A Method of Evaluating and Improving the Quality of Medical Care

EDWIN F. DAILY, M.D., F.A.P.H.A., and MILDRED A. MOREHEAD, M.D., M.P.H., F.A.P.H.A.

In any program area of community health administration there is frequent need for measuring quality, as well as quantity of services rendered. The study experience reported upon here may have some suggestions for the projective reader concerned with the true productiveness of his own little area.

The Health Insurance Plan of Greater New York is an organized medical care plan financed by premiums paid by insured persons and their employers. The services are provided by 30 medical groups approved by and under contract with HIP with approximately 1,000 physicians who give the home, office, and hospital care to the 480,000 insured men, women, and children. The Board of Directors, the Medical Control Board, the Medical Department of HIP, and the medical groups together have the obligation of seeing that the insured subscriber population receives medical care of high standards. This is in contrast to the private insurance company and many Blue Shield plans which pay cash indemnities for specified services, but assume no responsibility for the quality of medical care provided.

The Board of Directors of HIP, composed of community leaders, has twice authorized studies of the quality of medical care provided by the medical groups. The first was made in 1949 * and the methods of this study were reported to the American Public Health Association in 1950. The study now in progress was started in 1954 and will be completed for all the 30 medical groups in 1956. The proposal by the Board of Directors for such a study was endorsed by the medical group directors. There was agreement that the findings concerning the study of each group would be considered confidential between HIP and the medical groups.

The purposes of the study were: (1) to evaluate the quality of medical care provided by individual physicians and by each medical group and to compare the findings of the several medical groups; and (2) to make recommendations to each medical group, based upon the study findings in an effort to bring about changes which would improve group practice of medicine and the quality of medical care received by insured persons.

Planning the Study

The qualifications set for the director of the study were that he be a physician held in high esteem by the profession as a clinician and teacher and one who had long been interested in the quality of medical care in the community.† To assist in its planning and direction a full-time physician with an M.P.H. and years of experience in epidemiology was assigned from the HIP staff.

* By Dr. Henry Makover.

† Dr. I. Ogden Woodruff, professor emeritus of clinical medicine at the College of Physicians and Surgeons, Columbia University, former chief of medicine at Bellevue Hospital, and former president of the Health Council of the City of New York, was selected and employed on a half-time basis to direct the study.

Among the early decisions were that the scope of the study should encompass the Departments of Medicine, (consulting internists and family physicians), Surgery, Obstetrics-Gynecology, Pediatrics, Radiology, and Pathology in each of the medical groups. These six departments provide over 80 per cent of all services in this medical care plan. The Departments of Dermatology, Neurology and Psychiatry, Ophthalmology, Otolaryngology, Orthopedics, and Urology provide less than 20 per cent of the professional services. Another decision was that the evaluation of quality should be carried out by recognized specialists in the respective fields and should be made objectively from a review of clinical records and case discussions. The study team physicians were to have no association with any of the medical groups affiliated with HIP.

The director of the study undertakes the study of internists and family physicians and to assist him in the work he has two younger Diplomates of the American Board of Internal Medicine who hold appointments in teaching hospitals and are also engaged in clinical practice in the community. The surgeon, obstetrician-gynecologist, and pediatrician selected as members of the study team are all senior attending physicians of their respective departmental staffs at different large teaching hospitals. The pathologist is an assistant professor of pathology in one of the local universities, and the radiologist is the chief of his department in a large voluntary hospital.

None of the study team members had worked together previously. They are all enthusiastic about the opportunity to participate in the study and to contribute to the improvement of medical care in New York. They agreed to devote 10–20 hours of their time per week visiting medical group centers and physicians' offices to examine case records and discuss this material with each physician interviewed. They further agreed to present all findings and recommendations to the entire medical staff of each medical group at an evening meeting.

The studies of each of the six departments in a particular medical group were carried out at about the same time. Trial runs showed that it would take from four to six weeks to assemble the required data from each medical group and prepare the reports for presentation.

Methods

The basic problems in methodology were to determine the type of material to be studied, the criteria for evaluation, the methods of recording information, and the relative weights to be given to the various factors. The approaches in all fields, except radiology and pathology, are essentially similar. For the Departments of Medicine, Surgery, Obstetrics and Gynecology, and Pediatrics the technic consists of a study of medical care provided by the individual physicians in four major areas: (1) clinical records of the management of cases of serious illnesses and discussion of these cases with the physicians; (2) clinical records of preventive services (except in surgery); (3) questioning the physicians in regard to factors which influence the operation of the medical group as a whole, for example, hospitalization policies, work loads, or hours on call; and (4) evaluation of office facilities and record systems.*

* Since in most of the medical groups there are family physicians and pediatricians with offices outside of the medical centers, because of the large geographic areas covered by most groups, this creates a problem of maintaining duplicate sets of medical records.

Dr. Daily is vice-president and Dr. Morehead is associate medical director, Health Insurance Plan of Greater New York, New York, N. Y.

This paper was presented before the Medical Care Section of the American Public Health Association at the Eighty-Third Annual Meeting in Kansas City, Mo., November 15, 1955.

Selection of Cases to Be Studied—The central office of HIP requires just one reporting procedure of its physicians, namely, that every time a physician renders a service, he enter the name of the patient, identifying data, and tentative diagnosis on a special form called Med 10.

The Med 10's for each physician are reviewed for a three-month period and from these cases of serious disease are chosen. The criterion for selection is a diagnosis suggesting the need for fairly extensive diagnostic procedures, the involvement of specialists, or other group facilities. No strict attempt has been made to standardize the selection as it has not always been possible to select the same number of the similar type of case for each physician.

For medicine 10 cases of illness are studied for each internist or family physician. These are primarily cases of hypertension, diabetes, peptic ulcer, anemia, or kidney disease. Cases of carcinoma and liver disease are included when encountered.

For gynecology five cases of major surgery and 10 cases of sterility, cervical lesions, endocrine disorders, or other gynecological conditions are studied.

For obstetrics 10 cases are studied, primarily those with complications.

For surgery 10 cases of major surgery and five cases of office consultations are studied.

For pediatrics, because of the infrequency of serious conditions, more minor conditions are studied. Of the 10 cases studied those of rheumatic fever, kidney disease, or congenital abnormalities are selected when available. Several cases of asthma, behavior problems, or secondary anemias are generally included.

Recording Data—When the study team specialist in medicine, surgery, obstetrics and gynecology, or pediatrics goes to the group physician's office he takes with him a list of the names of patients whose care is to be reviewed and a casework card * for each patient. On the front of the card is a list of all items to be studied and space to rate each item. The possible ratings are "good," "fair," or "unsatisfactory." He makes a brief summary of each case on the back of the card, noting the reasons for the rating he has given.

A set of criteria for family physicians was formulated to apply in general terms to the items to be rated; for example, it lists the elements a history must include to be considered good or fair. Similar criteria with appropriate modifications are used by the other interviewing specialists.

One of the premises that influenced the planning of methods of study was that although it is certainly desirable that adequate records be kept, it is possible that a physician may render adequate medical care with a minimum of recorded data. Therefore, in addition to a review of the patient's record, each case is discussed with the physician involved and if he has additional information regarding the case he is given credit accordingly in all sections except records.

Our interviewers were advised that when they had completed the review of the selected cases for a physician they were to study additional cases if they felt there was a considerable variation in the management of the cases. This has seldom been necessary. The range of scores of each physician's cases is generally narrow. That is, it has been our experience that when a physician handles several cases well, he is apt to handle the others in a similar manner; the same is true of a physician doing less capable work.

The three major areas that are studied for each case are: (1) records (this

* Copies of the casework card and criteria for evaluation of family physicians can be obtained by writing to H. I. P., 625 Madison Ave., New York 22, N. Y.

includes the quality of the history and physical examination, the adequacy of progress notes, the presence and organization of supplementary reports, including hospital transcripts and the appropriateness of the recorded diagnosis); (2) diagnostic work-up (this includes an evaluation as to whether the time interval in ordering procedures was satisfactory, whether indicated consultations, laboratory, x-ray, and other diagnostic procedures were ordered and an evaluation of the over-all management of the diagnostic period); (3) treatment and follow-up (this includes evaluation of the treatment, whether indicated laboratory procedures were ordered during this period, whether follow-up of the patient was satisfactory, and a general evaluation of total patient care).

The rating for many of these items is, of course, in large measure dependent on the judgment of the interviewing physician. The importance of the selection of competent impartial interviewers for the study team is, therefore, obvious. Several cross-checks were made between the two interviewing internists by having them interview the same family physicians. The differences in the scores of the family physicians based on these separate ratings did not exceed 7 per cent.

Scoring—The clinical material obtained from the case studies and the interviewer's rating is then checked by the assistant director of the study to determine accuracy, and numerical weights are assigned for scoring.

For cases of illness: The maximum score for any single case is 100. This represents the total score in each category where the maximum for records is 30, for diagnostic work-up 40, and for treatment and follow-up 30. Each item within a category has a numerical weight assigned according to whether it has been judged good or fair. An item rated unsatisfactory receives no credit.

The first step in analyzing the work of an individual physician is to obtain an over-all score for him based on the average of the cases of illness studied. The over-all score of each physician studied in a medical group is then placed in one of arbitrary scoring categories: 0–45, 46–60, 61–75, 76–100. We have assumed that those physicians in the highest scoring bracket are practicing a good quality of medical care, those in the second category an adequate quality of medical care, the third class is considered below average, and the fourth class is rendering a poor quality of medical care. We have found that we could clearly demonstrate to a medical group the differences in the quality of medical care provided by its physicians.

Another important use made of the scoring is to determine areas of strength and weakness in the group as a whole by analyzing the total score for all physicians for certain items. For example, in one group where 125 cases were studied it was noted that the lowest scoring on an item basis was on ordering the indicated laboratory procedures in the follow-up period. This point was stressed in presentation of the findings to the medical group.

For preventive services: Study of these services is primarily based on the presence or absence of expected preventive procedures. In a health examination we expect as a minimum a good history and physical, hemoglobin, urine examination, serology, and chest x-ray. In obstetrics we expect that good prenatal care will include certain standard procedures, registration before the third month of gestation, periodic visits, a complete delivery record, and postpartum visits. For pediatrics we expect a good initial history and physical, at least eight visits during the first year of life, immunization, feeding advice, etc. The preventive phase is not studied for surgery.

For the Departments of Radiology

and Pathology: The major emphasis is placed on equipment, work-spread, methods, training and experience of the technicians, and administrative factors, such as recording, appointment systems, and patient load. In radiology we also review 15 films from the standpoint of the methods used, quantity and quality of the films, and the group radiologist's report as to the adequacy of its description, interpretation, and final impression. Criteria are established for good, fair, and poor, and weights assigned. In pathology we also review 10 cases each in the fields of hematology, kidney disease, liver disease, and excised skin lesions. These charts are studied from the standpoint of adequate use of the laboratory and appropriateness and quality of the procedures used. We also are sending unknown specimens, glucose, hemoglobins, total protein, and BUN (blood urea nitrogen) for comparison with controls.

At the conclusion of the study of a medical group we have scores for each department and for each physician in preventive services and case management of illness. These can be averaged to give an over-all score of each department. Comparing the same departments of different medical groups we have found there is little difference in the scores of the departments of surgery, obstetrics and gynecology, pediatrics, laboratory, and x-ray. They have been, with occasional exceptions, uniformly good. The greatest variations both within a group and between groups have occurred in the scores of the family physicians on their case management of illness. This has been felt to be the most effective index for comparing medical groups.

Presentation—When the study material has been analyzed and reports prepared for each specialty surveyed we arrange for a mutually convenient evening with the medical group for the presentation of the findings. All members of the group and of the study team attend this meeting. Each interviewer reads the report of the department he surveyed. The study director introduces the presentation with a statement of our premises regarding group practice and he concludes with a general over-all discussion and recommendations.

For the four clinical specialties of Medicine, Obstetrics-Gynecology, Pediatrics, and Surgery our presentation to the medical group consists of general findings and detailed case illustrations to demonstrate specifically any weakness which has been noted in diagnosis, treatment, or follow-up. This is followed by our analysis of the ways the members of the department relate themselves to over-all patient care in group practice. Specific recommendations are then made. Very little emphasis is placed on the actual scoring, although we do give the number of those physicians who were found to be practicing good, fair, or unsatisfactory quality of medical care. We concentrate instead on emphasizing clinical examples, using the case summaries on our casework cards. The names of patients or of physicians are never mentioned; a code sheet identifying cases and physicians, however, is given to the medical director of the group.

We present slides which summarize for the Department of Medicine the types of record systems, the number of essential diagnostic procedures employed in cases of hypertension, peptic ulcer, diabetes, and health examinations. For Obstetrics-Gynecology and Pediatrics there are similar summary slides of the care of obstetrical patients and for well babies.

We also discuss the policies and problems of hospitalization of patients, the postgraduate education of physicians, including clinics and postgraduate courses and hospital appointments, the use of the Visiting Nurse Service and community resources, and the presence

or absence of a tumor board or diagnostic board. The functions of the department chiefs in supervising the work of physicians in their departments are outlined, including such items as periodic review of clinical records and evaluation of their use of the resources of the medical group.

For the Departments of Radiology and Pathology we discuss our general findings and any deficiencies encountered. Slides are presented illustrating our evaluation of the x-ray films and the radiologist's reports which were surveyed.

Although we have noticed antagonistic, self-protective reactions on the part of some group members at the presentations, when they have had an opportunity to review the material further they have felt, in general, that the findings and recommendations were fair.

It is the impression of the interviewing team that the quality of medical care provided by the 18 medical groups studied to date is, on the whole, considerably above the prevailing standards in the community.

Problems in Methods—We have felt that the methods we have used have been well suited to our objectives. However, we are aware of several problem areas which we would attempt to change in a future study. Among these are:

1. We have concentrated on those cases of serious disease which received an initial diagnosis. It might be desirable to study more cases diagnosed as minor illness and to endeavor to determine if more serious illness existed.

2. The system of scoring which we have been using is rigidly allocated among the three major categories—records, diagnostic work-up, treatment and follow-up. However, as there is considerable interdependence between these categories, it would be desirable to have the total case score reflect this in greater measure. This could be accomplished by either enlarging the scope of the criteria for each item or by having the numerical weights for each category dependent upon an acceptable minimum in the other categories.

3. Accurate comparisons of the quality of medical care under prepaid group practice and under fee-for-service solo practice would prove valuable for comparative purposes if they could be carried out in the same community.

4. Direct observation of a physician's work, evaluation of doctor-patient relationships, and average time per office service are areas of investigation which might be considered for other studies.

Follow-Up of Study

A team of five physicians—a chief internist and a chief pediatrician from different medical groups, selected because of their long experience and great interest in group practice, and from HIP the director of the Division of Administrative Practice, the director of the Division of Preventive Medicine and Health Education and the director of the Division of Professional Services meet with the Executive Committee of each medical group a week or two after the findings of the study are presented. This team offers its services to help the groups in any possible way to carry out the study team's recommendations. Often, several long meetings have been requested by the Executive Committee of the groups.

All members of this team of consultants draw upon their knowledge of "what works, and what does not work" based on the years of experience with the medical groups affiliated with HIP. This follow-up consultation which takes several weeks for each medical group has been found to be effective and appreciated.

Effectiveness of Study

Many specific examples of improved methods of practice by individual physicians have been noted. All recommendations concerning laboratory and x-ray departments (these were usually minor) have been promptly corrected. Improved use of consultation and diagnostic services available in group prac-

tice has been repeatedly observed. A better understanding of the potentialities of teamwork in group practice is evident in many of the physicians. Several rechecks of clinical records have shown them to be much improved a few months after the groups are studied.

The medical groups have terminated the services of several physicians with low scores.

Following the survey chiefs of the various departments, often for the first time, assume full responsibility for directing and supervising the work of their associates and integrating their work with other departments.

Conclusions

The methods of the study have been found to identify clearly those physicians who are providing good, fair, or poor medical care.

The direct presentations to the medical groups of the findings and recommendations by distinguished specialists, and the intensive follow-up does bring about many improvements in medical care.

We believe other medical care programs would find the methods described, adjusted to local circumstances, useful in evaluating and improving the quality of medical care.

S. pyogenes Infection Among Newborn

A single strain of Staphylococcus pyogenes, phage type 42B/47C/44A/52, has been established as responsible for all lesions in a recent epidemic of staphylococcal infections among the newborn in the nurseries of the Ohio State University Hospital. (A report is to be published in Pediatrics.) The same type organism has been found in subsequent studies of cultures from similar outbreaks in Portsmouth and Lancaster, Ohio; Lepeer, Mich.; Seattle, Wash.; Hartford, Wis., and Philadelphia, Pa. In all but one instance these strains have shown identical patterns of antibiotic sensitivity.

The university would appreciate receiving cultures from pustular lesions of infants or purulent lesions of mothers obtained during nursery outbreaks of epidemic proportions to continue these studies. Reports of bacteriologic findings on such cultures will be transmitted promptly.

Cultures and communications should be addressed to Thomas E. Shaffer, M.D., The Phage Typing Laboratory, Department of Bacteriology, Ohio State University, Columbus 10, Ohio, who has supplied the above information.