## Isidore Altman \*

To set the stage for this presentation and to explain its rather different character -- in the sense that it is different from most presentations at an A. S. A. meeting -- let me quote from my letter to our program chairman, Mr. Elijah White.

"Some years ago, I published an annotated bibliography on methodology in evaluating the quality of medical care, 1955-1961. <sup>(1)</sup> A supplement covering 1962 through 1967 (now 1968) is . . . nearing completion. If you . . . feel something of this sort -- a description of what people are doing in this field (no statistics, no equations) -- might be suitable, I would be glad to submit an abstract."

I thought it might be of a little interest to a group like this, the Social Statistics Section, to hear of attempts to measure in some statistical way an intangible concept like quality of care -to measure something which cannot even be defined, except that we all probably have some notion about what it is; and good care is something we all devoutly wish for when we get sick. I shall not attempt to define it myself, except to quote from Dr. Donabedian: ". . . the definition of quality may be almost anything anyone wishes it to be although it is, ordinarily, a reflection of values and goals current in the medical care system and in the larger society of which it is a part." (2)

The basis for my remarks is a bibliography I have been compiling (or, more accurately, have had excellent young women assistants compiling) since 1955. We began our work in 1961 but made 1955 our starting point for search of the literature. The University of Pittsburgh Press has published a collection of abstracts covering the years 1955-1961, inclusive, and we hope to publish another collection next Spring for the years 1962 through 1968.

If you reflect for a moment, it may occur to you that every paper that has to do with treatment or handling of patients has some relationship to quality. Writers on the subject of quality have had to wrestle with the problem of scope. We tried, therefore, to set some rules or criteria for selection. In the preface to our first publication, we said this:

"There are a great many papers appearing in medical journals which make a contribution to medical care because they present new drugs, new instruments and new procedures that are proposed and evaluated in terms of assisting in the diagnosis, treatment, or alleviation of a specific condition. Such items were generally omitted from this bibliography. For example, a paper describing the value of a new antibiotic would be excluded, but a paper which contained a method for determining the extent to which patients were being treated with antibiotics unnecessarily would be included. We are concerned in this bibliography with methods which would have more general applicability than the effect upon one specific condition. / The next is important because here writers differ./ Studies of efficiency, costs, or utilization of services have been excluded unless it was the intent of the author to demonstrate the bearing of these aspects on quality. While the focus was on personal health services, many psycho-social aspects, nursing procedures and mental health were omitted."

As you might suppose, there were and are many borderline situations and difficult decisions to be made. We can only try to be reasonably consistent. And when you are charging a fee for your collection of abstracts, the customer does not mind a few extra abstracts; we have never had any complaints on that score.

Now let me, as it were, dispose of two more kinds of papers, one of which might be called non-statistical and the other "general statistical." As to the first kind, there are many papers in the literature which discuss one or more of the problems involved in measuring quality, with no data being utilized, except perhaps casually. While we try to avoid these in our bibliographies, a number of them do get in -- because of the cogency of their arguments or because they describe new systems, such as progressive patient care. Some are valuable review papers which decidedly warrant inclusion. A more or less typical abstract of a nonstatistical paper reads like this:

"Changes in the social order such as population mobility, shifts in age composition and disease prevalence, and commercialization of the professions are discussed in relation to their effects on patient care and the search for better patient care. The author sees the need for cooperation among members of the health professions in order to improve the quality of care and he sets forth possible approaches toward the

<sup>\*</sup> Professor of Biostatistics, Graduate School of Public Health, University of Pittsburgh. Presented before the Social Statistics Section, Americal Statistical Association, August 21, 1968.

achievement of this cooperation. "(3)

The second kind of paper, what I have termed the "general statistical", represents the usual or most common approach where the collection of data is involved. Such a paper might describe a trial or an experiment where observations are made on two groups -- one treated the other untreated; one subjected to a new pattern of care, the other not -- and then the proportion of recoveries, or whatever the appropriate measure might be, is calculated. The analysis is based on simple rates, computed for different age groups or the sexes. etc. Sometimes a test of significance for the difference between proportions is applied, which all too often serves chiefly to clutter up the tabulations. (I am jesting, but only in part.) One of the best studies, which went beyond simple rates, says this:

"The methods employed throughout this study for testing the statistical significance of the observations to rule out the possibility of chance happenings consisted of standard procedures. These included contingency  $X^2$  corrected for continuity, t-test for difference between two means, analysis of variance and covariance, t-test for significance of regression and correlation coefficients, and non-parametric procedures such as the sign test. Results stated as statistically significant were such that the probability of their occurrence by chance alone was always 5% or less. In some cases the exact probability was stated." (4)

I am going to kill two birds with one stone, in a way, by supplying illustrations within the classification scheme we have been following. Most attempts to assay the quality of some aspect of medical care fall into one of two broad groups: (1) observation of performance, directly or indirectly, and the application of some scheme of judgment, or (2) a gauging of the results presumably achieved. (Our published bibliography has two other sections, one calling attention to published sets of established standards of performance and the other containing a few unclassifiable items.) Time permits only an example or two from each of 5 sub-heads under the two major headings.

Under the heading <u>Elements of Perfor-</u> mance we had as a first sub-head "Audit, review, and evaluation." It seems only right to begin with a study that is becoming a classic in this field - - the study of general practice in North Carolina by Dr. Osler Peterson and his associates, from which I have already quoted. <sup>(4)</sup> This major study consisted basically of actual observations of physicians by other physicians as they carried on their practices. Practice was broken down into components which were scored -- by observing judges, in effect. Then, as part of the study, the findings on quality of work were shown to be associated with certain elements of education and training and other factors; that is to say, certain characteristics of the physician were predictors of professional performance (within limits).

Our second sub-head under Elements of Performance was "Suitability of the modality of care." The word "modality" refers to such kinds of care as home care, nursing home care, intensive care, physical and occupational therapy, and social services.

Here the before-and-after kind of approach appears to have proven the most useful. The functions the patient is able to perform or the activities he is able to handle might be coded periodically and measures of functional improvement obtained in some way.

The third sub-head under Elements of Performance is "Screening and case-finding." Here we generally have in mind community efforts to improve quality through prevention and timely treatment. Statisticians who have involved themselves with health examinations of substantial numbers of people have investigated the occurrence of false-positive and false-negative findings, with a view to determining the sensitivity and specificity of the tests which are employed. (5) Sensitivity "is measured by the percentage of all diagnosed cases that are screened positive by a particular test." Specificity "is measured by the percentage of the persons without the disease that are screened negative by the test in question." One statistician and his colleagues found that "The 'cheaper' tests frequently emerged as the most expensive ones when viewed in terms of effectiveness in identifying disease." (6)

Our second major category was the Effects of Care. In its most elementary aspect, the question one asks is, Did the patient recover? If he did, then presumably he received good care. (As an aside, it is possible that he recovered in spite of poor care or no care.) On a higher level, the approach has been through the examination of mortality and morbidity statistics--often the comparison of rates after the institution of some program of treatment or intensive study with the rates that had prevailed previously. Our first sub-head under Effects of Care is therefore "Mortality and morbidity rates".

One notable paper here was an evaluation of a measure (the perinatal mortality rate) which is commonly accepted as an index of quality of maternity care. If only one measure had to be used, this study showed, the perinatal mortality rate was the figure to be used. But the figure itself needed to be adjusted for differences in population composition, and the user was warned to be wary of over-simplification resulting from the use of a single index. <sup>(7)</sup>

Our second sub-head "Patient expectations and satisfactions" gets over into human behavior, emotions, psychology, and the like. If the patient is happy about his care and <u>he</u> thinks or believes that he was the recipient of quality medical care, then ipso facto his care has been good. This seems to be a generally accepted notion, as judged from the papers we have abstracted -and I would not quarrel with it.

A good deal of attention is paid these days to satisfaction with nursing care in the hospital. The most intensive study of this kind, probably, went at it by correlating satisfaction with different specific elements of nursing. Over 8000 patient forms and over 9000 personnel forms (usable ones) were collected in the course of the study. Patients and personnel other than nurses, it turned out, did not see eye to eye with the nurses themselves on which items of care best make for a happy patient. <sup>(8)</sup> But I get the impression that patient satisfaction is a very difficult thing to measure, and that the investigators are struggling.

As I indicated earlier, our final section was "Other" -- actually called, <u>General Ap-</u> <u>proaches</u>. Few studies were so unclassified, mostly discussions. But one study here tried to get a picture of general quality of medical care in the community by assessing the amount of unmet meed. This was done through a household survey which inquired about symptoms. As you can begin to gather, quality is a most elusive concept.

Finally, I must call attention to two e writers in this field and their works, with apologies to the other fine students whom I have not singled out. The first of these I consider our leading thinker on the subject of quality of medical care, Dr. Avedis Donabedian, Professor of Medical Care Economics at the University of Michigan, whom I mentioned earlier. Two very thoughtful and definitive papers of his (from which I am constantly borrowing) have appeared, respectively, in the Milbank Memorial Fund Quarterly July, 1966, Part 2, and in Medical Care, May-June, 1968.<sup>(9)</sup>

The other writer, Dr. Mindel C. Sheps, Professor of Biostatistics, Columbia University, School of Public Health, published her paper, "Approaches to the Quality of Hospital Care" back in 1955 in <u>Public Health Reports.</u> <sup>(10)</sup> But no paper since has approached it for clarity of exposition on the methodology of the kinds of studies we have been discussing.

I might mention that the "biggest study of them all" is currently under way at Yale's School of Public Health. Top-flight physicians in the School of Medicine are being questioned in great depth on how a long list of specific diseases should be diagnosed and treated. Out of this will evolve, I am sure, much clarification of what good medical care should consist of.

As I read our abstract and quite a number of the papers themselves, I get un uneasy impression of a lack of attention to statistical refinements in the studies which have been made; randomness, bias, reliability, validity, etc., do not receive the attention they should. But the field is in an early stage and the methods are still crude and rough and ready. It is a complex but intriguing field. These are "constraints" which may or may not attract you.

## References

 Anderson, Alice J., and Altman, Isidore. Methodology in Evaluating Quality of Medical Care: An Annotated Selected Bibliography, 1955-1961. Pittsburgh: University of Pittsburgh Press, 1962.

(2) Donabedian, Avedis. Evaluating the Quality of Medical Care. <u>Milbank Memorial</u> Fund Quarterly 24:166-206 (Part 2), July, 1966.

(3) Simmons, Leo W. Change in the Social Order as it Affects Health Services. <u>Hospital</u> <u>Progress</u> 45:81-86, June, 1964.

(4) Peterson, Osler L., Andrews, Leo P.,
Spain, Robert S., and Greenberg, Bernard G.
An Analytical Study of North Carolina General
Practice, 1953-54. J. Medical Education 31(12):
1-165 (Part 2), Dec. 1956.

(5) Kurlander, Arnold B., Hill, Elizabeth H., and Enterline, Philip E. An Evaluation of Some Commonly Used Screening Tests for Heart Disease and Hypertension. J. Chronic Diseases 2:427-439, Oct. 1955.

(6) Clark, Thomas W., Schor, Stanley S., Elsom, Katharine O., Hubbard, John P., and Elsom Kendall A. The Periodic Health Examination: Evaluation of Routine Tests and Procedures. <u>Annals of Internal Medicine</u> 54:1209-1222, June 1961.

(7) Pakter, Jean, Erhardt, Carl L., and Jacobziner, Harold. Perinatal Mortality Rates

as an Aid in Assessing Maternity Care. Amer. J. Public Health 45:728-735, June 1955.

(8) Abdellah, Faye G., and Levine, Eugene. Effect of Nurse Staffing on Satisfactions with Nursing Care. Hospital Monograph Series No. 4 Chicago, American Hospital Association, 1958. 82 pp.

(9) Donabedian, Avedis. Promoting Quality through Evaluating the Process of Patient Care. <u>Medical Care</u> 3:181-202, May-June 1968.

 (10) Sheps, Mindel C. Approaches to the Quality of Hospital Care. <u>Public Health Reports</u> 70:877-886, September 1955.