Implementation of a Peer Review System for Ambulatory Care

Experience of the Health Insurance Plan of Greater New York

THE NEED TO EVALUATE AND IMPROVE MEDICAL CARE has been recognized by the Health Insurance Plan of Greater New York (HIP) and its affiliated medical groups since the plan's beginning in 1947.

HIP is a prepaid health plan that contracts with and assists 28 affiliated medical groups in the delivery of comprehensive health care to 750,000 enrollees in the five boroughs of New York City and in Nassau and Suffolk Counties on Long Island. These medical groups use 1,000 physicians and 1,500 other health personnel to provide primary and specialty medical care in 45 centers and subcenters. Enrollment in the medical groups ranges from 5,000 to 125,000 persons, and 18 to 104 physicians serve the enrollees of a medical group.

As early as 1949, HIP conducted special studies in medical care evaluation of a peer review type. The Makover study (*1*) was a pioneering effort that involved the review of physicians' performance by examining medical records. Physicians from the medical groups reviewed patients' records and conducted interviews with other practicing physicians to evaluate their professional performance. The services of those whose performance was considered unsatisfactory were eventually terminated. Daily and Morehead (*2*) used a more sophisticated design in their study. Consultant specialists performed a medical audit of specific medical conditions and reported back to the medical groups. Standards of medical care, however, were not defined, and Daily and Morehead used a limited nonrandom sample of patients' records for review. The program was in operation from 1954 until 1956, and sporadic efforts followed. During the same period special studies were also undertaken to determine patient satisfaction with the health services provided by the affiliated medical groups.

In addition, HIP focused on improving the delivery of health services by offering financial incentives to the medical groups for providing defined kinds of medical services, services that were considered to be indicators of improved performance. For example, supplemental payments were made to medical groups offering such preventive services as the Papanicolaou test for uterine cancer and immunizations for children. Increasing the availability of services, for example, by providing evening and Saturday hours, also was rewarded with additional payments. In addition, special studies were also conducted by HIP to compare the delivery of health care to HIP and non-HIP populations (*3,4*).

As a result of the ongoing commitment of HIP and the affiliated medical groups to improve the quality of care to HIP patients, a contract monitoring program to evaluate the organization and delivery of services at the medical groups was put into operation in 1972. In 1973, the Peer Review Program (PRP) was established to evaluate physician performance. Both of these quality assurance programs, through contractual agreement, are subject to review by the New York City Department of Health. In this paper, the implementation of the Peer Review Program from 1974 through 1977 is described.

Selection of Method

After an extensive review of the available methods of assessing the quality of patient care, HIP selected Kessner and Kalk's tracer methodology as originally

JEANNE MAGAGNA DEUSCHLE, MA, DONALD N. LOGSDON, MD,
WILLIAM SOLLECITO, MS, WILLIAM STAHL, MD,
HARRY SMITH, Jr, PhD, MONA SONNENSHEIN, RRA,
and MARCIA KREITZER

described (*5*) to serve as a basis for its peer review system. Specifically, the HIP approach focuses on assessing the quality of physician performance by applying explicit process criteria formulated by peer committees for the conditions selected as being important in terms of the frequency of their occurrence in medical practice and the impact of medical management on the persons affected. Data required for application of these standards are collected from patients' records by professionally trained and supervised lay abstractors. This approach, which is workable in the HIP setting, provides a practical handle to use in objective assessment of the quality of the medical process and permits the documentation of improvements in physician performance within a reasonable period.

Organization

THE HIP review mechanism includes the Peer Review Committee, Medical Standards Subcommittee of the Peer Review Committee, a formal Advisory Committee, and an operational peer review staff (fig. 1).

The functions of the Peer Review Committee, which is comprised of six physicians from HIP and the affiliated medical groups, are to (*a*) adopt standards for professional services, (*b*) review staff findings on physician performance, (*c*) assess the quality and define the problems of health care, and (*d*) report unresolved problems to the Joint Committee for Improved Medical Care, the enforcement component of the mechanism.

The function of the Medical Standards Subcommittee, which is composed of six physicians who practice in the medical groups and have medical school or teaching hospital appointments, is to formulate standards of medical care for adoption by the Peer Review Committee. These physicians represent the specialties of internal medicine, pediatrics, surgery, and obstetrics and gynecology.

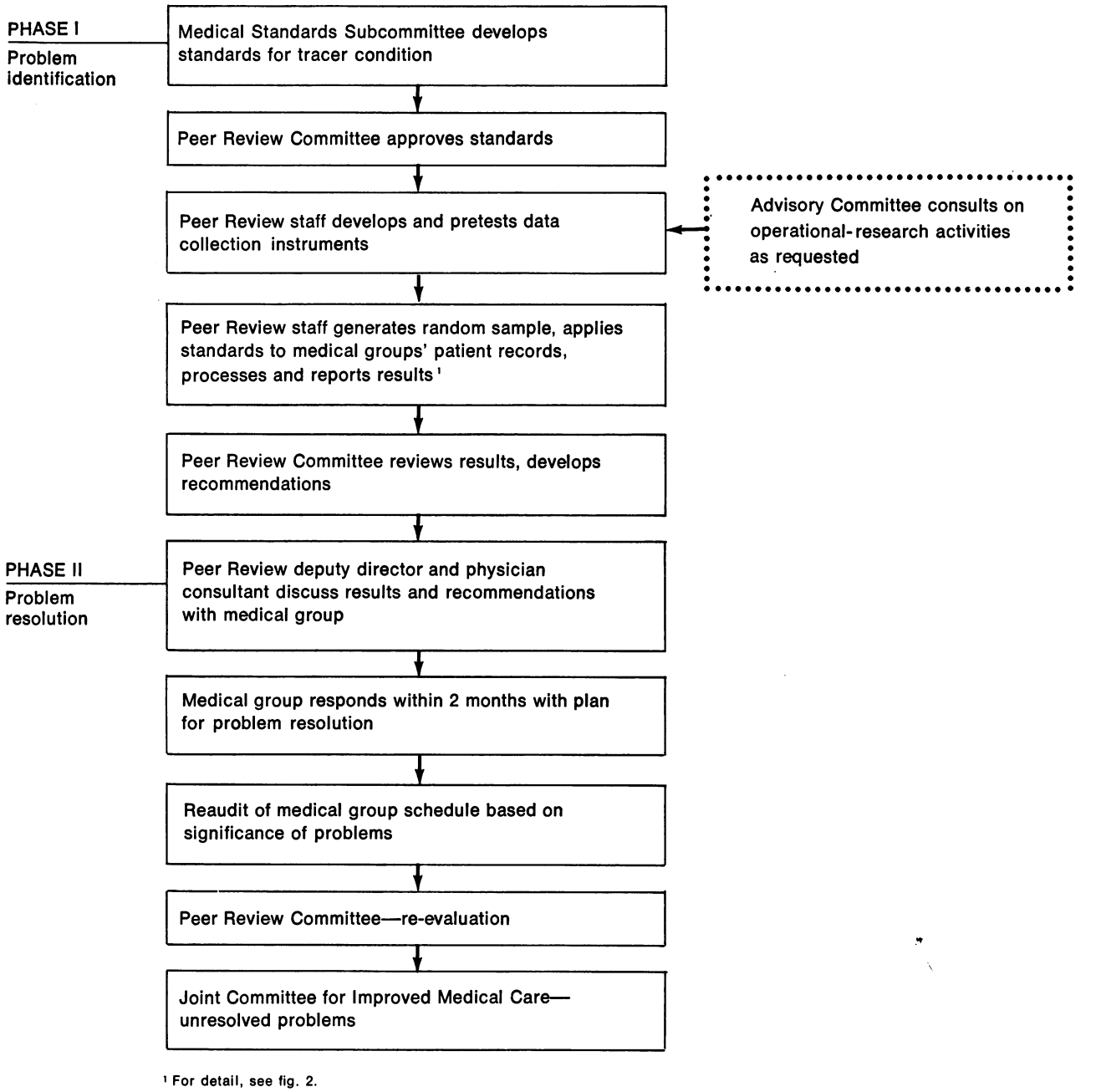
The formal Advisory Committee, which is comprised of five experts in medical care evaluation, serves as consultant to the Peer Review Program on operational and research activities.

The operational peer review staff consists of a deputy director-administrator, an assistant administrator-field supervisor, a physician consultant, a biostatistician, a biostatistical consultant, a records coordinator, four field abstractors, two record clerks, a control clerk, and a secretary. The medical director of HIP serves as director of the program and also as a member of the Peer Review Committee and Medical Standards Subcommittee.

☐ *Mrs. Deuschle is deputy director, Mr. Sollecito is a biostatistician, Mrs. Sonnenshein is the assistant administrator, and Mrs. Kreitzer is records coordinator of the Peer Review Program, Health Insurance Plan of Greater New York. Dr. Logsdon is medical director of the plan. Dr. Stahl is chief of surgery, Metropolitan Hospital Center, and professor of surgery at New York Medical College. Dr. Smith is chairman of the department of biostatistics, Mt. Sinai School of Medicine, New York City.*

Tearsheet requests to Mrs. Jeanne Magagna Deuschle, Peer Review Program, Health Insurance Plan of Greater New York, 625 Madison Ave., New York, N.Y. 10022.

Figure 1. Flow of activities in Peer Review Program



[1] For detail, see fig. 2.

Implementation

Medical care standards. In early 1974, physicians were selected to serve on the Peer Review Committee and Medical Standards Subcommittee. The Medical Standards Subcommittee began its task by focusing on developing standards of care for three tracer conditions—hypertension, acute otitis media in children, and breast lesions. It met approximately twice a month for 3 to 4 months, in addition to holding consultations with outside specialists, and submitted final draft standards for approval by the Peer Review Committee in June 1974. The Peer Review Committee, which met approximately once a month, modified the standards, and they were finally ap-

proved and sent to the medical group directors approximately 6 months later with a request for comment from group physicians who disagreed with any of the criteria. A more desirable time to have requested such input would have been before the criteria were approved by the Peer Review Committee. However, because of the large number of physicians in the system, such prior approval appeared impractical in view of the time constraints.

Essentially, no comments were forthcoming following this first distribution of the standards for the three conditions. One can only conjecture that the standards did not reach all medical group physicians or that any disagreement aroused was not expressed. However, following feedback of the results of the review (discussed later), frequent responses, both in agreement and disagreement with the standards, were obtained from the physicians.

Data collection instrument. A separate data collection instrument is designed for each condition for which standards have been approved. Each criterion contained in the standards has to be defined and refined so that all members of the abstracting staff will collect data uniformly.

In the first instruments designed for hypertension, acute otitis media, and breast lesions, data for the diagnostic process (that is, patient history, physical examination, and laboratory tests) require the abstractor to record physician compliance by coding from the patient's record the presence or absence of each criterion. However, for the more complex criteria relating to the therapeutic process, for example, specific treatment and followup, abstractors code only the raw data required for computer assessment of compliance. For both diagnostic and therapeutic processes, selected additional raw data are also recorded as a validity check. These instruments, which were modified as experience in data collection was gained, provide for flexibility in data collection and the method of analysis.

Sampling. Procedures have been set up for selecting a representative sample of patient records in each of the 28 HIP medical groups, a sample that would provide information to permit assessment of the quality of care provided for the conditions being reviewed. Each encounter between a physician and a patient in a HIP medical group is recorded on a Physician Visit Report form (MED 10), which includes space for entering the principal reason for each visit as well as patient identification information. The MED 10s are stored in hard copy form according to specialty and medical group in a central file and are available for review by the clerical staff of the Peer Review Program. From these forms, a random sample is drawn of patients recorded as having had a visit for one of the three conditions initially under review (hypertension, acute otitis media, and breast lesions).

Broad glossaries of the terms physicians use on the MED 10s to describe these conditions have been compiled so that the greatest possible yield of cases will be available for review. More cases are identified than are actually reviewed since, during the identification process, some cases are found to be ineligible for review (for example, patient has transferred to another medical group), and additional cases are excluded in the field upon examination of the medical records. (The reasons for exclusions are discussed later in the subsection "Collection of data.") A random sample of the identified cases is then selected, based on the expected number of exclusions and the total number of cases to be reviewed. The size of the sample for each condition being reviewed is based on the actual number of services provided by each medical group, its total enrollment, and estimates of the prevalence of the condition in each group. The actual number of cases selected is determined by use of a simple random sampling formula of the minimum sample needed to achieve an estimate (of weighted total scores) with a 95 percent confidence interval of 10 percentage points.

Training of abstractors. The program has four medical record abstractors; each has a college degree. All abstractors are professional in their dealings with medical group staffs and are able to work well independently.

The assistant administrator-field supervisor of the Peer Review Program is a registered records administrator with experience in a hospital medical records department. This expertise greatly facilitates the training of the staff in use of the patient record as the data source and in understanding the way in which a medical records department functions.

Abstractors are trained initially at the Peer Review Program office, where they become familiar with procedures and forms, in particular, the data collection instruments. They are then supervised daily at one of the medical groups until their accuracy and production are at a level comparable to that of experienced abstractors.

Before pretesting each medical care standard and data collection instrument, the abstractors attend a training session at the Peer Review Program office,

which includes a discussion of the epidemiologic and clinical aspects of each condition. The medical care standards and data collection instruments are then reviewed in detail. The program staff is of the opinion that this intensive preparation of field staff is time well spent in terms of the quality of the data abstracted. Moreover, the entire operational staff is thereby brought together for periodic feedback and exchange of ideas. Quality control procedures (discussed later in the subsection "Quality control") also serve as a tool for continuing staff education.

Approximately a month before review of a medical group, the field supervisor meets with the medical group administrator to discuss medical group procedures, policy, and facilities related to a PRP review. Such information facilitates the review by anticipating potential snags. For example, if a unit record system is not used, arrangements must be made to have all records of care for each patient included in the sample available to the abstracting staff for review. This meeting also provides PRP staff an opportunity to respond to questions concerning the program and the conduct of the review.

Collection of data. In most instances, two abstractors are assigned to review a medical group in order to facilitate the quality control procedures and to serve as a technical resource for one another. Because physical facilities at the medical groups in many instances are limited, abstractors are frequently assigned small quarters. Abstractors are required to keep a daily log of review activities, including problems encountered, such as difficulty in obtaining patients' records. This information is useful to the field supervisor in understanding changes in staff productivity between field visits.

Experience has shown that on the average, 4 to 6 weeks are required to complete a medical group review for the three conditions—hypertension, acute otitis media, and breast lesions. The median time for review of a hypertension case is 32 minutes and for an acute otitis media case, 12 minutes. Information on the review time for breast lesions is not yet available. It is obvious from these data that review for an episodic illness such as acute otitis media is less expensive than for a chronic illness such as hypertension, for which the medical care standards require the recording of more extensive and complex diagnostic and treatment data. In addition, an adult's medical record for chronic disease is often voluminous in comparison with a pediatric record.

Cases are excluded from review at the field level primarily because criteria for application of standards are not met. For example, since hypertension standards apply only to cases diagnosed within the 3-year period before review, patients under treatment before this time are excluded. Exclusions at this level are costly since the cases have been prepared for review by the clerical staff at the program office, records have been retrieved at the medical group, and abstracting time has been expended in excluding the cases. A detailed record of exclusions is therefore maintained for future analysis. The exclusion rates vary for each of the three conditions—53 percent for hypertension, 19 percent for acute otitis media, and 48 percent for breast lesions. For all three conditions, the median time required to exclude a case is approximately 4 minutes.

In those instances in which the abstractor finds the clinical data in the medical record ambiguous, the abstracts are held for review by the field supervisor, who visits the group twice weekly. If necessary, the field supervisor discusses these cases with the physician consultant. As anticipated, the number of problem cases that the physician consultant has to review decreases as abstractors' experience with application of the standards increases and they encounter a greater diversity in the patterns of care. As specific problems occur, it is necessary to further refine the criteria and set up additional definitions and instructions for the abstracting staff. Such modifications are given to the abstractors in the form of memoranda, which each abstractor is responsible for maintaining in a book to which he or she constantly refers. Again, the purpose is to achieve uniform abstracting.

In spite of the well-documented problem of the illegibility of physicians' notes on patient records, the abstractors develop unusual skill in interpreting physicians' recordings. Thus, illegibility codes are rarely used. In both acute otitis media and hypertension cases (breast lesion data have not yet been analyzed), the problem of illegibility has been encountered most frequently when a recording of the drug and dosage is required, and even then the illegibility rate for these items is only 1 percent. Entire cases are rarely excluded from review because of indecipherable records; such exclusions total less than 1 percent.

To avoid undue inconvenience to the staff of the medical group record room, abstractors request only as many records as they can review each day, and they return all records at the end of each work day.

Quality control. A significant effort has been made to insure the quality of the work being performed by the staff in all phases of the Peer Review Program. Procedures have been established and implemented

in the ongoing system to measure and assure the reliability and accuracy of the data collected (fig. 2).

To measure the reliability of the abstracted data, the field supervisor selects daily a random sample (approximately 5 percent) of the patient records that have already been abstracted and resubmits them to be read a second time by another abstractor. In addition, a subsample of these duplicate records is selected periodically and read a third time by the field supervisor. Preliminary data, based on these procedures, indicate that the reliability rate for abstracted items exceeds 90 percent, a rate consistent with the reproducibility rate for lay abstractors of medical records reported by Lyons and Payne (*6,7*).

The accuracy of the data coded on the abstract forms is checked manually as well as by computer. Because of the design of the abstract forms, the coded data can be proofread by comparing them with information recorded in the body of the form. Therefore, each case is proofread by a second person, and any errors detected are reviewed by supervisory personnel and corrected before the data are keypunched.

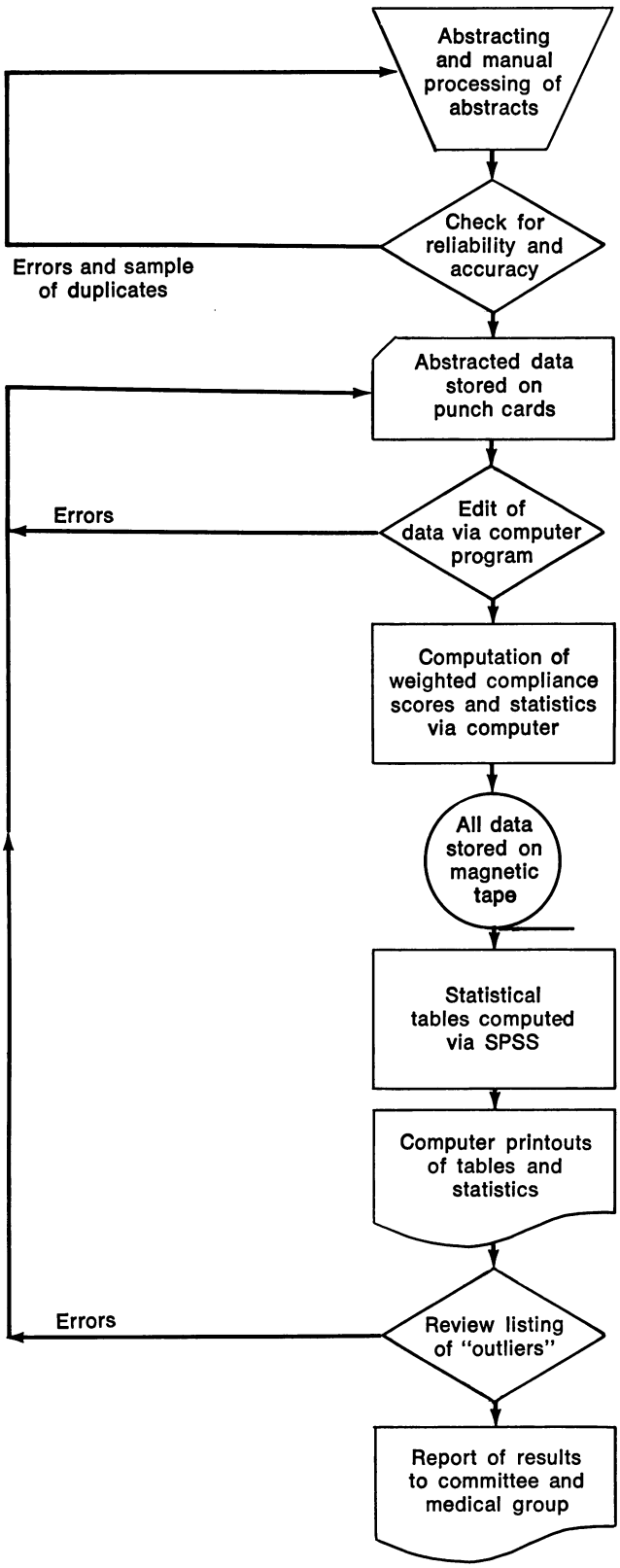
Further checks on the accuracy of the data being collected are accomplished by computerized programs. As with the proofreading, this is done for 100 percent of the data. All data coded on abstract forms are keypunched, verified, and then fed into a computerized editing system, which checks the validity of the coded data and, also, wherever possible, the consistency of codes betweens sets of items. Errors detected in this process are printed out on a computerized listing, which is reviewed by a data control clerk and checked and corrected on the abstract form by supervisory personnel. Then corrected punchcards are resubmitted to the computer system.

The final stages in the quality control process entail the review of the computerized listings of all data stored on tape as well as a review of summary statistics and tables in order to identify "outliers" and errors that may not have been detected in the previous stages.

Method of Analysis

Assignment of item importance weights. Preliminary analysis of the data for acute otitis media and hypertension raised two important issues: Should a summary index of physician scores for compliance with each medical standard be used to evaluate the care provided, and if so, should this score be based on the assignment of unequal item importance weights to each item collected? Lyons and Payne (*6*) have discussed these questions in detail and pointed out that when a summary index is used without assigning

Figure 2. Production flow and quality control of abstracted data in HIP Peer Review Program



unequal item importance weights to each item collected, each item will have a unit (equal) weight. Thus, subcategories of the particular process being reviewed would be weighted by the number and not in terms of the importance of the items within the subsegments. This fact was clearly seen in pretest data presented to the Peer Review Committees. Therefore, in order to make comparisons among HIP medical groups, summary indices were computed for each case, and unequal item importance weights based on clinical significance were assigned to each item in the standards. For example, although recording the temperature of a child with acute otitis media is desirable, it is not as important as prescribing an acceptable antimicrobial at an adequate dosage level. This approach to analysis of the results of peer review was discussed with the Advisory Committee at its first meeting in April 1975, and the committee members concurred with this plan.

In order for the individual medical group to determine specific areas in need of improvement, unweighted scores for individual items are also used. Thus, weighted scores are used to make comparisons between medical groups, while scores on individual items are most useful in pointing out specific deficiencies within each medical group. Novick and Dickenson (*8*) have shown this system to be effective in assessing the quality of care provided for iron deficiency anemia.

The table presents the item importance weights that were assigned to each of the major categories and subcategories for the first three conditions that were reviewed. Medical care standards are divided into two major components of care: (*a*) the diagnostic process, including early disease detection, history taking, physical examination, and laboratory and other tests and (*b*) the therapeutic process, including specific treatment and followup. In weighting an item, the category was first weighted, then the subcategories, and finally, the individual items. For example, for acute otitis media, the total diagnostic process was assigned a weight of 25 points and the total therapuetic process, a weight of 75 points, since the therapeutic process was deemed to be more significant in terms of medical management. Obviously, the result is that physician performance relative to treatment and followup items will contribute much more to the overall individual case score than diagnostic items.

Data processing and reporting. Because the data reporting system was still under development, all data for each medical condition from each medical group reviewed had to be stored in the most flexible system possible. For this reason, it was decided early in the program that all data should be stored on magnetic tape and that these data would be analyzed and retrieved by means of the HIP computer system (which consists of an IBM 370/145 that operates under DOS/VS).

The data are keypunched directly from the self-coded abstracting instruments and then copied onto magnetic tapes by means of computer programs that also edit the information being entered. After the editing, computer programs are used to apply item importance weights to the compliance items abstracted from the patients' records and to compute weighted compliance scores, as well as to provide the summary statistics needed for analyzing each patient's results. These compliance scores and the summary statistics are stored on tape along with all raw data for each patient.

All tables and statistics needed to review the results for each condition and each medical group are generated by means of the SPSS (*9*) system of computer programs, which is available in the HIP computer system. The "canned" programs (SPSS) have proved invaluable in performing the data analysis because the subset of tables needed to review the results of each condition has been modified as the program developed.

The exact set of tables used for analysis of each medical group has evolved since the beginning of the peer review program. The set of tables finally agreed upon for presentation of the data in a concise and meaningful way is made up of the percentage of physician compliance with each criterion as well as the median physician compliance scores for each medical group, including comparisons with all groups previously analyzed. Following analysis, the PRP staff presents tables of results, including a narrative analysis stressing areas of high and low physician compliance, to the Peer Review Committee for discussion and formulation of recommendations. In this presentation, the medical group is not identified except by code letter. In an effort to assure the Peer Review Program's educational objectives, the Peer Review Committee, early in the program, had established a policy of confidentiality in respect to medical group reports. For example, copies of medical group reports are letter-coded, reviewed by Peer Review Committee members, and returned to locked files. However, copies of the reports are sent to the medical group's director for reproduction and distribution at his discretion. Following discussion by the members of the committee, any changes considered appropriate are made in the narrative report. This report, which

Item importance weights assigned to each subcategory of medical standards for care of acute otitis media, hypertension, and breast lesions

Item	Acute otitis media	Hypertension [1]	Breast lesions [2]
Total	100.0	100.0	100.0
Diagnostic process	25.0	40.0	30.0
Early disease detection	([3])	4.0	6.0
History taking	8.0	14.0	3.0
Physical examination	12.0	7.0	17.0
Laboratory and other tests	5.0	15.0	4.0
Therapeutic process	75.0	60.0	70.0
Treatment (management)	50.0	30.0	0.0
Followup	25.0	30.0	70.0

[1] Weights apply to mild and moderate cases only. Severe hypertension cases are weighted as follows: treatment 60.0 and followup 0.0.

[2] Weights apply to lesions that are classified as benign throughout and that are never biopsied during entire followup period.

Lesions initially classified as suspicious that are found to be benign after biopsy are weighted as follows: treatment 50.0 and followup 20.0.

Lesions initially classified as suspicious that are found to be malignant after biopsy are weighted as follows: treatment 70.0 and followup 0.0.

[3] Not applicable.

includes recommendations for corrective action, item compliance scores, and total median scores, is then sent to the medical group director. Since evaluation is based on the total care delivered by the physicians within the medical specialty, results for individual physicians are not provided unless requested by the medical group.

After the director of the medical group has had an opportunity to become familiar with the report, the deputy director of the Peer Review Program, the field supervisor, and the physician consultant meet at the medical group with its representatives to discuss the report and the recommendations of the Peer Review Committee. Without exception, these meetings have been free and open discussions at which the medical group physicians have exhibited a positive attitude toward the need for corrective action in areas of low compliance. At the meetings, the physicians are encouraged to comment on the medical care standards. They have made suggestions concerning the desirability of certain criteria and raised questions about the importance of others. The PRP policy of keeping the data confidential is also stressed. The medical group director is reminded that a letter of response to the recommendations of the Peer Review Committee is required within 2 months and that it must include the medical group's plan to improve physician compliance with the medical care standards.

Problem Solving and Reaudit

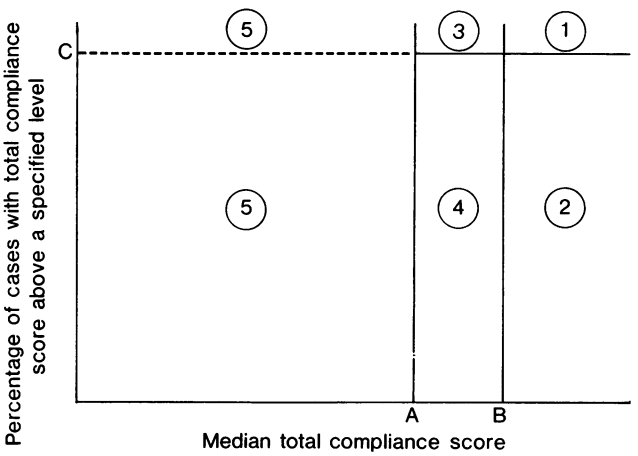
Methods for problem solving and for reaudit are being explored. An educational approach is planned to motivate and assist the medical group physicians in addressing the problems identified through peer review. A period of at least 2 months is probably necessary to permit reeducation and implementation of such efforts. Pilot reaudit reviews to assess improvement in physician compliance with medical care standards are currently being conducted at selected medical groups.

A management tool has been devised and is being used on an exploratory basis to determine the significance of identified medical care deficiences (fig. 3). Two parameters are considered important: first, the medical group's overall compliance score as measured by the group's median compliance score, and second, the percentage of individual cases in the group that are judged to be above a specified compliance level. The grid is divided into five areas based on the following management criteria:

Level A—an acceptable median group total score
Level B—an outstanding median group total score
Level C—an acceptable percentage of individual cases with total scores above a specified level

After the peer review audit, each group is characterized by estimates of the two parameters just mentioned, that is, its median total score and the percentage of individual scores within the group that are above a specified level. These coordinates locate the

Figure 3. Criteria for determining significance of medical care deficiencies



NOTE: Because the median total compliance score is a function of the percentage of cases above a specified level, some areas of the grid (upper left and lower right) are not mathematically possible.

group on the management grid. Management action is determined by the area into which the group falls. For example, if it were to fall into area 5, that is, a median score less than the acceptable level A and a percentage of individual scores less than the acceptable level C, the group would be assigned a high priority in the scheduling for reaudit. On the other hand, if a group were to fall into area 1, that is, an outstanding median score and a percentage of individual cases above the acceptable level C, it would be assigned a lower priority in the reaudit schedule.

Costs

The Peer Review Program's cost over the 3 years 1974–76 was $500,000. Approximately one-half of this figure was for development and one-half for operations. This outlay ranged from 100 percent for development in the first year to 50 percent in the second year and 40 percent in the third. The continued high proportion of costs assignable to development can be attributed to the continuing formulation and testing of standards for new tracer conditions and the setting up of related computer systems. Based on total program expense (excluding rent, capital equipment, telephones, and printing costs) for the 3 years, costs were distributed as follows: full-time staff 75 percent, committees and consultants 10 percent, computer systems and data processing 12 percent, and travel to medical groups and miscellaneous 3 percent.

To assess the cost per enrollee, the third year of the program was used as being most representative of ongoing operations. The total Peer Review Program's cost for 1976 was $260,000, or 35.5 cents per enrollee. This sum represents less than one-half of 1 percent of the total HIP per capita cost for health care during the same period.

Discussion

In addition to 1 year for program development (1974), a 3-year period (1975–77) was required to complete the review for the three conditions that were included in the first round of the Peer Review Program operation. A sample of approximately 7,000 patient records from the 28 medical groups (3,000 records of patients with hypertension, 2,100 records of patients with acute otitis media, and 1,900 records of patients with breast lesions) were included in the first review of HIP physician compliance with medical care standards. The 3 years that were required included the phasing in and training of the abstracting staff, as well as the refinement of the data collection and quality control procedures.

Because of the lack of data processing staff in the early stages of the program, as well as the program's technical complexity (for example, application of item weights and computation of weighted scores), it took more time than had been anticipated to complete the computer systems that were needed for analysis of the results of the review. This long period resulted in a delay in reporting the review results to the Peer Review Committee and to the medical groups. Thus, for acute otitis media and hypertension, presentation of the results of review both to the Peer Review Committee and to the medical groups was completed in late 1977. The computer system for breast lesions was completed in January 1978, and presentation of these data to the Peer Review Committee and to the medical groups began in the spring of 1978.

The Medical Standards Subcommittee and the Peer Review Committee approved standards of care for three additional conditions for initiation of a second round of reviews in mid-1977. These included cervical neoplasia, anemia in children, and urinary tract infections in adults. The source of cases for all three conditions will include either HIP Physician Visit Reports (Med 10) or HIP Centralized Laboratory Services reports and medical group laboratories reports, as applicable. These procedures have already been set up and tested, and it has been found that use of the laboratory as a source for case identification is feasible. However, use of this source appears to be more costly because of difficulties in patient identification and in the way laboratory data are filed. Data collection instruments and pretesting for cervical neoplasia, anemia in children, and urinary tract infections have been completed, and the conditions appear to present no special difficulty in data collection. It will be of interest to compare the level of physician performance for two conditions within the same specialty area, for example, in pediatrics—acute otitis media and anemia. Some spillover effect is anticipated from feedback of the results for the first three conditions analyzed.

The Advisory Committee to the Peer Review Program has met with the peer review staff, the Medical Standards Subcommittee, and the Peer Review Committee each year since the Peer Review Program became operational and has provided invaluable assistance in reviewing progress, suggesting modifications in methods and in priorities for implementation of activities, and finally, in helping to identify areas to explore in the future.

Specifically, future planning includes: (*a*) modification of the first three medical care standards, (*b*)

exploration of the potential for including some kind of intermediate outcome assessment, and (*c*) exploration of ways to control the program's cost by reducing exclusion rates and the time required to identify cases for review.

In terms of setting up new medical care standards, the Peer Review Program committees in the future plan to explore, in lieu of standards for assessment of the total care process (which has been the approach to date), what Kessner and associates (*10*) described as "subsets of the care process as a whole . . . a set of tracers . . . that highlight each major medical care activity, using symptoms, problems, diagnosis, drugs or procedures that are common in a given practice or community."

Conclusion

From a methodological point of view, the Peer Review Program is successful, as demonstrated by the activities that are now being carried out on an ongoing basis. However, the basic question of the program's effectiveness is yet to be determined. That will be done at the conclusion of the next phase of the program, which addresses the basic issues of how to assist medical group physicians in the correction of deficiencies and how to assess, through reaudit, improvement in physician performance. The assumption is that modification of physician performance will result in improved patient care.

References

1. Makover, H. B., The quality of medical care. Am J Public Health 41: 824–832 (1951).
2. Daily, E. F., and Morehead, M. A.: A method of evaluating and improving the quality of medical care. Am J Public Health 46: 848–854 (1956).
3. Shapiro, S., Weiner, L., and Densen, P. M.: Comparison of prematurity and perinatal mortality in a general population and in the population of a prepaid group practice medical care plan. Am J Public Health 48: 170–187 (1959).
4. Shapiro, S., et al.: Patterns of medical use by the indigent aged under two systems of medical care. Am J Public Health 57: 784–790 (1967).
5. Kessner, D. M., and Kalk, C. E.: A strategy for evaluating health services. Contrasts in health status. National Academy of Sciences, Washington, D.C., 1973, vol. 2.
6. Lyons, T. F., and Payne, B. C.: The use of item importance weights in assessing physician performance with predetermined criteria indices. Med Care 13: 432–439 (1975).
7. Payne, B. C., et al.: The quality of medical care: evaluation and improvement. Health Services Monograph Series. Hospital Research and Educational Trust, Chicago, Ill., 1976.
8. Novick, L. F., and Dickenson, K.: Assessment of ambulatory care—Application of the tracer methodology. Med Care 14: 1–12 (1976).
9. Nie, N., Bent, D., and Hull, C.: Statistical package for the social sciences. McGraw-Hill Book Company, New York, 1970.
10. Kessner, D. M., et al.: Assessing ambulatory care: The use of an enriched encounter form. Health Services Research Office, Department of Community Medicine and International Health, Georgetown University, Washington, D.C., 1975.

SYNOPSIS

DEUSCHLE, JEANNE MAGAGNA (Health Insurance Plan of Greater New York), LOGSDON, DONALD N., SOLLECITO, WILLIAM, STAHL, WILLIAM, SMITH, HARRY, JR., SONNENSHEIN, MONA, and KREITZER, MARCIA: *Implementation of a peer review system for ambulatory care: Experience of the Health Insurance Plan of Greater New York. Public Health Reports, Vol. 93, May–June, 1978, pp. 258–267.*

To evaluate and improve physician performance, the Health Insurance Plan of Greater New York (HIP) initiated a peer review system in 1974. HIP is a prepaid health plan with 28 affiliated medical groups that provide comprehensive health care to 750,000 enrollees.

Under the plan's peer review system, physician committees formulate explicit medical care standards for selected tracer conditions. The data required for application of these standards are collected from patients' records by professionally supervised lay abstractors. Following analysis of data, results are reviewed and recommendations are prepared by the Peer Review Committee. These reports are discussed with the medical group physicians affiliated with HIP, who are requested to devise a plan for correction of medical care deficiencies cited in the recommendations. Between 1975 and 1977, all 28 affiliated medical groups had had their medical care reviewed for three tracer conditions—hypertension, acute otitis media, and breast lesions, for a total of approximately 7,000 cases.

Methods for solving the problems identified by peer review and for reaudit are still in an exploratory stage. However, HIP has selected an educational approach in seeking to motivate and assist the medical groups in addressing any medical care deficiencies revealed.

The cost of the HIP Peer Review Program over the first 3 years of operation was $500,000. Approximately one-half of this sum was for development of the program and one-half for its operation. In spite of the complexities inherent in a peer review system, the progress HIP has achieved with its program should be encouraging to others initiating similar programs.