Physician Preference for Criteria Mapping in Medical Care Evaluation

Sheldon Greenfield, MD, Sherrie H. Kaplan, MPH, George A. Goldberg, MD, Mary Ann Nadler, MPH, and Rosalyn Deigh-Hewertson, MPH
Los Angeles, California

This study was designed to determine which of three quality assessment methods most validly identifies deficient care. Process criteria were developed to assess outpatient care for urinary tract infection using each of three methods: a limited "list" of seven criteria, an extensive "list" of 40 criteria, and a criteria map (CM) which uses branching logic to identify applicable criteria according to the specific needs of each case. Defining deficiency as compliance with less than 60 percent of criteria, the extensive list found all 66 cases deficient; the limited list, 27 (41.0 percent); and the CM system, 15 (22.7 percent). After excluding the extensive list because of its nondiscrimination, 23 discrepancies in rating remained between the limited list and the CM. Ten physicians unaware of the results reviewed all 23 cases. In 12 of these 23 cases, at least seven of the ten physicians preferred the rating of one method over another; the CM assessment was preferred in 11 of the 12 cases ($P < .01$). Criteria maps, providing a patient-specific approach, offer a more valid assessment of medical care than either the extensive or limited list.

Comparisons of medical audit scores with measures of outcomes of care[1-3] have often failed to demonstrate positive associations between medical process and patient outcome. In such studies, the scores from measures of patient outcome have usually exceeded those for medical process. This relatively consistent finding suggests a systematic difficulty with the "measurement" of process. As recently suggested by Brook,[4] one of the possible factors contributing to the process/outcome measurement disparity is the inherent inability of a list of process criteria to reflect the patient-specific medical decisions/actions which are most relevant to that patient's outcomes. The logical processes of the physician, who sequentially collects preliminary patient data, identifies positive findings, and takes specific actions based on those findings, are excluded from evaluations which rely on a criteria list.

The authors have developed a method called Criteria Mapping[5] which, by tracking physician logic for a given problem or diagnosis, limits the actual number of criteria applied to a given patient's case to those criteria which are relevant to that case. Criteria constructed in this way have greater potential for identifying those specific medical processes that will lead to discrete patient outcomes. Results using this method have been shown to correlate with discrete outcomes for at least one problem—chest pain as evaluated in an Emergency Department.[6]

This paper reports the findings of the first in a series of studies undertaken to compare representative explicit criteria lists with criteria map-

From the Department of Medicine, School of Medicine, and the School of Public Health, University of California, Los Angeles, California. Requests for reprints should be addressed to Dr. Sheldon Greenfield, Room 42-170, Department of Medicine, UCLA, Los Angeles, CA 90024.

Table 1. Long List Criteria for Quality of Care[7] First Section*

Acute Urinary Tract Infection in the Female

A. History (specific reference to vesical dysfunction):

Score: ____ 1. Frequency of urination
____ a. Day
____ b. Night
____ 2. Obstructive symptoms
____ 3. Pain
____ a. Nature
____ b. Location
____ c. Radiation
____ 4. Hematuria
____ 5. Pattern of incontinence
____ 6. Chronology of symptomatology
____ 7. Previous urologic disease
____ 8. Previous urologic treatment
____ 9. Previous urologic instrumentation
____10. Obstetrical history
____11. Gynecologic history
____12. Medication initiated for this illness prior to contacting physician
____13. Duration of symptoms before contacting physician
____14. Previous evaluation of genitourinary system
____15. Temperature
____16. Chills
____17. History of recent sexual contact
____18. History of other recent infection

* A total of 22 additional criteria for the physical examination, laboratory, therapy, and follow-up examination constitute the remainder of this list.

ping, to determine which method more accurately measures medical process. In this study, physician judgment is used as the standard of reference in the analysis of three process assessment methods. Quality-of-care assessment methods are designed to identify cases which would be considered by physicians to be inadequately managed. A method which more closely conforms to these physician judgments of adequacy can be considered of greater utility than one which shows less conformity with physician judgment.

Methods

Setting and Problem Selection

A family practice training unit affiliated with the University of California, Los Angeles (UCLA) served as the site for the study and supplied patients' records for evaluation. Urinary tract infection (UTI) in adult females was chosen because of the frequency of occurrence in this setting, the relatively important role of primary care in the control of acute morbidity, and the potential for prevention of chronic renal disease.

The Process Assessment Methods

Three evaluation methods were chosen for comparison: two criteria lists—one an extensive, comprehensive list, the other a more abbreviated list—and a criteria map. These three methods were selected as representative of the types of process assessment measures currently in use. The first method, an extensive list containing 40 items applicable to each case, was generated by the American Society of Internal Medicine.[7] A portion of this list appears in Table 1. It is clear from inspection that this list includes items for evaluation of the most complex cases of urinary tract infection. The more economical "abbreviated list" (Table 2) contained only items applicable to *all* cases. This list was developed by the family practice unit medical staff using the California Medical Association's Patient Care Audit method, and is representative of an abbreviated list which has limited ability to evaluate complex cases.

The third method, a criteria map, was initially developed by members of the Department of

Medicine at UCLA and was subsequently modified by the family practice unit. Because of the branching, patient-specific format of these criteria (as shown schematically in Figure 1), the actual items applicable to each case varied. The map contained a total of 97 items, but an average of five criteria applied to any given case. For example, different criteria would apply if the patient had vaginal symptoms, or was pregnant, or had a previous history of urinary tract infection, or was diabetic. In these cases, a more in-depth evaluation is required and is accounted for in the criteria map. However, for uncomplicated urinary tract infection—which accounts for the majority of adult female patients with UTI presenting at a family practice unit—only a few criteria apply (Figure 2). In addition, many "options" for diagnosis and treatment are provided in the criteria map (accounting for many of the total 97 items), such as alternative confirmation of the diagnosis by "any of bacteriuria, pyuria, or a positive urine culture." These options allow for accepted variations in clinical "styles" (eg, choice of range of antibiotics, diagnostic techniques) not accounted for by a criteria list. If any of these findings are present, the subsequent criteria for treatment and follow-up apply. If none of these findings are present, the abstractor need not proceed to that section of the map which deals with the subsequent (conditional) criteria. In general, the criteria map identified various subgroups of patients according to individual clinical findings (eg, chills, fever, blood pressure ≤80/60 mmHg, positive urine cultures). Subsequent criteria are then applied only to patients with the relevant predisposing clinical findings.

When development of the three criteria sets was complete, the content of each was reviewed by the principal investigator to ensure that the standards (not the criteria themselves) required by each set were comparable (eg, positive urinalysis meant ≥10 white blood cells/high power field for all methods).

Using each of the three methods described above, the charts of 66 patients presenting to the family practice unit with either symptoms of urinary tract infection or a positive routine urinalysis were abstracted. For the purpose of this study, an individual patient score was computed for each case, using each of the criteria sets. Those cases failing to meet an arbitrary level of 60 percent of the required criteria were rated "inadequate." (This very liberal limit was chosen to allow for reasonable variation in care based on the underlying assumption that many of the criteria commonly designated for urinary tract infection are of uncertain clinical value.)[8] If, for example, a case complied with 75 percent of the California Medical Association (CMA) criteria, 50 percent of the criteria-map criteria, and 40 percent of the American Society of Internal Medicine (ASIM) criteria, it would receive, respectively, "Adequate," "Inadequate," and "Inadequate" ratings.

Table 2. Abbreviated List (Complete)

1. Mention of presence or absence of any one or all of the following:
   a. dysuria
   b. frequency
   c. urgency
2. Urinalysis showing
   ≥10 white blood cells per high power field
3. Urinalysis showing
   ≤10 epithelial cells per high power field
4. Appropriate antibacterial agent to include *only one* of the following:
   a. sulfonamide
   b. ampicillin
   c. tetracycline
   d. nitrofurantoin
   e. trimethoprim-sulfa
5. Antibacterial agent given at least 10 days
6. Repeat urinalysis requested
7. Repeat urinalysis obtained 10 to 14 days after therapy is initiated

When discrepancies in ratings between the methods occurred, an external estimate of the quality of care was used to indicate which method more accurately reflected the quality of care. It was decided not to use the outcomes of care for this usually self-limited disease, since outcomes such as symptomatic relief may be unrelated to good care for urinary tract infection (that is, they are likely to correlate poorly with process). In addition, poor outcomes, such as persistent positive culture, are infrequent, and a prohibitively large sample size would have been needed to obtain a sufficient number of negative outcomes to show a significant correlation between processes and outcomes.[8,9] Lastly, while patient outcome is clearly

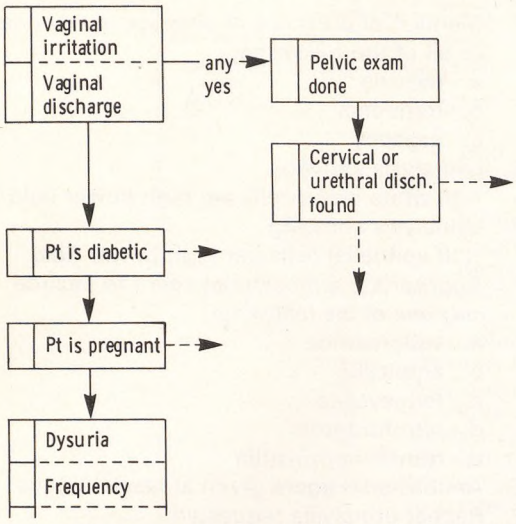

Figure 1. Boxes to the left of each item are provided for coding. Positive responses to an item lead to the next item on the right. Negative responses or "missing" information are directed down to the next vertical item. The abstractor looks only for the items to which he/she is directed by the arrows. Here, if neither vaginal irritation nor vaginal discharge were reported in the record, the abstractor follows the arrow down and does not seek information concerning the performance of pelvic examination or the discovery of cervical or urethral discharge.

one form of validation of medical process, and is currently the focus of considerable attention, it has limited feasibility as a method for validation of patient care audits. For these reasons, it was decided to validate the methods of medical process assessment by using physicians' independent evaluations of the adequacy or inadequacy of process.

Ten UCLA primary care physicians of varying ages and both sexes—six internists and four family physicians (none of whom was involved in the criteria development)—were presented with an abstract of each case with conflicting method assessments. Essentially all the information in these short ambulatory care notes was put into the abstract; abstracts were used to avoid the effect of poor handwriting and recognition of signatures. In addition, conflicting judgments regarding the adequacy of the care for that case and the corresponding reasons for the judgment rendered by each method were provided. An example of the abstract presented to the physicians is shown in Figure 3. Each physician was asked to read each case abstract and the accompanying evaluations by each method, and to decide which method most appropriately evaluated the care. The methods were labeled only "A," "B," or "C" in order to prevent easy identification. In addition, to reduce the effects of preexisting bias toward any of the methods, the physicians were not shown any of the criteria sets in advance of completing the case reviews.

Results

As illustrated in Table 3, of the total of 66 charts abstracted, using the standard of ≥60 percent as the requirement for adequate care, the long list judged all cases as having received inadequate care. The abbreviated (limited or "short") list found 41 percent of the cases to be inadequate, while the criteria map showed 22.7 percent (15 out of the 66) to be inadequate.

Because the long list failed to discriminate even minimally between adequate and inadequate care (ie, all charts were found to be inadequate), further comparisons were made only between the abbreviated list and the criteria map review results.

To determine whether using an "adequate" cutoff point of 60 percent could have been responsible for the differences in ratings by these methods, the data were reviewed for rank order correlations of method scores. No correlation between rank orders of the two methods was found, indicating that the differences found (22.7 percent vs 41 percent) were not artifacts of scaling. In addition, the data were analyzed using 50 percent and 75 percent adequacy cutoffs; use of these levels did not change the results significantly.

The review procedure is summarized in Figure 4. There were 23 cases in which the two methods, mapping and the abbreviated list, gave opposite ratings as to the adequacy of care. These 23 cases were then subjected to physician review. The results of the physician review of these 23 cases are presented in Table 4. In a total of 12 cases, physicians showed a definite preference for one method over the other, as evidenced by a high degree of concordance, ie, at least seven physicians agreeing with the rating of one method over the other for any case. Of these 12 cases, method A (criteria mapping) was chosen 11 times. In only one in-

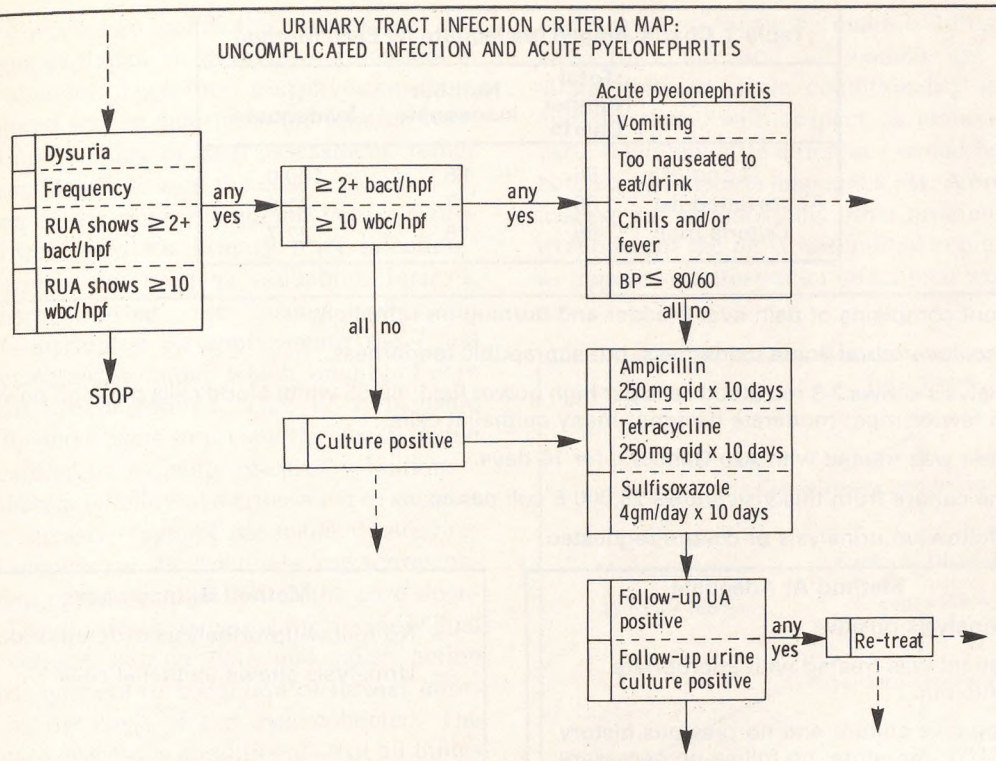

Figure 2. Boxes to the left of each item are provided for coding. Positive responses to an item lead to the next item on the right. Negative responses or "missing" information are directed down to the next vertical item. "STOPs" are placed after items for which subsequent items would exceed the limits of the criteria map. The abstractor looks only for the items to which he/she is directed by the arrows. Therefore, for any individual case, only a fraction of the available criteria is applied.

stance of definitive physician preference was method B (the abbreviated list) chosen over method A. If the two methods were equally likely to be preferred, the chance of seeing a preference this extreme is less than 1 in 100 ($P < .01$), by a test of proportions.

Examination of these 11 "highly preferred" cases revealed that 8 of the 11 either had complications or unusual presentations, all of which could be explained or included in the various branches of the criteria map, but could not be captured by the list. Five had unusual presentations such as lower abdominal pain, dyspareunia, nocturia, etc. Two had a history of urinary tract infection requiring more investigation and follow-up than a list could fully encompass, and one had a persistent infection which was documented by a positive culture and required follow-up. Criteria map assessment of the case illustrated in Figure 3 did not include follow-up urinalysis since the initial culture was negative.

For the remaining 11 cases of the 23, the physicians made no definitive choice between either of the methods. Review of these records revealed straightforward, uncomplicated cases in which physician choice would be expected to vary according to individual preference for specific criteria. In these cases, additional qualifying clinical information usually provided by a criteria map would not contribute to the adequate/inadequate decision.

To determine whether the choice of method was affected by the individual physician's tendency to judge cases consistently as adequate or inadequate, the physicians' choice of method was analyzed by method rating (adequate or inadequate) for each case. Physicians were just as likely to choose the criteria map assessment (method A) for a case which had been rated adequate as for one which had been rated inadequate. Similarly, physicians chose the list method as often when it judged a case adequate as

Table 3. Charts Judged Inadequate by Each Method

	Total Number Charts	Number Inadequate	% Inadequate
Long list	66	66	100.0
Abbreviated list	66	27	41.0
Criteria map	66	15	22.7

— Patient complains of pain over bladder and burning on urination.

— No costovertebral angle tenderness, but suprapubic tenderness.

— Urinalysis shows 2-3 red blood cells per high power field; 40-55 white blood cells per high power field with few clumps; moderate bacteria; many epithelial cells.

— Patient was treated with Azo Gantrisin for 10 days.

— Urine culture from this visit shows 25,000 E coli per cc.

— No follow-up urinalysis or culture requested.

Method A: Adequate

— Urinalysis positive

— Patient was treated with appropriate antibiotic

— Negative culture and no previous history of UTI; therefore, no follow-up necessary

Agree____

Method B: Inadequate

— No follow-up urinalysis ordered or done

— Urinalysis shows epithelial cells

Agree____

Figure 3. Case #213—Medical Record Abstract

when it judged it inadequate. The criteria mapping method was preferred over the list method regardless of rating.

Discussion

This study was concerned primarily with the relative accuracy with which a method could assess the quality of care rendered. Additionally, there was interest in determining whether a method could, by increased accuracy of evaluation, reduce the need for physician review of cases.

Consequently, attention was focused on the 23 discrepant judgments (35 percent). In those 11 cases (47.8 percent of the 23) for which the physician choice was discernible—that is, those cases where there was clear consensus—the explanation for that choice (as indicated by participating physicians) was that the method chosen (ie, criteria mapping) more completely explained the clinical details of the case. That is, when the cases were more complex, criteria mapping provided a more valid assessment of the care. Where the cases were relatively more simple and straightforward, there was no basis for either system to provide a more discriminating evaluation, and therefore physicians had no strong preference for either method.

The choice of method was not based on whether the method resulted in an adequate or inadequate rating; preference for criteria mapping was consistent regardless of the rating this method rendered. Subjective opinions or implicit judgments about quality of care in the absence of objective evaluations have proven to be unreliable in the past.[10] One factor which may have contributed to the variability of these judgments is that each physician-judge in an implicit review may apply his or her own unique standards of care to the evaluation effort. In an attempt to control this problem, physicians in this study were specifically asked to

choose between two conflicting ratings and accompanying explanations in reviewing the data in the case abstract. Therefore, the physician judgment required was to determine the relative completeness and validity of each assessment, rather than independently to rate the case.

Because of the poor correlation between process and outcomes for urinary tract infection,[8] outcomes were not used as validation. Instead, this approach—that of using physician judgment—attempted to approximate the "real life" circumstances under which quality-of-care evaluations are performed. That is, it is usual practice for physicians to review the questionable cases identified by a quality-of-care evaluation.

The inherent inability of a criteria list to account for the progressive logic of the medical care process can account for the relatively low scores obtained from applying either list to the care documented. The physician, in caring for the individual patient, collects certain data and takes action (either management or collection of further information) on the basis of the data collected. The medical care process is conditional—not all things are done for all patients with the same disease or complaint. The single explicit list, intended for application to all cases of a specific disease or diagnosis, fails to account for this conditionality. Thus, routine cases requiring only a few criteria fail to comply with extensive lists of criteria designed to account for very complicated cases; similarly, more abbreviated lists designed to apply a limited number of criteria to all cases fail to provide a meaningful evaluation of the more complicated cases. It is not surprising, then, that low performance scores are frequently found when lists are used to evaluate medical care. However, the criteria map applies only relevant criteria, in sequential fashion, to an individual case, so these criteria are more patient-specific than disease-specific. It would be expected that when these conditional, decision-oriented criteria are used to measure care, a more valid assessment of care results.

The choice of urinary tract infection for a comparison of these two methods put the criteria mapping method at a considerable disadvantage. The strength of criteria mapping lies in the measurement of care of those diagnoses which have multiple divergent subgroups of patients, more complicated cases, or more options for diagnosis and treatment. Since the majority of cases of urinary tract infection in women are relatively straightforward, little conditionality is required and therefore, with respect to ability to assess care, relatively little difference would be expected between the criteria map and a list. A comparative assessment of almost any other problem (with the exception of the most self-limited conditions such as upper respiratory tract infections) would be ex-

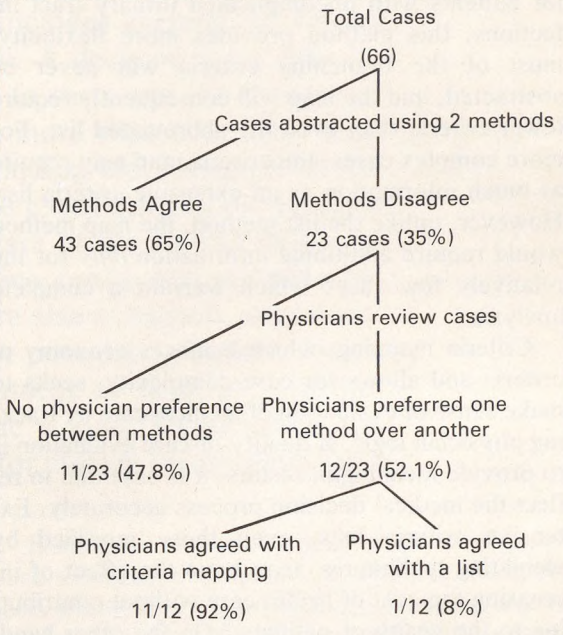

Figure 4. Review procedure. The methods disagreed in a substantial percentage of cases (35 percent). When physicians preferred one method's decision over the other, they nearly always agreed with the criteria mapping decision. Thus, criteria mapping comes closer to judging quality of care in the same way as physicians would judge it.

Table 4. Method Choice in Cases with High Concordance (≥7/10 Physicians)

	Cases with High Concordance
Method A	11
Method B	1
Total	12
Total Cases	23

pected to show even greater differences between evaluations of the methods. Nevertheless, the fact that the criteria mapping method was consistently preferred in the assessment of this problem suggests that even for simple problems, medical practice varies sufficiently to require flexibility in the evaluation of care.

Although criteria mapping may appear to be more involved, it is neither more complicated nor more extensive than a criteria list. It is only as complex as the individual case requires. Indeed, for patients with uncomplicated urinary tract infections, this method provides more flexibility: most of the branching criteria will never be abstracted, and the map will consequently require fewer criteria than even the abbreviated list. For more complex cases, the criteria map may require as much information as an extensive criteria list. However, unlike the list method, the map method would require additional information *only* for the relatively few cases which warrant a complete analysis.

Criteria mapping, which balances economy of criteria and allows for case complexity, seeks to make sense of evaluation of medical care by tracking physician logic. If quality-of-care evaluation is to provide meaningful results, it is essential to reflect the medical decision process accurately. Extensive criteria lists, even those modified by weighting techniques, may have the effect of increasing the cost of health care without contributing to the health of patients.[2] On the other hand, abbreviated lists may not provide enough of the essential information to permit a valid evaluation of medical care, and as a result many records may require subsequent (and costly) physician review.

Current trends in quality-of-care assessment reflect increased awareness of the need to incorporate the idea of logic or conditionality into assessment measures. Some studies have accounted for conditionality by subcategorizing patients into basic/inclusive subgroups (ie, diabetics, individuals over 40 years of age, males/females, etc).[11] Criteria mapping has attempted to incorporate both broad subgroupings of patients *and* clinical variations in patient presentations in the formulation of a method capable of accurate and efficient medical care evaluation.

It may be concluded that even for an uncomplicated, rather standard outpatient problem with minimal recording, the criteria mapping approach is as feasible, and more discriminating, than either a simple or a complex list. It corresponds better with the actual process of medical care, offers an alternative to the more rigid and less satisfying list, and shows potential for ultimately narrowing the gap between process and outcome measures.

Acknowledgement

Supported by grant number HS 01320, from the Center for Health Services Research and Development, United States Public Health Service. The authors wish to thank Dr. Shan Cretin, Dr. Robert H. Brook, Ms. Linda Worthman, and Ms. Nancy Solomon for careful review of the manuscript. We are also grateful to Dr. Charles E. Lewis for continuing support and advice.

References

1. Nobrega FT, Morrow GW Jr, Smoldt RK, et al: Quality assessment in hypertension: Analysis of process and outcome methods. N Engl J Med 296:145, 1977

2. Brook RH, Appel FA: Quality-of-care assessment: Choosing a method for peer review. N Engl J Med 288:1323, 1973

3. Romm FJ, Hulka BS, Mayo F: Correlates of outcomes in patients with congestive heart failure. Med Care 14:765, 1976

4. Brook RH: Quality—Can we measure it? N Engl J Med 296:170, 1977

5. Greenfield S, Lewis CE, Kaplan SH, et al: Peer review by criteria mapping: Criteria for diabetes mellitus. Ann Intern Med 83:761, 1975

6. Greenfield S, Nadler MA, Morgan MT, et al: The clinical investigation and management of chest pain in an emergency department: Quality assessment by criteria mapping. Med Care 15:898, 1977

7. Hare RL, Barnoon S: Medical care appraisal and quality assurance in the office practice of internal medicine. American Society of Internal Medicine, National Center for Health Services Research and Development, US Department of Health, Education, and Welfare, HSM 110-70-420, 1973

8. Lindsay MI, Hermans PE, Nobrega FT, et al: Quality of care assessment: Part I: Outpatient management of acute bacterial cystitis as the model. Mayo Clin Proc 51:307, 1976

9. Brook RH, Davies-Avery AD, Greenfield S, et al: Assessing the quality of medical care using outcome measures: An overview of the method. Med Care 15(Suppl):1, 1977

10. Richardson FM: Peer review of medical care. Med Care 10:29, 1972

11. Jacobs CM, Jacobs ND: The PEP Primer: The JCAH Performance Evaluation Procedure for Auditing and Improving Physician Care. Chicago, Quality Review Center, Joint Commission on Accreditation of Hospitals, 1974