

QUALITY ASSESSMENT AND MONITORING

Retrospect and Prospect

AVEDIS DONABEDIAN

*School of Public Health
University of Michigan*

A commentary on the papers in the two special issues on Quality Assurance which shows how they contribute to some important conceptual, methodological, and policy issues. These issues pertain to the definition of quality; the distinction between quality assessment and program evaluation; the role of the market in regulating the quality of care; the role of more direct consumer participation in defining and assessing quality; professional responsibility for quality; the applicability of the structure-process-outcome paradigm to quality assessment; the format, methods of formulation, and validity of the criteria; probability sampling and purposive selection of the topics to be assessed; problems of scaling and measurement; the applicability of industrial control methods to quality monitoring; bringing about behavior change in response to the findings of quality monitoring; measuring the costs and benefits of quality assessment methods, including a consideration of their screening efficiency; and the relationship between quality monitoring and cost containment through competition or by other means.

AUTHOR'S NOTE: My more recent work on quality assessment and monitoring has been supported by the National Center for Health Services Research (Grant 1-RO1-HS-02081), by the Commonwealth Fund, and by the W. K. Kellogg Foundation. While I acknowledge my indebtedness to these generous sponsors, I must emphasize that in this article, as in all my published work, I speak only for myself.

EVALUATION & THE HEALTH PROFESSIONS, Vol. 6 No. 3, September 1983 363-375
© 1983 Sage Publications, Inc.

What a privilege and pleasure it has been to read the succession of stimulating articles assembled in these two issues of *Evaluation & the Health Professions*, and to have an opportunity to share my impressions with their other readers. My purpose is to offer an appreciation rather than a critique, to share with you the thoughts these studies most immediately evoked, and in so doing to reveal perhaps more about my own biases and preoccupations than about the papers themselves. So much the better if while I do that, I am also able to identify perennial concerns, recurring themes, and still-beckoning opportunities.

Naturally, I have had to impose some kind of order on my initially rambling thoughts and, from time to time, to indicate how they came about. But my references to particular papers are merely meant to show how the thoughts were occasioned and, generally, to acknowledge indebtedness, rather than to signal disagreement. If I sometimes fail in the attempt, it may be partly because several of the articles I read were only drafts of their final versions, and some were available only as abstracts. In any event, the attention I give to the several papers has little to say about their relative merits, or even about their relative contributions to my commentary.

Much of the diversity in our approaches to quality assessment and assurance arises from different conceptions of quality itself. In fact, as Baker points out, there are, many definitions of quality and within the domain of one definition a large number of more detailed specifications. I believe, however, that there is a framework that encompasses all the definitions and specifications; a formulation that explains how these variants arise, how they relate to each other, and what their consequences to measurement, monitoring, and overall policy are. It is a simple formulation that defines quality as an approved or preferred relationship between means and ends. The means are essentially the strategies of care; the ends are the changes in "health" brought about by these strategies.

What is included under the rubric of "health," and how the changes brought about in it are valued, depend on who does the defining and the valuation: the health care practitioner, the client, or some instrumentality of the social purpose. As to the means, the differences in the definition of quality arise from different views of the legitimacy of the means and from the partition of the cost of care between the individual patient and society at large. As to the relationship between means and ends, the definitional differences arise mainly from the way in which monetary cost enters into the comparison: whether cost matters not at

all, whether it enters only as a requirement of parsimonious clinical practice, or whether it also includes the trade-off between benefits and costs.

While various combinations of these different classificatory elements can lead to several corresponding definitions of quality, I believe that only two defensible definitions emerge. One requires that clients be active participants with the practitioner in balancing the costs to themselves against the benefits to themselves, as they value the costs and the benefits. This is an essentially "individualized" definition of quality. The other definition, which can be called "social," recognizes that society must, in the end, value all the costs and benefits, including the costs and benefits to those other than the patient, and do so with due regard to some principle of equity (Donabedian, 1980; Donabedian et al., 1982).

An important component among the many differences between quality assessment and program evaluation that Baker explicates is the applicability of the individualized definition of quality to the former, and of the social definition to the latter. It is not surprising, therefore, that the two assessments may lead to different conclusions. More important, however, is that the tension—even conflict—between these two legitimate definitions needs to be recognized and handled (Donabedian, 1983). As to the myriad specifications of the individualized definition of quality, these express what Hemenway, after the manner of economists, calls "tastes." It is both legitimate and necessary that these differences be recognized. But any differences that reflect what Getzen calls the "asymmetry of information" between client and practitioner must be rectified. Other differences in valuation that arise from differences among people in their ability to pay or in their access to information should at the very least cause the gravest social concern. I would be hard put to accept as quality care anything scaled down to patients' ability to pay, when for the more fortunate a higher standard prevails.

These considerations are, of course, directly pertinent to the nature and locus of the responsibility for promoting and safeguarding the quality of care. In this regard, the free market remains a forever beckoning ideal. Cast in utopian splendor, it tugs constantly at our hearts and minds as some long-lost Eden we hope someday to find again. While we are doomed to disappointment, we should not conclude that the market, in its much less perfect everyday manifestations, does not have a role to play. The individualized definition of quality demands, in fact, the exercise of personal choice, provided the opportunity for choice is offered and the choice is informed. But even if

perfect competition were a reality, quality monitoring, far from becoming irrelevant, would yield the information that producers need to operate efficiently and the purchasers also to regulate the market through their choices. To the extent that the market falls short of its idealized images, however, it must be reformed, supplemented, or replaced.

Getzen reminds us how successful consumers can be in bringing about those changes in health care that they understand and desire, while being ineffectual in influencing the more unfamiliar aspects of technical care. Of course, we should not assume that consumers, simply because they choose on the basis of convenience and congeniality, are unmindful of the technical quality of care; they are simply less able to judge the latter and may have despaired of getting the necessary information. Partly, consumers must rely on a trusted medical advisor. A more determinedly independent stance is nevertheless also necessary. How refreshing, therefore, to encounter Rodale's sponsorship of a "People's Medical Society," and his insistence that the considerable information about quality that we have, and which we now so jealously conceal, be made more accessible to the public. The PSROs, for example, have accumulated a considerable fund of hospital and physician "profiles." Hospitals, nursing homes, and other institutions undergo repeated visitations for accreditation, certification, and licensure. Should the findings of all these be disclosed? Should a prospective client at least know how many operations of a certain kind a surgeon or an institution has performed in the recent past? None of this necessarily means that some practitioners and institutions should prosper while the others perish. Rather, the goal would be to match the need for care to the qualifications of the provider, so that each can make the appropriate contribution to the general welfare.

The selection of a source of care is only the beginning of consumers' participation in influencing the quality of care. Consumers must know the available strategies of care and their consequences so as to choose the course of action that is most in keeping with their means and aspirations. Whether the chosen strategy succeeds or fails also depends on patients' cooperation in its implementation. By revealing their satisfactions and dissatisfactions, moreover, consumers can contribute to a formulation of the relevant dimensions of quality, as well as to their assessment. They can also contribute to the professional assessments of quality by providing information on the process and outcomes of care. For all these reasons, it has been suggested that patients ought to see their medical records regularly and, perhaps, be able to add their own comments to it (Kelman, 1976; Marshall, 1977; Westin, 1976; Shenkin and Warner, 1973; Steven et al., 1977).

It is difficult to oppose the informed, active participation of a patient in his or her own health care. Some of the objections to granting the consumer a much more active role in quality assessment and monitoring cannot be easily dismissed, however. But one can insist that if consumers are not to have all the information they need, and if the market, for that reason or another, fails to regulate the quality of care, some other means must be found to do so. Society in general, and the health care professions in particular, must then shoulder more of the burden. It is timely and encouraging, therefore, to be reminded by Dr. Affeldt of the distinguished past record and the praiseworthy current efforts of the Joint Commission on Accreditation of Hospitals. The JCAH understood very early that good system design is the bedrock on which quality rests, but that the ongoing review of health care activities is a necessary additional safeguard. Accordingly, the commission now requires each institution to develop a method of monitoring suited to its own circumstances, so that one can have evidence of performance as well as assurance that successful corrective action has been taken, if required. Professional concern and leadership are particularly needed to guard against the possibly hurtful effects of cost-containment and to promote the higher standards of quality that necessarily cost more money to achieve. To champion quality is the distinctive role and the sacred obligation of the health care professions.

Having defined quality and discussed the responsibility for it, our attention shifts to the more operational issues of assessment and monitoring. Demlo, in her remarkably comprehensive article leads us through the length and breadth of "qualityland," while several other papers give us more detailed views of particular methods. Along the way, one encounters recurrent themes that deserve some comment in this summary. I am impressed, for example, by the remarkable vitality and durability of the structure-process-outcome paradigm, as well as by the many misunderstandings it seems capable of producing. Structure, process, and outcome are not, of course, definitions of quality; they are simply approaches to assessment that can be adopted irrespective of which definition of quality one accepts. Because of these approaches has characteristics that can be interpreted as strengths and weaknesses in particular circumstances, I hesitate to say that any one is superior in all instances, or in most. I believe, in particular, that outcomes are no more valid a measurement of quality than is process, since validity resides not in the outcomes or processes themselves but in the causal linkages between outcome and process. Because outcome and process are in so many ways mirror images of each other, I have found it difficult, as I

have shown elsewhere, to say something about the one that does not have an analogue in the other (Donabedian, 1980). It is not correct to say, for example, that outcome standards encourage parsimonious care whereas process standards encourage redundancy. Outcome standards may encourage efficiency under some special circumstances; under others, the pursuit of outcomes without regard to the means can produce an unbridled outpouring of excess that may be only inefficient, or both inefficient and self-defeating.

Having encountered the structure-process-outcome paradigm used in some form or another by many investigators in many different situations, I must conclude that there is something sound or satisfying about it. I hope, for example, that I am not simply indulging a personal whim when I say that in Hemenway's¹ formulation, outcomes would correspond to outputs, process to actual inputs, and structure to potential inputs. His "market variables" and "survey results" I would consider to be not approaches to quality assessment but methods of obtaining objective or subjective data about the three basic approaches. But is not important that my reformulation either in this instance or others be correct; structure, process, and outcome are the servants of quality assessment, not its masters.²

More specific aspects of method surface repeatedly in these articles. The formulation of criteria occupies a central position among these aspects, since the criteria and standards are the means by which the abstractions embodied in the definitions of quality are translated to more concrete measurement tools. Ryge and DeVincenzi show us that the more specific objectives of certain aspects of dental care are amenable to the formulation of explicit criteria, and that these in turn lead to high levels of reliability in judgments. By contrast, as Cohen and Stricker suggest, when the clinical material is more varied, and not easily classifiable into homogeneous categories, explicit criteria can at best serve only as a screening device; more definitive judgment requires an assessment using implicit criteria as well. While the use of implicit criteria may decrease reliability, it is hoped that validity is improved. But because validity is difficult to substantiate, the reviewers should always be asked to give reasons for their adverse judgements.³

The validity of the criteria is, of course, a matter of the greatest importance. When explicit criteria are formulated, validity depends to a large extent on the format of the criteria, on the choice of the panel charged with their formulation, and on the method for reaching consensus (Donabedian, 1982). When implicit criteria are used, the choice of reviewers is the key determining factor. Cohen and Stricker

show us how the professional background of reviewers influences their judgments. Explicit criteria have also been shown to bear the distinctive imprints of those who formulate them (Thompson and Osborne, 1974; Wagner et al., 1976). Command over the criteria confers considerable power, influence, and responsibility on those who exercise it.

The use of implicit criteria is particularly suited to the assessment of representative samples of care, provided one can assemble the necessary range of expert reviewers. By contrast, the decision to use explicit criteria implies the necessity of confining the assessment to a preselected set of "referents," by which I mean the things to which the criteria pertain. There is, moreover, a close relationship between the precision with which the referent can be defined and the specificity of the criteria. Subsequently, as cases are reviewed, their precise matching to the referent for which the criteria were originally developed becomes critical to the validity of the judgments obtained (Donabedian, 1982).

Schemes for random sampling are, of course, necessary if one wishes to obtain an accurate view of how an institution or a system performs. But a more purposive method of selecting cases is required if one wishes to identify efficiently the cases that are most likely to have been mismanaged. Though almost all who have used explicit criteria for research have told us what guided them in their choice of conditions to study there is no truly comprehensive conceptual model that one might use as a guide.⁴ Such a model would include, for example, the finding by Thompson et al., in this series, that the cost of assessment varies according to the condition chosen.

The construction of quantitative measures of quality is another unsolved problem. A precisely graded measure may not be necessary if we only need to identify cases that are probably mismanaged, so as to study them in greater detail. In some cases, as Ryge and DeVincenzi point out, the presence of one major defect is sufficient to reach a conclusion—in this case, that a dental restoration needs to be replaced. But for research purposes, and some monitoring applications as well, we do need a quantitative scale that takes into account the different weights to be assigned to the several elements of care, and to the interrelationships among these elements.

The pertinence of these issues of sampling and measurement to administrative monitoring (as distinct from research) is evident in the article by Knapp and Miller on statistical controls.⁵ But as I see it, the more interesting questions prompted by this article are normative and conceptual. One must decide, for example, whether average performance or some other standard is to serve as the norm of practice. Even more

interesting is the meaning of the upper and lower control limits. It is reasonably clear what falling below the standard means. But when the quality of care is better than expected, it is not clear whether the standard should be raised, the quality of care reduced, the price of care increased to correspond to the higher quality, the relationship between cost and quality examined to see whether the added quality is worth the corresponding increment in cost, or whether the entire process should be reviewed to make sure that one is measuring what one set out to measure.⁶ In any event deviations beyond the control limits would call for further investigation.

It is a long-established principle that any aberrant finding concerning the quality of care should stimulate efforts to understand and correct the fault.⁷ But acting successfully in obedience to this principle has been far from easy (Donabedian, 1969). In this half of the special issue, Donaldson and Keith propose a way of structuring the internal monitoring apparatus so as to improve its ability to bring about institutional change. Sanazaro carries the theme one step further by discussing why physicians perform at less than optimal levels, and by suggesting methods that might influence performance. The latter include "adequate resources," "a professional environment conducive to a high level of performance," "detailed specification of expected performance," and various "direct interventions." Quality monitoring, by identifying pervasive problems as well as individual failures, obviously plays a central role in any effort to bring about system reform and behavioral change. We know something, therefore, about the intricate machinery for accomplishing our objectives. What remains is to mobilize the system of incentives that, by fueling the machine, provides the energy to make it run.

At the broadest level, these incentives arise out of ethical norms, market forces, and legally enforced obligations. Closer to the delivery of care, organizational commitment derives partly from the foregoing influences as they act upon health care institutions, and partly from the educational, service, and research activities of these institutions. But these activities are themselves driven by competition in another kind of market, one whose rewards are not only recognition and prestige but also all the material riches that flow from these. Finally, individual health practitioners are not likely to alter their behavior unless it is a demonstrable fact that in every way and at every stage their professional and material welfare is dependent on their adherence to the norms that define quality.

The success of quality assessment and assurance in improving performance depends on their linkages to the incentives described

above. When the incentives are there, the simplest or crudest methods of quality monitoring will bear fruit. Without the incentives, the most elegant of quality monitoring mechanisms will fail to work. Given the incentives, however, the design of the quality monitoring mechanism will influence both its effectiveness and its cost. In this series, the article by Thompson et al. provides a method for computing the costs of assessment and monitoring. The measurement of effects can be both easy and difficult. It is easy if one can focus on a discrete behavioral change or a specific manifestation of health status. But if many behavioral changes occur, and these produce corresponding changes in several aspects of health, a single measure cannot be constructed without weighting the several components. Such weights—whether based on individual preferences for the several health states or on the monetary losses attributed to each state—introduce technical and ethical issues of great complexity.

The ability to detect faulty care, while not a measure of effectiveness in itself, is an interim assessment related to both cost and effectiveness. It is, therefore, a useful test of system design. As such, it is influenced by the sampling scheme that identifies the cases to be reviewed in the first place. In some systems of monitoring that are reputedly “outcome-oriented,” substandard outcomes are used not to assess the care but only as a screening device to identify cases that require a more definitive review of process.⁸ In this series, Ryge and DeVincenzi describe a two-tiered method for first identifying dentists whose performance is more likely to be questionable, and then for examining samples of their work, using explicit criteria to judge it. Cohen and Stricker describe a similar strategy for assessing mental health services, one that uses explicit criteria for screening cases and implicit criteria for arriving at more definitive judgments. The set of activities entrusted to the PSROs offers the possibility of more steps. For example, “profile” analysis could lead to the identification of deviant cases, which could then be assessed by explicit criteria, followed by more definitive review if the case fails to pass the screen of explicit criteria.

In all multi-tiered designs, the screening test is subject to two errors, one of passing a case when it should have failed, and another of failing a cases when it should have passed. Often it is possible to calibrate the screening device—for example, by varying the percentage of explicit criteria met that signifies the line of division between pass and fail,—and to select the most efficient cut-off point, screening efficiency being dependent first on the balance of false positives and false negatives, and then on how serious it is to miss a case with poor care as compared to

how costly it is to review cases that are later found not to require it (Greenfield et al., 1982). Similar considerations should guide the choice of when and how often to intervene in implementing a strategy of concurrent review. For each intervention, one needs to know how many cases of inappropriate care are detected, and how many cases are missed. But no matter what screening or case-finding method is used, one also needs to know how often corrective action can be taken and what the effects or benefits of that action are, so that these consequences can be compared to the effort or cost of review (Donabedian and Wyszewianski, 1977; Averill et al., 1977; Dittman and Magee, 1980).

There is one possible consequence of quality monitoring that is rather difficult to detect, unless one makes special plans to do so. This is the "sentinel effect," a change in behavior brought about by the mere expectation that one's work may be reviewed. It is a moot point whether or not monitoring can be so repressive as to inhibit the institution of useful care, the recommendation to have surgery being one example. It has also been suggested that certain kinds of explicit criteria of process not only allow for but may actually encourage redundancies in care. Thus, while we fear, and often find, that quality monitoring is ineffectual, we sometimes claim that it can be too effective, or effective in the wrong way. We obviously have a great deal to learn about the uses and limitations of quality assessment and monitoring.

The application of the general principles of quality assessment and monitoring to the special needs of particular kinds of care is an obvious direction for further growth and experimentation. This series of articles provides excellent examples that demonstrate the universality of the general principles as well as their capacity to adapt to particular situations. In devising a method for assessing dental care, for example, Ryge and DeVincenzi take full advantage of the opportunity to inspect the product so as to judge the details of technical execution, using explicit criteria. By contrast, Cohen and Stricker show how the method of assessment takes into account the ambiguities of diagnosis and treatment in the mental health field, and how the results of assessment reflect these ambiguities (Riedel et al., 1974). Similarly, in their review of the nursing literature, Lang and Clinton demonstrate the generalizability of the structure-process-outcome paradigm while also showing the adaptation of the many measures under each of the three headings to the special features of nursing care. Zimmer's paper on long-term care also emphasizes the need for such adaptations—in this case to the rather harsh realities of the particular situation. I see no contradiction,

however, between conceptual sophistication on the one hand and the crudities of care on the other. On the contrary, it is quality assessment by rote (which Zimmer does *not* propose) that is more likely to prove inapplicable. In my opinion, the assessment of long-term care is precisely the kind of challenge that calls for a more complete, more sophisticated conceptual framework from which the most appropriate method of assessment can derive.

The articles in these special issues, taken as a whole, demonstrate a remarkable commitment to and concern for quality care, even under seemingly adverse circumstances, when so much of our attention is being forced to focus on cost containment. Fortunately, properly understood, cost containment and quality assessment must be an inseparable pair. For example, we cannot make progress in producing care more efficiently unless we can measure the product, and the product cannot be measured properly if its quality is not known. Furthermore, knowledge about quality is essential if certain cost-containment strategies are to succeed without a tragic distortion of their social purposes. Accurate, widespread knowledge of quality is necessary if competition, driven by consumer choice, is to regulate the market properly. Similarly, quality must be monitored assiduously if health care institutions subjected to the constraints of case-based reimbursement are to resist the temptation to pare down quality.

To the extent that care that makes no contribution to health, or has a net harmful effect, can be eliminated, or at least reduced in frequency, quality control and cost control are the closest of allies. There is thus a large element of cost-saving in quality monitoring and control. There is, however, a cost-generating component as well, since improvements in quality may require more and better care, as long as the increments in benefits are worth the additions to cost.

The health care professions, supported by an informed public, must accept stewardship over all aspects of the quality of the care they provide. But the cost-generating component of quality should be their particular concern, orphaned and abandoned as it is in these cold and cruel times.

NOTES

1. David Hemenway's manuscript (entitled "Quality assurance from an economic perspective: a taxonomy of approaches with applications to nursing home care") will appear in the December issue.

2. For a demonstration of how several other alternative formulations correspond to the structure-process-outcome paradigm, see Donabedian (1980: 85-100).

3. Morehead, who is the principal proponent of implicit criteria, has incorporated this feature in her method from its earliest days. See, for example, Morehead et al. (1964). The superior validity of a method that uses implicit criteria is attributable mainly to its ability to adapt more precisely to the special characteristics of any given case by including more information about these characteristics and, consequently, modifying the explicit criteria, or bringing additional criteria to bear on the subject.

4. The two formulations I find most attractive are the "maximum achievable benefit" proposed by Williamson (1978) and the "tracer method" proposed by Kessner et al. (1973). Brauer discusses the selection of case for individual review in Reidel et al. (1974).

5. I have often wondered why the model of industrial statistical control has received so little attention among the methods of quality monitoring. It may be because of the diversity of cases (the material processed) and the corresponding variability of the objectives of care (the product), so that the tendency has been to individualize rather than to aggregate cases into categories. For early examples of interests in variants of this model, see Metzner (1953) and Wolfe (1967). Williamson's method of "health accounting" also has affinities to this approach of being oriented to system assessment (as contrasted to the review of individual cases) and in calling for a review only when the discrepancy between performance and standards exceeds statistically defined limits (Williamson, 1971).

6. These questions do not come up, of course, when the phenomenon being measured is "inflected," by which I mean that both too much and too little are bad. Hospital length of stay is a good example.

7. An important early example is the "bi-cycle" model as described in Brown and Uhl (1970).

8. For example, the health accounting method proposed by Williamson (1971) does this by identifying groups of cases while the "problem status index" described by Mushlin and Appel (1980) identifies individual cases. These investigators also document the screening performance of their method.

REFERENCES

- AVERILL, R. F. and L. F. McMAHON (1977) "A cost benefit analysis of continued stay certification." *Medical Care* 15 (February): 158-173.
- BROWN, C. R. and H.M.S. UHL (1970) "Mandatory continuing education: sense or nonsense?" *J. of the Amer. Medical Assn.* 213 (September 7): 1660-1668.
- DITTMAN, D. A. and R. P. MAGEE (1980) "Optimal investigation policies under selected PSRO procedures." *Medical Care* 18 (October): 1032-1047.
- DONABEDIAN, A. (1983) "Quality, cost, and clinical decisions." *Annals of the Amer. Academy of Political and Social Sci.* 468 (July): 196-204.
- (1982) *Explorations in Quality Assessment and Monitoring, Volume II: The Criteria and Standards of Quality.* Ann Arbor, MI: Health Administration Press.
- (1980) *Explorations in Quality Assessment and Monitoring, Volume I: The Definition of Quality and Approaches to Its Assessment.* Ann Arbor, MI: Health Administration Press.
- (1969) *A Guide to Medical Care Administration, Volume I: Medical Care Appraisal—Quality and Utilization.* Washington, DC. American Public Health Association.

- and L. WYSZEWIANSKI (1977) "The numerology of utilization control revisited: when to recertify." *Inquiry* 14 (March): 96-102.
- DONABEDIAN, A., J.R.C. WHEELER, and L. WYSZEWIANSKI (1982) "Quality, cost and health: an integrative model." *Medical Care* 20 (October): 975-992.
- GREENFIELD, S., S. CRETIN, L. C. WORTHMAN, and F. DOREY (1982) "The use of an ROC curve to express quality of care results." *Medical Decision Making* 2 (Spring): 23-31.
- KELMAN, H. R. (1976) "Evaluation of health care quality by consumers." *Int. J. of Health Services* 6, 3: 431-442.
- KESSNER, D. M., C. E. KALK, and J. SINGER (1973) "Assessing health quality—the case for tracers." *New England J. of Medicine* 288 (January 25): 189-194.
- MARSHALL, C. L. (1977) *Toward an Educated Health Consumer: Mass Communication and Quality in Health Care*. Washington, DC: Government Printing Office.
- METZNER, C. A., S. J. AXELROD, and J. H. SLOSS (1953) "Statistical analysis as a basis for control in fee-for-service plans." *Amer. J. of Public Health* 43 (September): 1162-1170.
- MOREHEAD, M. A. et al. (1964) *A Study of the Quality of Hospital Care Secured by a Sample of Teamster Family Members in New York City*. New York: Columbia University School of Public Health and Administrative Medicine.
- MUSHLIN, A. I. and F. A. APPEL (1980) "Testing an outcome-based quality assurance strategy in primary care." *Medical Care* 18 (May Supplement): 1-100.
- REIDEL, D. C., G. L. TISCHLER, and J. K. MYERS (1974) *Patient Care Evaluation in Mental Health Programs*. Cambridge, MA: Ballinger.
- SHENKIN, B. N. and D. C. WARNER (1973) "Giving the patient his medical record: a proposal to improve the system." *New England J. of Medicine* 289 (September 27): 688-692.
- STEVEN, D. P., R. STAGG, and I. R. MACKEY (1977) "What happens when hospitalized patients see their own records." *Annals of Internal Medicine* 86 (April): 474-477.
- THOMPSON, H. C. and C. E. OSBORNE (1974) "Development of criteria for quality assurance of ambulatory child health care." *Medical Care* 12 (October): 807-827.
- WAGNER, E. H., R. A. GREENBERG, P. B. IMREY, 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.
- WESTIN, A. F. (1976) "Medical records: should patients have access?" *Hastings Center Report* 6 (December): 23-28.
- WILLIAMSON, J. A. (1978) *Assessing and Improving Health Care Outcomes: The Health Accounting Approach to Quality Assessment*. Cambridge, MA: Ballinger.
- WILLIAMSON, J. W. (1971) "Evaluating quality of patient care: a strategy relating outcome and process assessment." *J. of the Amer. Medical Assn.* 218 (October 25): 564-569.
- WOLFE, H. (1967) "A computerized screening device for selecting cases for utilization review." *Medical Care* 5 (January-February) 44-51.