospitals may be able to shift costs but diminutions in quality as well. The Medicare program, if it confines its attention only to how its own beneficiaries fare, could do a serious disservice to health care as a whole.

As to the very immediate, totally Medicare-centered assessments of quality, I have already commented on the severe limitations of using case fatality and readmissions as measures of quality. However, if we confine our attention to these rough measures and accept the provisional nature of conclusions based on data for so short a period, we arrive at a verdict that is closer to "not proven" than to either "guilty" or "not guilty."

Let us first see what our contributing authors conclude. According to Eggers, "Early results from utilization and mortality statistics do not suggest . . . problems of access to inpatient care or . . . increases in mortality." Gaumer, Poggio, and Sennett, based on data that precede the PRO's (but not the professional standards review organizations), are a little more willing to assign blame. "Based on these findings," they say, "we conclude that PR programs may be increasing elective surgical mortality."

A pessimist, looking more closely at the findings, would easily find cause for disquiet. First, one observes evidence that prospective payment has been associated with lower admission rates and shorter lengths of stay, even though case-mix severity appears to have increased. Does this pattern suggest merely greater efficiency or lower quality as well? The findings reported by Eggers suggest the former.

For the set of "elective" surgical procedures studied, Gaumer and colleagues report an excess fatality of only 0.9 death per 1,000 patients at 15 days following admission, but this relative excess increases progressively as the period of observation lengthens until the excess attains a level of 5.5 deaths per 1,000 patients. This progression may indeed mean, as the authors suggest, the inability of their method of case-mix adjustment to account completely for the greater severity of cases in hospitals subject to prospective payment. Might it not also mean that care has been less complete? The need for further study is obvious and, indeed, accepted by all.

## Research agenda

Fortunately for everyone, the Health Care Financing Administration has taken on with remarkable skill and vigor the task of answering many of the questions concerning the impact of prospective payment. The long list of research projects that Eggers enumerates and briefly describes constitutes an eloquent testimonial to its sense of responsibility and

purpose. In time, we shall know not only how Medicare patients fare but also, through some comparisons with others (for example, in the studies that use data from the Commission on Professional and Hospital Activities) what more pervasive changes in hospital care have taken place.

During a recent conversation with visitors from abroad, I pointed out with some pride how much research is being done in this country to elucidate the consequences of per-case payment for hospital services covered by Medicare. My listeners were amazed, for in their country, they said, there was a long history of governmental responsibility for health care, with frequent changes in policy, unaccompanied by evaluation. I explained that, in our case, the pressures of public opinion, the workings of our democratic institutions, our tradition of accountability, our sense of fairness, the rich resources of our research establishment, and our inclination to self-doubt and self-examination have combined to lead us down a happier path.

Perhaps the time will come when every piece of proposed legislation will be seen as a social experiment pursuant to hypotheses that require verification. Plans for legislation and plans for evaluation would then proceed hand in hand. How confidently would we then stride into the brighter future that we all so ardently desire.

## References

- Aday, L. A., and Andersen, R.: A framework for the study of access to medical care. *Health Services Research* 9(3):208-220, Fall 1974.
- Barnett, O. G.: The application of computer-based medical-record systems in ambulatory practice. *New England Journal of Medicine* 310(25):1643-1650, June 21, 1984.
- Brewster, A. C., Karlin, B. G., Hyde, L. A., et al.: MEDISGRPS: A clinically based approach to classifying hospital patients at admission. *Inquiry* 22(4):377-387, Winter 1985.
- Brook, R.H.: *Quality of Care Assessment: A Comparison of Five Methods of Peer Review*. DHEW Pub. No. HRA-74-31,000. Bureau of Health Services Research and Evaluation, Public Health Service. Rockville, Md. Health Resources Administration, July 1973.
- Brook, R. H., Davies-Avery, A., Greenfield, S., et al.: Assessing the quality of care using outcome measures: An overview of the method. *Medical Care* 15(9) Supplement: 1-165, Sept. 1977.
- Bundesen, H. N.: Effective reduction of needless hebdomadal deaths in hospitals. *Journal of the American Medical Association* 157(16):1384-1399, Apr. 16, 1955.
- Codman, E. A.: *A Study in Hospital Efficiency*. Boston. Thomas Todd Co., Printers, circa 1917.
- Donabedian, A.: Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly* 44(3, Part 2):166-206, July 1966.
- Donabedian, A.: *A Guide to Medical Care Administration, Volume II, Medical Care Appraisal—Quality and Utilization*. Washington, D.C. American Public Health Association, 1969.
- Donabedian, A.: *Explorations in Quality Assessment and Monitoring, Volume I, The Definition of Quality and Approaches to Its Assessment*. Ann Arbor, Mich. Health Administration Press, 1980.
- Donabedian, A.: *Explorations in Quality Assessment and Monitoring, Volume II, The Criteria and Standards of Quality*. Ann Arbor, Mich. Health Administration Press, 1982.
- Donabedian, A.: The quality of care in a health maintenance organization: A personal view. *Inquiry* 20(3):218-222, Fall 1983.
- Donabedian, A.: The Price of Quality and the Perplexities of Care. The 1986 Michael M. Davis Lecture. Chicago. The Center for Health Administration Studies, University of Chicago, 1986.
- Donabedian, A., Wheeler, J. R. C., and Wyszewianski, L.: Quality, cost, and health: An integrative model. *Medical Care* 20(10):975-992, Oct. 1982.
- Georgopoulos, B. S., and Mann, F. C.: *The Community General Hospital*. New York. The Macmillan Co., 1962.
- Hammons, G. T., Brook, R. H., and Newhouse, J. P.: *Selected Alternatives for Paying Physicians Under the Medicare Program: Effects on the Quality of Care*. Santa Monica, Calif. The Rand Corporation, June 1986.
- Krakauer, H.: Outcomes of In-Hospital Care of Medicare Patients in 1983-1985: The Medicare Experience. Draft Report. Office of Medical Review, Health Standards and Quality Bureau, Health Care Financing Administration. Baltimore, Md. 1987.
- Lembcke, P. A.: Medical auditing by scientific methods. *Journal of the American Medical Association* 162(17):646-655, Oct. 13, 1956.
- Lohr, K. N., Brook, R. H., Goldberg, G. A., et al.: *Impact of Medicare Prospective Payment on the Quality of Medical Care: A Research Agenda*. Santa Monica, Calif. The Rand Corporation, Mar. 1985.
- McDonald, C. J.: Protocol-based computer reminders, the quality of care and the non-perfectability of man. *New England Journal of Medicine* 295(24):1351-1355, Dec. 9, 1970.
- McNeil, B. J., Weichselbaum, R., and Pauker, S. G.: Fallacy of the five-year survival in lung cancer. *New England Journal of Medicine* 299(25):1397-1401, Dec. 21, 1978.
- McNeil, B. J., Weichselbaum, R., and Pauker, S. G.: Speech and survival: Tradeoffs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine* 305(17): 982-987, Oct. 22, 1981.
- Moses, L. E., and Mosteller, F.: Institutional differences in postoperative death rates: Commentary on some of the findings of the National Halothane Study. *Journal of the American Medical Association* 203(7):492-494, Feb. 12, 1968.
- Mushlin, A. I., Appel, F. A., and Barr, D. M.: Quality assurance in primary care: A strategy based on outcome assessment. *Journal of Community Health* 3(4):292-305, Summer 1978.
- Payne, B. C., Lyons, T. E., Dewarshius, L., et al.: Episode of Illness Study. In their *The Quality of Medical Care: Evaluation and Improvement*. Chicago. Hospital Research and Education Trust, 1976.

Roos, L. L., Cageorge, S. M., Austen, E., and Lohr, K. N.: Using computers to identify complications after surgery. *American Journal of Public Health* 75(11):1288-1295, Nov. 1985.

Scott, W. R., Flood, A. B., and Ewy, W.: Organizational determinants of services, quality and cost of care in hospitals. *Milbank Memorial Fund Quarterly/Health and Society* 57(2):234-264, Spring 1979.

Scott, W. R., Forrest, W. H., Jr., and Brown, B. W.: Hospital structure and postoperative mortality and morbidity. In Shortell, S. M., and Brown, M., eds. *Organizational Research in Hospitals*. Chicago. Blue Cross Association, 1976.

Shortell, S. M., Becker, S. W., and Neuhauser, D.: The effects of managerial practices on hospital efficiency and quality of care. In Shortell, S. M., and Brown, M., eds. *Organizational Research in Hospitals*. Chicago. Blue Cross Association, 1976.

Slee, V. N.: PSRO and the hospital's quality control. *Annals of Internal Medicine* 81(1):97-106, July 1974.

Williamson, J. W.: Evaluating quality of patient care: A strategy relating outcome and process assessment. *Journal of the American Medical Association* 218(4):564-569, Oct. 25, 1971.

Williamson, J. W., Alexander, M., and Miller, G. E.: Priorities in patient-care research and continuing education. *Journal of the American Medical Association* 204(4):303-308, Apr. 22, 1968.

Wolfe, H.: A computerized screening device for selecting cases for utilization review. *Medical Care* 5(1):44-51, Jan.-Feb. 1967.