Approaches to the Quality of Hospital Care

By MINDEL C. SHEPS, M.D., M.P.H.

INCREASING ATTENTION is being paid to the problems of improving and of appraising the quality of health services in general (*1–8*) and of hospital care in particular. The general problems of measurement and evaluation in all these areas are similar (*9–16*). Basically, they involve finding valid and reliable measurements of quality and interpreting these measurements when made.

Purposes of Hospital Evaluations

Evaluations of hospital quality may have different purposes. The methods and standards selected must be related to the particular purposes for which they are being applied. The most familiar purpose is regulatory. Such an appraisal is designed to match an institution against specified standards that determine its acceptability for the purposes of the regulatory or accrediting agency. It is intended to correct abuses and raise the general level to an acceptable minimum. Various national bodies, now united in the Joint Commission for the Accreditation of Hospitals, have evolved minimum acceptable levels of facilities, equipment, administrative and professional organization, and professional qualifications. They have also made use of some numerical indexes of organization and performance. All of these have been set forth in the commission's Standards for Hospital Accreditation.

The requirements of licensing boards, health insurance funds, and other organizations serve a similar purpose. Regulatory appraisals set minimum or desirable levels by excluding institutions that fail to qualify. At the same time, this provides protection for patients and students who go to acceptable institutions (*17–20*). As yardsticks, these standards have one division that divides hospitals into two classes only—good enough or not good enough.

Improvement of Quality

The second purpose of appraisals is a closely related one—that of serving as a stimulus for the improvement of quality. Licensing and accrediting assessments serve this purpose as well as that of regulation. Hospitals also make self-appraisals for this purpose. The standards used may be minimum or optimum levels, similar to those used for regulatory purposes, or may provide for comparisons among physicians in the same institution. One of the chief instruments for such appraisals has been the medical audit (*21–25*).

The basis of the audit is a review of hospital records according to such criteria as qualitative judgments of the care given and examination of diagnostic errors (*26–29*), as well as such numerical indexes as mortality rates, rates for the

Dr. Sheps is research associate, department of biostatistics, Harvard School of Public Health, and lecturer on biostatistics, Boston University Medical School. *She presented this paper at a joint session of the Medical Care Section of the American Public Health Association and the American Association of Hospital Consultants at the 82d annual meeting of the APHA in Buffalo, October 14, 1954. A news summary of Dr. Sheps' paper appeared in the February 1955 issue of Public Health Reports, p. 198.*

incidence of specific complications, removal of normal tissues at operations (*30–32*), consultations, cesarean rates, and the rates for certain tests by diagnostic category. The term "professional service accounting" has been suggested for the compilation of these rates and "medical audit" for their evaluation.

Program Evaluation

The third, and more recently recognized, purpose for quality appraisal, is to study the effects of specific programs or procedures on the quality of care. Generally we may refer to this purpose as program evaluation. Certain procedures, such as clinical-pathological conferences, are believed to improve the quality of hospitals. Large-scale complex programs, such as regionalization, are under trial. It is essential that their effectiveness be examined. We need to know in detail the effect of such procedures on the care received by patients. Program evaluation is the attempt to study this effect, by seeing whether a difference or an improvement in quality is associated with a certain procedure or program.

In other situations the purpose of a program evaluation may be to see whether a certain institution is giving "good" care. Such an appraisal is basically an evaluation in which the judge compares what he finds with what in his mind seems desirable and possible.

For regulatory purposes, it is desirable to establish criteria which actually do differentiate between acceptable and unacceptable institutions. They must be discriminating at the level where the regulating agency feels the line must be drawn. To help in improving the quality of hospital care, the criteria used must discover the most important problems and reflect progress in meeting them. Criteria used for these purposes may be useful in program evaluation, but more refined measuring tools are also needed. A scale which only says "good enough" or "not good enough" is inadequate. What is needed is a scale that measures values along a continuum extending from one extreme to the other. The measurements used must be sensitive to those aspects of hospital quality that may be affected by the program.

Reference has been made to some of the standard methods in use. It has recently been stated that the statistics used in judging hospitals are "usually meaningless, often illogical and frequently unscientific" (*33*). For example, postoperative mortality ratios are based on deaths within 10 days divided by the total number of operations performed. Thus, deaths after 10 days are omitted, and on the other hand procedures of varying risks, such as dental extractions and neurosurgery, are indiscriminately lumped together.

Moreover, there is a question regarding the validity of the standards by which some of the indexes are evaluated (*33–36*). A top limit of 3 to 4 percent is set for cesarean sections, but current clinical practice and results justify consideration of a higher level. Many of the standards were derived empirically, and their validity was not adequately established. Progress in clinical practice, in any case, calls for frequent revision and revalidation of the standards. The standards to be used, therefore, should be at levels which move according to changing medical knowledge.

The American College of Surgeons and the Professional Activity Study Group in Michigan are cooperating on a new approach to the audit as a measuring device and on the development of new indexes (*36*). A number of new indexes of hospital or medical care have recently been described (*37–39*). Studies currently in progress in the Rochester Regional Hospital Council, the North Carolina General Practice Study, and the Boston Evaluation Study may produce other indexes.

Other methodological developments of interest have been the application of the time study technique to hospital nursing (*40*) and a statistical analysis of the items included in hospital licensing regulations (*41*).

Problems of Measurement

Hospital care is multidimensional. It is a service provided by a coordinated group of professional, technical, and other workers under the direction of a physician. The quality of the care received by patients is affected by the adequacy of the hospital facilities and their maintenance, by the administrative and professional organization of the hospital, by the competence of the personnel, and by the inter-

personal relations among the staff as well as between the staff and the patients.

Any consideration of evaluation, therefore, must recognize the large number of factors involved in patient care. It has been frequently suggested that an appraisal form for hospital quality be developed. Some have extended this concept to include deriving a final score or number to represent the quality of a given hospital. Such a composite index would obscure important differences. Moreover, it would be impossible to choose the items to be included and the relative weights for them on a basis that was generally applicable. On the other hand, a number of measurements can be made and each of these allowed to stand by itself, thus producing a profile of the hospital (*42*). We do not try to represent the health status of an individual by a single figure such as 90 percent of the optimum, but rather, we say something like:

"This patient's health, in general, is excellent, except for mild obesity and a hemoglobin of 10 grams."

Similarly, would it not be meaningful and helpful for the final report of a study to state not that "the quality is good" but rather something of this sort:

"Differences found in the following indexes were highly significant . . . No differences were observed in . . . The quality of physicians' services was significantly higher . . . The differences in social service were not significant . . ."

An exhaustive discussion of all the aspects of hospital care that could be included in such a profile will not be attempted. Instead it is proposed to deal with some of the methodological problems involved.

The main techniques used in appraisals of hospital quality can be divided into:

- The examination of prerequisites or desiderata for adequate care.
- Indexes of elements of performance.
- Indexes of the effects of care.
- Qualitative clinical evaluations.

By Set Standards of Care

According to the first approach, it is assumed that it is possible to select prerequisites for adequate care and that improvement of these factors leads to improved care. These prerequisites are minimum or optimum levels of facilities, equipment, professional training, and organization. As examples we may mention:

- The provision and appropriate maintenance of adequate physical facilities.
- The existence of special facilities, such as blood bank, bone bank, special laboratory and diagnostic facilities, premature nursery, and artificial kidney.
- An effective organizational structure, both administrative and professional.
- Standards and functioning of service departments such as records, laboratories, and libraries.
- Numbers of personnel by size of hospital—interns, residents, nurses, social workers, physical therapists, nutritionists, technicians.
- The availability of specialized personnel for consultation and of facilities for consultation with others, as in certain regional programs.
- Arrangements for ward rounds, refresher courses, continuing education.
- Minimum qualifications of personnel.
- The existence and functioning of internal controls, such as tissue committee, obstetrical committee, and medical audit.

The use of this approach implies the hypothesis that, given certain facilities and standards, the desirable quality of care is achieved. This hypothesis should be recognized and tested explicitly so that valid criteria can be used in a more informed fashion and to better purpose.

Consideration of the norms used for these desiderata raises such questions as, should they be national averages, regional averages, minimum adequate levels or optimum levels, or, should the search for standards be abandoned and the findings on various hospitals simply compared with each other? The answer depends partly on the purpose of the appraisal. In program evaluation, for example, it may be preferable to use flexible indexes applicable to different types of hospitals and to different periods, rather than to adopt any fixed standard of desirability.

Similar considerations led Stouman and Falk (*10*), in their proposals for international health indexes, to abandon the search for standards. If one has valid measurements for a characteristic, then intelligent, directed appli-

cation of these measurements in some situations will provide useful information as to their variability and their significance without a norm.

The argument will perhaps be made clearer by an example. When we use an index such as weight of children, we need a knowledge of norms to assist in diagnosis of a particular child. However, if we want to test the effect of a certain vitamin on weight, we need to compare the gains made by children receiving the vitamin and those not receiving it. Knowledge of the norms is here irrelevant.

By Elements of Performance

The second approach to quality uses indexes intended to reflect one or more elements of performance. Indexes may be defined as "one or a set of measures . . . used to measure indirectly the incidence of a characteristic that is not directly measurable" (*43*). Patient care is such a characteristic. Its quality cannot be measured quantitatively, except by the arbitrary allotment of a certain number of points to a qualitative judgment. It is difficult to define; it is complex and intangible. It is therefore natural and logical that much of the effort to evaluate this quality has been focused on the development of indexes, for example:

- Utilization rates for specific procedures by category, such as admission chest X-rays as an index of preventive services, or rectal examination in specified groups of patients.
- Utilization rates of certain laboratory and other diagnostic procedures, by category.
- Indexes which would reflect the promptness and discrimination with which new procedures or drugs are used in the hospital.
- Referral rates and patterns.
- Autopsy rates.
- Cesarean rates.
- Pathological reports on surgical specimens.
- Correlations between preoperative and postoperative diagnoses and between ante-mortem and post-mortem diagnoses.
- Accuracy of diagnostic procedures.
- Average length of hospital stay by diagnosis. It may be hypothesized that in a well-organized service where the staff members work with purpose and integration, hospital stay will be shorter, on the average, for certain types of cases.
- Listing specific diagnostic and therapeutic procedures expected for each type of case and matching records against them.

Good indexes are objective, reliable, and valid. By the reliability or precision of a measurement, we mean the degree of agreement among repeated measurements of the same things. Numerical indexes, such as rates or ratios, would appear to be relatively precise—anyone counting the number of autopsies done out of all hospital deaths should get the same rate provided that a hospital death is defined without ambiguity. Reliability is more difficult to achieve in indexes that require measurement rather than counting, as has been shown when different physicians measure the same enlarged livers (*44*), or the same reactions to tuberculin tests (*45*). Agreement on interpretation of X-ray films, even when repeated by the same physician, is far from perfect (*46*). Quality indexes that depend on such assessments or on diagnosis, however made, must be tested for precision. It is an error to assume that because they are numerical they are precise.

The use of these performance indexes depends on an assumption that they are valid for an assessment of hospital quality. Presumably this involves the hypotheses (*a*) that each index does measure an element of patient care, and (*b*) that one or a number of these indexes are highly correlated with the intangibles of care. These hypotheses can and should be examined in the light of clinical information and by statistical analysis.

Again, the desirability and validity of the yardstick for these indexes must be examined. We can all agree that the higher the proportion of correct diagnoses, the better the quality of medicine. But what proportion ought we to expect in 1955 in different hospitals?

Other standards are even less definite. There can be too many laboratory tests as well as too few; too few cesareans as well as too many. Discriminating standards would be based on studies of the relation between good patient care and the resultant number of procedures by category of case. The development of appropriate yardsticks thus involves studies to estimate desirable levels for the various indexes and the expected variation. These levels should be adaptable for different types of hospitals

and capable of reflecting progress in clinical medicine.

By the Effects of Care

A third approach to appraising hospital quality is the use of indexes intended to measure the effects of quality of care on patient health. The outcome of specific therapy is influenced by many factors in addition to that of treatment, factors such as age, sex, nutrition, stage of the disease, and the emotional state of the patient. The use of these indexes may therefore be highly misleading. It necessitates a careful evaluation of concomitant factors and an attempt to control them.

Some instances of such indexes are in fairly common use. They include:

- Specific mortality ratios such as postoperative, puerperal, and neonatal.
- Survival rates of premature infants in a specified weight range.
- The incidence of preventable complications such as postoperative infections.

These indexes would seem to be relatively precise if their basis is carefully defined. A postoperative mortality rate depends only on the careful definition of postoperative deaths and of the types of surgery counted. However, the validity of this rate as an index of the quality of surgery would still need to be established. The definitions used are highly relevant to this validity. But even then questions remain:

Does the death rate within 10 days after major chest surgery really reflect the quality of care?

What about the effect of such patient characteristics as diagnosis, complicating illnesses, age, sex, and nutritional status?

Objective indexes of the kinds discussed therefore require careful definition and evaluation. They hold considerable promise for the appraisal of hospital quality, but they are not yet the ideal measuring tools. At best, they are indirect and partial indicators of the basically intangible characteristic with which we are concerned. In a given organizational setting, with access to given standards of consultation and services, doing a certain number of laboratory examinations, one can have numerous shades in the range of quality of care, depending on the skill, the judgment, the experience and the character of the persons involved and on their relationships with patients. This end product is what one really is trying to evaluate in any assessment of patient care. It is for this reason that some workers have tackled the problem of evaluating quality directly through the fourth approach, namely a clinical evaluation.

By Clinical Evaluations

Makover (*47*) included a clinical evaluation in his study of medical groups associated with the Health Insurance Plan of New York. The continuing HIP evaluation program has adapted some of his techniques as well as adding new ones. According to a preliminary unpublished report by M. A. Morehead, delivered at the 1954 New England regional meeting of the American Public Health Association, records from six clinical fields are selected and scores assigned on the completeness of records, diagnostic management, treatment, and reporting. Although performance is measured against prepared standards, the final score depends on the judgment of the consultant making the evaluation.

More recently, another type of combined appraisal was made by Goldmann and Graham (*48*). To an analysis of the availability of service and utilization of service, they added qualitative ratings of the efficiency of service organization and of a random sample of patient records.

Clinical evaluations are in the end subjective and thus less precise than some of the indexes previously discussed. However, they may be more valid since they are a direct approach to the characteristic we want to evaluate, the quality of patient care. Quality, though intangible, is not an abstraction. Nor is judgment of quality capricious or a purely personal whim. While agreement could not be expected to be complete, there are, in numerous situations, widely accepted concepts of what is meant by good care.

The reliability of qualitative judgments can be tested, and they can be subjected to statistical analysis (*49–51*). To make relatively precise estimates of quality, it is necessary to have the findings of several independent judges. This permits an assessment of the degree of agreement among judges (*52*, *53*); it allows for an

objective test of the reliability of the evaluations and diminishes the effect of individual bias. Although more difficult than assessing the precision of interpretations of objective tests, it is the same in principle.

At times, agreement on judgmental evaluation is better than agreement on definitions and criteria. Reynolds (*44*) describes such a case. A group of clinicians attempting to outline criteria for ambulation of patients with infectious hepatitis could not reach agreement after a lengthy discussion of weighting procedures. But analysis of "paired comparisons" by which each man gave his preferences between the criteria in all possible combinations of two, resulted in good agreement on the ranks to be assigned them. Thus, it may be that several clinicians would agree more readily on the relative ranks of a number of hospitals than they would on the weights to be given various objective indexes in computing a quality score.

Judgmental assessments are resorted to in other fields. In some food industry situations the taste of a product must be evaluated. Objective tests give some information, but to answer the important question, "how does a particular process affect the taste," one must in the end resort to tasting the food. If it is necessary to study the effect of different ingredients and preparation procedures on the palatability of ice cream, the only way out is to ask a number of judges which product tastes better to them.

From such tests, valid results may be obtained notwithstanding the problems related to nonagreement among a group of judges and the lack of consistency on the part of an individual judge. The validity is also dependent on such controls as randomization of the tests and the judge's ignorance of which particular product he is testing.

Accordingly, it would appear that appropriate experimental and statistical procedures would enable more meaningful results to be obtained from qualitative evaluation of hospital care. In general, qualitative judgments are expressed through either ranking or scoring.

In ranking, a number of units are placed by each judge in order of his preference and the various ranks analyzed for consistency.

In scoring, a scale of quality is established, and each judge assigns the score that he considers appropriate.

These techniques may be combined in various ways. Thus, individual scores can be given on quality in different clinical fields, and the findings combined into one overall score. The subjects may then be ranked according to the scores obtained.

As already emphasized, the use of these techniques depends on replication, that is, on securing at least two separate evaluations of the same set of units. This is not the same as asking a committee to make a combined appraisal. Only through separate evaluations is it possible to assess the consistency of the individual judgments and to arrive at a relatively unbiased estimate. The value assigned results from combining the different judgments, and is more reliable and objective than the opinion of a single individual. A committee of experts would emerge with one final appraisal, but this would not allow the internal checking suggested. It is even possible with such a panel that one or two members could influence the others so that the final assessment would not be truly a consensus.

If qualitative appraisals of hospital care are made, the particular aspect of care to be studied might be medical, surgical, obstetrical, nursing, social work, or a combination of these and others. Well-qualified, experienced practitioners in the field under study should make the evaluation. Although most clinicians, medical and other, would probably prefer to base their opinions on actual observations of patient care, a properly selected random sample of clinical records may provide adequate information. If various clinical fields were reviewed, replication would be desirable in each field. The evaluations in the major clinical fields could then be crosschecked to test the consistency among them.

It is obvious that such qualitative judgments would have meaning only as comparisons among different hospitals or subdivisions of hospitals. The rank or score would have no absolute value but merely a relative value within the groups appraised. Direct comparisons are unavoidable when we operate without a scale or yardstick. However, this only makes explicit the fact that comparison lies at the basis of all meas-

urement. Measurement has been defined as "the assignment of numerals to things so as to represent facts and conventions about them . . . under a consistent set of rules" (*54*).

The operations to which any measurements can be subjected depend on the rules that can be made and on the validity and reliability of the values obtained. However, with any measurements we make comparisons. If we measure height, we compare the height of the subject with the markings on an arbitrary scale. These markings have meaning only in terms of established norms, of earlier results on the same subject, or of readings obtained from another subject. When we don't have a ruler, we can stand two subjects next to each other and make the comparison directly. Similarly, any index of quality may be measured against a yardstick, or the values obtained on several hospitals may be directly compared.

The Basis of Comparison

Any of the four approaches to quality, therefore, involve comparisons, either indirectly through the use of standards or directly. In program evaluations the basis of comparison is vital. If a specific procedure has an effect on quality, it must be revealed in differences. To find such differences, the hospital under study must be compared with something, either with other hospitals or with itself before inception of the program. When differences exist, it is of paramount importance to isolate differences related to the program itself from differences owing to such other factors as changes in time, economic differences, cultural differences, and so on. The selection of a basis of comparison or control is crucial in this attempt.

The basis of comparison and the indexes to be used, in fact the plan for evaluation of a new procedure or program, should go hand in hand with planning the program itself. Ideally, before the new procedure is instituted, several similar hospitals would be chosen. By random selection, half would become experimental units for the new program, and half, controls (*55*). A careful study of patient care would be made in all the hospitals before instituting the new program and again at a suitable time after the program was in effect. Medical and social progress occurring during this period might produce changes in both experimental and control units. But the changes might be different, and it is these changes that would be compared.

Such an experiment would be relatively simple if it tested only one procedure at a time. Thus one could, for example, estimate the effect of providing small hospitals with special courses for laboratory technicians by comparing changes in the accuracy of their results in certain procedures. More complex programs should also be amenable to this type of experiment.

However, there are cases where this does not apply. The effects of an existing program are to be evaluted, or a hospital plans to embark on a new program, and a study is to be made of quality before and after. The institutions then are self-selected and thereby are different from other hospitals. The before-and-after case does not make the problem simpler. A comparison within the one hospital at two different periods necessitates control of secular changes related to the passage of time or changes owing to extraneous factors that may have come into play.

This situation has parallels in population studies, in public health research, and in clinical research. Analytical studies of what exists, rather than of a planned experiment, call for a more critical evaluation of the findings, and the conclusions must be more tentative. This is even truer when the analyses are retrospective, being made after the fact (*56–59*). However, in some cases it may be possible to select suitable controls for comparison (*60*) and to test their suitability by an examination of variables that might affect the result. The careful selection and critical analysis of the controls (*61*) are basic to the validity of the findings.

An important step in program evaluation, therefore, is the search for suitable controls and for methods of eliminating some of the many variables that affect the quality of care. This is especially difficult because hospitals are highly complex institutions, and patient care is an intangible quality, influenced by many variables.

It is therefore possible that no comparable

units could be found for nonexperimental situations. In the event of failure to establish acceptable controls, or even as a complement to comparisons made with controls, there is another avenue of approach that might prove fruitful. One might formulate a hypothesis such as:

"Characteristic x is usually found only in teaching hospitals. A random sample of nonteaching hospitals in comparable communities will show low values for x, but an examination of hospitals in our group will reveal significantly higher values."

Characteristic x might be laboratory services of a certain type, or rehabilitation services, or one of the quantitative indexes considered. Such an approach, of course, adopts the rest of the country as controls. As in other situations, purely descriptive studies may be used as a basis for planning future experiments.

Many references have already been made to testing the validity of various indexes. This could be done in various ways. One would be through a statistically controlled clinical analysis of each index. Another would be through seeking correlations among various measurements (*37*, *62*) including qualitative appraisals. Thus, correlations might be sought among two or more measurements which are believed to measure practically the same thing. It might be found that a relatively simple, inexpensive objective test showed a high correlation with the results of the qualitative judgments, or, more likely, that a combination of such tests did so. Once the validity and applicability of such indicators were established, there would be many instances where they could be used instead of more difficult, expensive, and cumbersome techniques.

Conclusions

Most of the work done to date in the appraisal of hospital quality has been related to the purposes of correcting abuses, setting minimum standards, and stimulating improvements in quality. The field of program evaluation is just beginning to be explored.

Techniques used in quality evaluations vary with the purposes of the particular study. The quality of care can be evaluated through a multidimensional approach which results in a profile of the hospital. The main basis of the appraisal can be the use of one or a combination of: examinations for prerequisites for good care, indexes intended to measure elements of performance, indexes intended to measure the effect of care by results obtained, and qualitative clinical judgments.

Any indexes and standards used in such appraisals would be clearly defined, based on comparable data, and examined for their reliability and validity. Qualitative clinical appraisals should also be tested in a similar fashion and statistical controls and analysis applied to them as well.

Correlations among different indexes and judgments used should be attempted.

Appraisals which are intended to examine the effects of specific procedures or programs should be planned before the inception of the program.

The selection of an appropriate basis of comparison is crucial to program evaluation.

The development of practical and valid methods of measurement will involve the expenditure of considerable money and time. However, in view of efforts and money now being spent on programs to raise quality, it would seem essential to direct some of those resources toward the development of appropriate methods with which to judge their effects. Collaboration of clinicians, administrators, and statisticians is necessary for such a development.

A critical analysis of the particular methods used should be an explicit objective of a quality evaluation.

REFERENCES

(*1*) Reed, L. S., and Clark, D. A.: Appraising public medical services. Am. J. Pub. Health 31: 421–430, May 1941.

(*2*) Goldmann, F.: The adequacy of medical care. Yale J. Biol. & Med. 19: 681–688, March 1947.

(*3*) The quality of medical care in a national health program. A statement of the Subcommittee on Medical Care, American Public Health Association. Am. J. Pub. Health 39: 898–924, July 1949.

(*4*) American Public Health Association. Subcommittee on Medical Care: The quality of medical care: Concepts, standards, methods of evaluation and improvement. Bibliography No. 9. New York, The Association, 1954. Processed.

(5) Weinerman, E. R.: Appraisal of medical care programs. Am. J. Pub. Health 40: 1129–1134, September 1950.
(6) Improving the quality of medical care. Dublin, T. D.: The training of personnel; Clark, D. A.: Group medical practice; Kaiser, A. D.: Regionalization of hospitals; Daily, E. F.: Sound principles of administration. Am. J. Pub. Health 39: 314–339, March 1949.
(7) Lee, R. I., and Jones, L. W.: The fundamentals of good medical care. Chicago, University of Chicago Press, 1933.
(8) Evaluation and health practice. Editorial. Am. J. Pub. Health 40: 868–869, July 1950.
(9) Youmans, J. B.: Experience with a postgraduate course for practitioners: Evaluation of results. J. A. Am. M. Coll. 10: 154–173, May 1935.
(10) Stouman, K., and Falk, I. S.: Health indices. A study of objective indices of health in relation to environment and sanitation. Quart. Bull. Health Organ., League of Nations 5: 901–1081, December 1936.
(11) Methods in public health research. Proceedings of a conference held under the auspices of the Public Health Study Section, National Institutes of Health, in conjunction with the Graduate School of Public Health, University of Pittsburgh. Am. J. Pub. Health 41: 1–117, August 1951, pt. 2.
(12) Evaluation studies which have contributed to school health services and education. Am. J. Pub. Health (Year Book 1951–52) 42: 125–129, May 1952, pt. 2.
(13) Yankauer, A.: Designs for evaluation needed in the school health field. Am. J. Pub. Health 42: 655–660, June 1952.
(14) American Public Health Association. Medical Care and Statistics Sections. Joint Committee on Medical Care Statistics: Medical care statistics. New York, The Association, 1953. Processed.
(15) Sheps, M. C., and Sheps, C. G.: Assessing the effectiveness of programs in operation. Study group reports, Committee IV on research, National Conference on Care of the Long-Term Patient. Baltimore, Commission on Chronic Illness, 1954, pp. 93–104. Processed.
(16) Cowles, J. T.: Current trends in examination procedures. J. A. M. A. 155: 1383–1387, August 14, 1954.
(17) Commission on Hospital Care: Hospital care in the United States. New York, the Commonwealth Fund, 1947.
(18) Gonzalez, J.: Medical staff organization vital to accreditation. Hospitals 27: 104–106, April 1953.
(19) Crosby, E. L.; The goal of accreditation. Mod. Hosp. 78: 74, May 1952.
(20) Gundersen, G.: Benefits of hospital accreditation to the medical profession. J. A. M. A. 154: 917–918, March 13, 1954.
(21) Ponton, T. R.: The medical staff in the hospital. Revised by M. T. MacEachern. Chicago, Physicians Record Co., 1953.
(22) Hill, F. T.: The staff audit and the consultation ratio. J. Maine M. A. 42: 58–60, February 1951.
(23) Mortrud, L. C.: The control of professional practice through the medical audit. Hospitals 27: 91, September 1953.
(24) Krause, C. D.: The merits of a medical audit. Mod. Hosp. 81: 85–86, December 1953.
(25) Perdew, W. C.: Are your patients properly cared for. A medical audit, or analysis, helps the staff answer this question. Mod. Hosp. 71: 84–88, July 1948.
(26) Cabot, R. C.: Diagnostic pitfalls identified during a study of three thousand autopsies. J. A. M. A. 59: 2295–2298, December 28, 1912.
(27) Swartout, H. O.: Ante-mortem and post-mortem diagnoses. New England J. Med. 211: 539–542, September 20, 1934.
(28) Pohlen, K., and Emerson, H.: Errors in clinical statements of causes of death. Am. J. Pub. Health 42: 251–260, March 1942; 33: 505–516, May 1943.
(29) Redlich, F. C., Dunsmore, R. H., Brody, E. B.: Delays and errors in the diagnosis of brain tumor. New England J. Med. 239: 945–950, December 16, 1948.
(30) Miller, N. F.: Hysterectomy: Therapeutic necessity or surgical racket. Am. J. Obst. & Gynec. 51: 804–810, June 1946.
(31) Doyle, J. C.: Unnecessary ovariectomies: Study based on removal of 704 normal ovaries from 546 patients. J. A. M. A. 148: 1105–1111, March 29, 1952.
(32) Doyle, J. C.: Unnecessary hysterectomies: Study based on 6,248 operations in 35 hospitals during 1948. J. A. M. A. 151: 360–366, January 31, 1953.
(33) Myers, R. S.: Professional activity study. Hospital statistics don't tell the truth. Mod. Hosp. 83: 53–54, July 1954.
(34) Letourneau, C. U.: The legal and moral aspects of unnecessary surgery. Hospitals 27: 82–86, May 1953.
(35) Robson, S. M.: Why pick on the surgeons? Mod. Hosp. 82: 57–58, February 1954.
(36) Professional activity study. Slee, V. N.: Statistics influence medical practice; Mooi, H. R.: Doctors do take records seriously; Hoffmann, R. G.: We must ask the right questions to get the right answers; Erickson, W.: Small hospitals benefit by the new approach; Van der Kolk, B.: Did they have pneumonia—Or didn't they; Farr, V.: Record librarian lists advantages; Eisele, C. W.: Opinions are no basis for objective analysis. Mod. Hosp. 83: 53–64, July 1954.
(37) Ciocco, A., Hunt, G. H., and Altman, I.: Statistics on clinical services to new patients in medi-

cal groups. Pub. Health Rep. 65: 99–115, January 27, 1950.

(*38*) Furstenberg, F. F., Taback, M., Goldberg, H., and Davis, J. W.: Prescribing, an index to the quality of medical care: A study of the Baltimore City medical care program. Am. J. Pub. Health 43: 1299–1309, October 1953.

(*39*) Lembcke, P. A.: Measuring the quality of medical care through vital statistics based on hospital service areas. I. Comparative study of appendectomy rates. Am. J. Pub. Health 42: 276–286, March 1952.

(*40*) Wright, M. J.: Improvement of patient care. A study at Harper Hospital. New York, G. P. Putnam's Sons, 1954.

(*41*) O'Malley, M., and Kossack, C. F.: A statistical study of factors influencing the quality of patient care in hospitals. Am. J. Pub. Health 40: 1428–1443, November 1950.

(*42*) Kossack, C. F.: To measure the quality of a hospital. Mod. Hosp. 81: 77–79, July 1953.

(*43*) Hagood, M. J., and Price, D. O.: Statistics for sociologists. Rev. ed. New York, Henry Holt and Co., 1952, pp. 138–143.

(*44*) Reynolds, W. E.: Some problems of clinical measurement in the study of chronic diseases. *In* Research in public health. Papers presented at the 1951 annual conference of the Milbank Memorial Fund. New York, Milbank Memorial Fund, 1952, pp. 76–88.

(*45*) Meyer, S. N., Hougen, A., and Edwards, P.: Experimental error in the determination of tuberculin sensitivity. Pub. Health Rep. 66: 561–569, May 4, 1951.

(*46*) Birkelo, C. C., Chamberlain, W. E., Phelps, P. S., Schools, P. E., Zacks, D., and Yerushalmy, J.: Tuberculosis case finding. A comparison of the effectiveness of various roentgenographic and photofluorographic methods. J. A. M. A. 133: 359–366, February 8, 1947.

(*47*) Makover, H. B.: The quality of medical care: Methodology of survey of the medical groups associated with the Health Insurance Plan of New York. Am. J. Pub Health 41: 824–832, July 1951.

(*48*) Goldmann, F., and Graham, E. A.: The quality of medical care provided at the Labor Health Institute. St. Louis, Labor Health Institute, 1954.

(*49*) Savage, I. R.: Bibliography of nonparametric statistics and related topics. J. Am. Stat. A. 48: 844–906, December 1953.

(*50*) Bradley, R. A.: Some statistical methods in taste testing and quality evaluation. Biometrics 9: 22–38, March 1953.

(*51*) Mason, D. D., and Koch, E. J.: Some problems in the design and statistical analysis of taste tests. Biometrics 9: 39–46, March 1953.

(*52*) Moroney, M. J.: Facts from figures. Baltimore, Penguin Books, 1951, ch. 18.

(*53*) Bradley, R. A.: Some notes on the theory and application of rank order statistics. Blacksburg, Virginia Agricultural Experiment Station of the Virginia Polytechnic Institute, undated. Processed.

(*54*) Stevens, S. S.: On the theory of scales of measurement. Science 103: 677–680, June 7, 1946.

(*55*) Greenberg, B. G., Harris, M. E., MacKinnon, C. F., and Chipman, S. S.: A method for evaluating the effectiveness of health education literature. Am. J. Pub. Health 43: 1147–1155, May 1953.

(*56*) Hill, B. A.: Observation and experiment. New England J. Med. 248: 995–1001, June 11, 1953.

(*57*) Dorn, H. F.: Philosophy of inferences from retrospective studies. Am. J. Pub. Health 43: 677–683, June 1953.

(*58*) Cochran, W. G.: Matching in analytical studies. Am. J. Pub. Health 43: 684–691, June 1953.

(*59*) Greenberg, B. G.: The use of analysis of covariance and balancing in analytical surveys. Am. J. Pub. Health 43: 692–699, June 1953.

(*60*) Wright, J. J., Sheps, C. G., and Gifford, A. E.: Reports of the North Carolina syphilis studies. IV. Some problems in the evaluation of venereal disease education. J. Ven. Dis. Inform. 31: 125–133, May 1950.

(*61*) Densen, P. M., Padget, P., Webster, B., Nicol, C. S., and Rich, C.: Studies in cardiovascular syphilis. II. Methodologic problems in the evaluation of therapy. Am. J. Syph., Gonor. & Ven. Dis. 36: 64–76, January 1952.

(*62*) Platt, P. S.: The validity of the appraisal form as a measure of administrative health practice. New York, American Public Health Association, 1928.