
Chapter 7

Evaluation of Physicians' Performance: Care for Hypertension

CONTENTS

	<i>Page</i>
Introduction	145
Evaluations of the Outcomes of Care for Hypertension	148
Reliability of the Indicator	149
Validity of the Indicator	149
Feasibility of Using the Indicator	150
Evaluations of the Process of Care for Hypertension	151
Reliability of the Indicator	152
Validity of the Indicator	153
Feasibility of Using the Indicator	156
Conclusions and Policy Implications	158

Box

Box	Page
7-A. The Process of Medical Care for Hypertension	151

Figure

Figure	Page
7-I. The Process of Medical Care for Hypertension	147

Tables

Table	
7-1. Studies on Care for Hypertension Reviewed by OTA.	148
7-2. Prognostic Factors for Case-Mix Adjustment of Hypertension	
Outcome Data	150
7-3. Potential Sources of Information on the Process of Patient Care	152
7-4. Comparison of Generic Approaches to the Evaluation of Physicians' Performance: Care for Hypertension	159

Evaluation of Physicians' Performance: Care for Hypertension

INTRODUCTION

A major approach to assessing a physician's performance, especially since the 1950s, has entailed evaluation of the care provided for specific medical conditions (184,371). This approach has spread widely during the past two decades as researchers and clinicians have refined assessment techniques. Physicians and other medical professionals have increasingly participated in the review of their peers' performance through privately sponsored activities of hospitals, health maintenance organizations, group practices, medical associations, and third-party payers and through publicly funded programs of State and Federal governments.

This chapter examines the reliability, validity, and feasibility of using evaluations of physicians' performance in caring for a particular condition as an indicator of physician quality. Hypertension is used as a case study condition. Elevated blood pressure is one of the most prevalent and costly medical disorders in the U.S. population, and the effective detection and management of hypertension is one of the Nation's chief public health goals (372,662). Since about 30 percent of the U.S. population is hypertensive,^{*} an evaluation of the methods used to assess care for hyper-

tension and to provide information on its quality is important in itself. But evaluating care for hypertension may also illustrate a number of key considerations relevant to evaluating care for other conditions. At the same time, evaluation of the quality of care provided by a physician for hypertension might provide some insights into the quality of other aspects of a physician's services. Consequently, this case study provides a vehicle for analyzing many broader issues in evaluating the process of medical care.

The process of medical care for hypertension is outlined in box 7-A and figure 7-1. In borderline as well as more severe cases, hypertension is generally asymptomatic; its diagnosis depends on the use of blood pressure measurements in individuals who may appear well or who may be seeking care for unrelated health problems. In over 90 percent of cases, hypertension cannot be attributed to an identifiable pathologic cause and must be treated on a chronic, lifetime basis. Detection and followup are crucial, because long-term sequelae of uncontrolled hypertension include serious morbidity associated with strokes, renal disease, cardiac dysfunction, and increased risk of premature death (89). The efficacy of therapies designed to reduce blood pressure toward desired levels in significantly reducing the incidence of these complications was demonstrated in Veterans Administration trials in the early 1970s (676,677). The Hypertension Detection and Follow-Up Program, a 5-year randomized clinical trial with over 10,000 participants, found that a systematic "stepped-care" program for treatment to reduce high blood pressure was associated with significantly higher rates of pressure control and 5-year survival than was "usual" management (313).

OTA's selection of care for hypertension for analysis in this report was based in part on the

¹ Estimates of the prevalence of hypertension depend on the precise definition of hypertension adopted. On the basis of the outcome findings of large randomized controlled trials, the Joint National Committee on Hypertension has recommended that patients be diagnosed as hypertensive if the average of blood pressure measurements taken on at least three successive occasions is greater than or equal to 140 mmHg systolic over 90 mmHg diastolic (332). This definition represents a stricter standard than the previous one, which involved repeated measurements above 160/95. Some variation in the specific cutoff pressure levels used for patients in the "mild" hypertensive category still exists among clinicians, especially outside the United States (47). Further, some specialists have argued for diagnosing patients with isolated systolic hypertension as well (717). Because elasticity of the major arteries declines with age, the combined prevalence of isolated systolic and diastolic hypertension for persons aged 65 to 74 is estimated at 64 percent overall and 76 percent in blacks.

Box 7-A.—The Process of Medical Care for Hypertension

Medical care for hypertension—including screening for the disorder and managing therapy for it—is an example of medical care for a specific condition and can be described in terms of the spectrum of medical care presented in chapter 3. There is a high degree of consensus regarding the value of widespread population screening and patient adherence to therapies designed to control elevated blood pressure. Consequently, the basic clinical sequence for effective case finding, diagnosis, and management is especially well defined for hypertension (89,569). This sequence is illustrated in figure 7-1. The figure also notes the many possible stages at which inadequate access to care, discontinuities, and patient dropout can result in care failures and thus poor quality.

Appropriate case-finding procedures are particularly important for two reasons: because general preventive measures for essential hypertension have not been established, and because the disease is both asymptomatic and highly prevalent. The target population for case finding, via standard blood pressure measurements documented at least every several years, is the adult population. Confirming the diagnosis of hypertension requires repeated elevated measurements, taken in different limbs, on each of at least two subsequent visits. This requirement before initiating antihypertensive treatment is a consequence of the frequency of isolated hypertensive readings resulting from stress, daily variations, measurement errors, or other transient causes.

Patients whose diagnosis of hypertension is confirmed represent the target population for management, which involves treatment and followup. Although details may vary among clinicians, treatment typically consists of behavioral modifications coupled with drug therapy. The former include diet modifications to reduce obesity and sodium intake, exercise, cessation of smoking, reduced use of alcohol, and steps to reduce stress, each tailored appropriately to the individual case. Pharmacologic therapy has traditionally featured a “stepped-care” regimen in which more powerful medications are administered incrementally as the patient fails to achieve blood pressure control at a given level (313). These drugs include diuretics, beta blockers, and vasodilators.

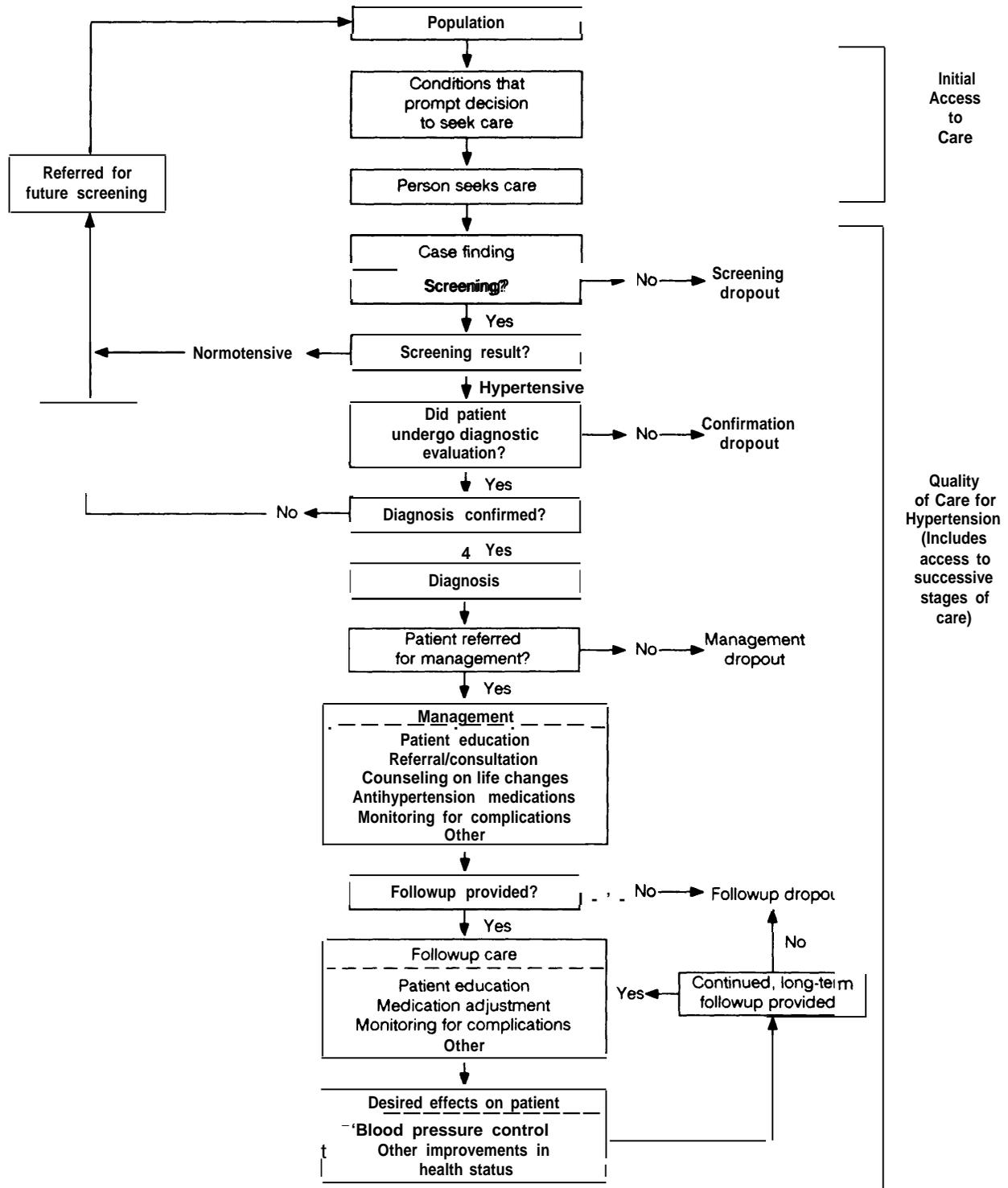
The use of stepped-care for certain subgroups of hypertensive patients is currently controversial. These subgroups include patients with mild hypertension (diastolic blood pressure 90-95 mmHg) and patients for whom a particular pathophysiologic mechanism more amenable to an alternative type of medication is suspected (425). The controversy is confined largely to mild hypertensives (and thus is related to controversy in defining hypertension) and to the choice of particular drugs. Broader issues are sufficiently well resolved to permit the elucidation of guidelines for appropriate care.

Because essential hypertension is a chronic condition requiring lifetime treatment, hypertensive patients generally receive care on an ongoing ambulatory basis unless evidence of acute pathological complications supervenes. These complications include strokes, renal disease, visual disorders, or severe headaches. Followup is crucial in management, because patients must adhere consistently to a set of potentially unpleasant behavioral and medical recommendations for many years.

fact that the efficacy of antihypertensive therapy has been well demonstrated and that there exists a fundamental clinical consensus on its effectiveness. The demonstrated efficacy of generally accepted procedures supports the validity of basing quality assessments on adherence to the procedures. Technical aspects of care for hypertension are critical to case finding and management; examples include appropriate screening and diagnostic procedures, proper drug prescriptions, and patient followup for monitoring the effects of ther-

apy and the possible development of complications. It is important to recognize, however, that interpersonal aspects of care for hypertension may be just as important as the technical aspects: hypertensive patients must be persuaded to comply with their medication schedule in spite of unpleasant side effects (196), lifestyle changes may be necessary, and behavior modifications must be maintained. Clearly, both technical and interpersonal aspects of care for hypertension must be considered in evaluating quality. Further, hyper-

Figure 7-1. – The Process of Medical Care for Hypertension



tension is generally diagnosed and managed by a physician in an ambulatory setting rather than in a hospital. Its treatment thus depends on a major segment of health care providers that many of the other potential indicators of quality evaluated in this report do not address.

Drawing on published and unpublished studies (see table 7-1),² this chapter analyzes the reliability, validity, and feasibility of using evaluations of physicians' care for hypertension as an indicator of quality. Two generic approaches may be used to evaluate physicians' care:

- evaluations of patient outcomes, and
- evaluations of the process of medical care through the use of explicit criteria or implicit judgment.

The reader should recall that hypertension is only an example and that many of the same concerns that arise may apply to evaluations of physicians' care for other conditions. Clearly, some issues transcend the specific case of evaluating care for hypertension. What are the advantages and disadvantages of using patients' medical records as the source of data for assessments of the process of care? And how can aspects of care that are poorly reflected in medical records best be evaluated? How can physician involvement, and thus medical expertise, be incorporated most effectively into evaluation techniques? How should specific criteria and standards be developed and applied to evaluate physicians' performance? Do evaluations of the process of care need to adjust for differences among patient groups, in disease

²Additional details on the studies reviewed can be found in OTA's technical working paper "Hypertension Screening and Management as an Indicator of Quality: Reliability, Validity, and Feasibility Issues" (415).

Table 7-1.—Studies on Care for Hypertension Reviewed by OTA^a

Assessments of patients' outcomes:
Brook, 1973 (99)
Schroeder and Donaldson, 1976 (557)
Shorr and Nutting, 1977 (569)
Fletcher, et al., 1979 (211)
Hulka, et al., 1979 (309)
Dove and Schneider, 1980 (188)
Keeler, et al., 1985 (343)
Assessments of medical process, implicit criteria:
Brook, 1973 (99)
Hulka, et al., 1979 (309)
Hastings, et al., 1980 (284)
McAuley and Henderson, 1984 (410)
Assessments of medical process, explicit criteria:
Brook, 1973 (99)
Shorr and Nutting, 1977 (569)
Hulka, et al., 1979 (309)
Deuschle, et al., 1982 (174)
Nutting, et al., 1982 (468)
Sheps and Robertson, 1984 (567)
Borgiel, et al., 1985 (86)
Keeler, et al., 1985 (343)
Combined assessments of patients' outcomes and medical process:
Palmer, 1983 (475)
McCoy, et al., 1988 (417)

^aNumbers in parentheses refer to numbered entries in the reference list at the end of this report.

SOURCE: Office of Technology Assessment, 1988.

severity or otherwise? Is the quality of care provided for one condition at all indicative of the overall quality of a physician's practice, or are no such generalizations possible? Most importantly, how can data obtained using these evaluative techniques be appropriately and effectively translated into information useful to consumers? The following analysis discusses these issues in the context of hypertension, but similar issues arise in any attempt to assess physicians' performance by evaluating the care rendered for a specific condition.

EVALUATIONS OF THE OUTCOMES OF CARE FOR HYPERTENSION

The most widely used measure of patients' outcomes in hypertension studies is a reduction in blood pressure levels or hypertension control rates; this is a proxy measure for longer term clinical complications. Actual measures of complications include specific morbidity rates and mortality differences associated with poor control of

hypertension. Few studies of patient outcomes have incorporated functional considerations or other measures related to the patients' quality of life. Further, few studies have based quality-of-care comparisons among different provider groups exclusively on outcome measures.

Reliability of the Indicator

The procedure used to measure blood pressure is a rapid and accurate procedure when performed by trained personnel. But single measurements of an individual's blood pressure often correlate poorly with that individual's typical blood pressure. Consequently, high false-positive rates (343) and false-negative rates (557) of hypertension have been reported when single measurements are used.³

In assessing physicians' performance, quality assessors use blood pressure readings noted in patients' medical records. This approach has the disadvantage of relying on outcome data provided by the physician practice being evaluated rather than by a more independent source (475). Most studies do not provide explicit or quantitative information concerning the reliability of these recorded measurements, because the procedure for measuring blood pressure is technically accurate when performed by qualified personnel and because a series of readings from successive visits is typically reported,

If blood pressure data are grouped into different outcome classes reflecting adequacy of control based on clinical consensus (309), variations in definitions of hypertension may reduce the comparability of results obtained, with identical measurements being categorized differently. Reflecting disagreements among expert judgments, this problem pertains especially to whether diastolic pressures in the borderline 90-9s mmHg range are considered controlled. If such expert classifications are used, a uniform system is required across all providers for reliability.

More innovative approaches to outcome assessment can create special reliability problems. For example, relying on judgments by a panel of experts as to whether a patient's outcome is "improvable" or "unimprovable" requires consideration of the same interrater and interrater reliability issues that arise in process measures (99).

³For this reason, the clinical diagnosis of hypertension requires elevated pressure recordings on successive visits, and perhaps several readings on each visit.

Typically, however, such problems arise only in assessment methods using implicit judgments of experts.

Validity of the Indicator

The use of blood pressure readings as a measure of the outcome of care for hypertension is intelligible to average consumers, because such readings are the clinical parameter with which hypertension case finding and management are ultimately concerned.

Even an outcome as immediate as blood pressure readings, however, is the result of a broad range of personal and environmental factors, many of which are beyond the influence of a physician's care. This validity problem can be corrected through standardization of a physician's patient mix based on relevant prognostic factors for desirable or undesirable outcomes. Such methods are analogous to the severity-of-illness adjustments described for patient characteristics in hospital mortality data (see ch. 4). Patient age and other variables that various studies have found to correlate significantly with blood pressure control are listed in table 7-2. Although only one of the studies listed in that table had the statistical power of a large prospective randomized trial (343), the studies collectively indicate that factors as such as the patient's age, race, initial blood pressure, weight, compliance with the prescribed regimen, and access to care can be used to help standardize outcomes across different patient samples.

Although statistical manipulations can increase the validity of the assessment results, a substantial portion of the observed variations in outcomes remains unexplained. Can this portion be attributed exclusively or primarily to the quality of physician care? In general, outcome measures provide little insight into what particular steps a provider may be taking—among a universe of uncontrolled factors in the long-term treatment of a chronic illness—that have a significant impact on the outcomes. This difficulty of attribution is a central problem for an assessment of quality that relies purely on outcomes. Consequently, most assessments use some type of process measure or combine process and outcome approaches (569).

Table 7.2.—Prognostic Factors for Case-Mix Adjustment of Hypertension Outcome Data

Study ^a	Significant factors
Keeler, et al., 1985 (343)	Initial blood pressure Age Sex Race Location
Dove and Schneider, 1980 (188)	Initial blood pressure Presence of alcohol abuse Weight No treatment in other clinics
Fletcher, et al., 1979 (211)	Lower age Initial blood pressure Patient compliance Prescription of certain medications
Nobrega, et al., 1977 (465)	Initial blood pressure Weight Age

^aNumbers in parentheses refer to numbered entries in the reference list at the end of this report.

SOURCE: Office of Technology Assessment, 1988.

These conclusions are based on a small number of mostly nonrandom and retrospective studies, a situation that limits analysis of the validity of hypertension outcome measures. The validity of any construct for measuring quality of care depends on the extent to which a statistically significant causal relationship exists between the specific processes performed by the physician and the ultimate patient outcomes observed (185). As in many other areas of quality assessment, additional well-designed studies are required for more powerful conclusions about the use of outcome data. Most importantly, further analyses of external factors that significantly influence observed outcomes for a physician's patients are necessary to develop valid adjustments for severity of illness and other patient characteristics.

Feasibility of Using the Indicator

The advantages of using blood pressure control rates or a related outcome to evaluate the quality of care for hypertension are similar to those of using hospital mortality rates to evaluate hospital care. Both are globally oriented measures, subsuming many aspects of care (and much else as well). With the strong emphasis on out-

comes in the general population, these measures are also relatively easy for the public to understand.

But many serious disadvantages accompany these measures. Mortality and morbidity rates for surgical and other inpatient procedures can be computed from data obtained over a relatively bounded time frame (e.g., 30 to 60 days after an operation), but the chronicity of hypertension may require data collection over years for valid assessments of management and control. Extended followup periods present practical methodological problems (557).

Another set of difficulties relates to the feasibility of using patients' medical records.⁷ A patient's medical record typically contains the most complete information available on the process of technical care and associated outcomes for patients in both hospitals and ambulatory facilities. It is also the legal record of care, and hospital medical records have been used extensively in evaluations of the quality of inpatient care (185).

The first potential obstacle to medical record review is that medical providers must agree to participate in the review. All experimental assessments have involved voluntary participation, with reported participation rates ranging from 30 percent to over 80 percent. Factors enhancing participation rates include persuasion by colleagues (84) or the involvement of physicians within the practice organization in the assessment process and its treatment as a team effort with constructive goals rather than as an adversarial process (99). Presumably, other incentives or compulsions could also enhance participation. Some studies have noted that physicians who have not been board certified or who are members of smaller practice groups are more likely to refuse to participate; this situation raises questions about the representativeness of the results obtained from these studies (309).

Another group of problems concerns practical issues in collecting data from records for ambulatory patients. Obstacles such as indecipherable

⁷The alternative way to develop a similar data stream is through ongoing independent collection of blood pressure measurements, a method that is expensive and logistically difficult.

handwriting and unretrievable records vary significantly by site and practitioner (479). Neuhaus and colleagues identified three types of difficulties in the data collection process: 1) obtaining a listing of all patient visits by diagnosis, 2) finding charts, and 3) dealing with miscoded or unretrievable records (459). Obtaining a list of visits by diagnosis was impeded by the absence of a uniform method of coding diagnoses, by the fact that practitioners generally did not order their records by diagnosis, and by the need to obtain drug listings from pharmacies in some cases. High miscoding rates may have resulted from clerical recording errors, the listing of a single diagnosis when several were under consideration, and the fact that a hypertension diagnosis may not be confirmed on repeat visits. Neuhaus and colleagues also noted that a pilot study of the office practice being assessed could estimate the amount and type of oversampling required to get an adequate number of "complete" cases for analysis (although these oversampled cases might not be representative).

Technical progress in the management of data bases and other information systems for recording and retrieving patients' medical records is making such records an increasingly useful source of information on physicians' performance. But as a consequence of current problems, cheap and reliable access to data from all providers remains

only a possible goal for the future. Moreover, not only has there been less research using patients' records for ambulatory care than for inpatient care, but also ambulatory records are more likely than inpatient records to be too incomplete to serve as an adequate data source. The consistent of medical record quality tends to be greater for large multiprovider organizations with computerized data bases, but most ambulatory care is delivered in small practices where recordkeeping quality may be much more uneven (475). Blood pressure and some key patient characteristics useful for severity-of-illness adjustments, however, are objective findings that are more likely to be recorded regularly than many details of the medical care process (475).

Although consideration of patient outcomes is obviously an important component of any review of the quality of care for hypertension, relying on blood pressure measurements alone—however easy to abstract from patients' records in comparison to elements related to the process of medical care—would probably require some type of independent auditing mechanism to confirm the accuracy of recorded measurements. Moreover, this approach would not directly encourage better adherence by physicians to effective case finding, diagnosis, and management for all hypertensive patients. That goal requires evaluations of the process of care.

EVALUATIONS OF THE PROCESS OF CARE FOR HYPERTENSION

All evaluations of the process of medical care involve the application of quality standards by experts (184). The types of criteria used in process evaluations span a continuum from purely explicit criteria (completely specified checklists) to purely implicit criteria (unstructured expert analysis). Between these extremes are many possibilities, e.g., the use of explicit guidelines for implicit evaluations by medical experts (284) or the use of a limited set of explicit criteria to target cases likely to be unsatisfactory for implicit review by medical experts (475).⁵ To a considerable extent,

the strengths and weaknesses of various approaches to evaluating the process of care can be analyzed in terms of trade-offs in reliability, validity, and feasibility along this implicit/explicit continuum.

Any evaluation of the process of care requires a data source that can provide adequate information on the processes used in the delivery of care. Possible sources of information are listed in table 7-3 (475,716). of the sources listed, only sets of medical records—containing histories of case

⁵An alternative approach could measure the percent of patients that complete each state of the treatment sequence (see fig. 7-1).

This approach would allow access issues to be incorporated into the assessment (569).

Table 7=3.—Potential Sources of Information on the Process of Patient Care

Sources that rely on data collected by providers	Sources that require independent collection of new data
Medical records	Patient interviews
Prescription records	Patient assessments
Claims forms	Taping/videotaping of patient encounters
Appointment books	Direct observation by experts
Patient tracking systems	"Simulated" patients ^a

Incident reports

assessors are trained to give a standardized presentation of a clinical problem and (undetected) to evaluate a physician's management of the condition (716).

SOURCE: Office of Technology Assessment, 1966; adapted from R.H. Palmer, *Ambulatory Health Care Evaluation/Utilization: Principles and Practice* (Chicago, IL: American Hospital Publishing, Inc., 1963).

management recorded by the health providers involved—are usually detailed enough and accessible enough to be used for evaluating care for specific conditions, such as hypertension.^b

Reliability of the Indicator

Variations in judgment over time or among physicians represent an obvious problem for a method of quality assessment that uses relatively unstructured expert opinion. Thus, implicit evaluations of the process of medical care must address reliability issues (99). Low interrater reliability may result from systematic bias, with some raters having an inherent tendency to rate cases more stringently than others. These variations can be moderated by adjusting the results statistically to obtain identical mean scores among reviewers (309).

Alternatively, reviewers may simply have different expectations or standards. Various steps can be taken to reduce these interrater differences: selecting physicians who are motivated to participate in the quality review or who are experienced in such assessments, including them in the development of the study, providing clear instructions and guidance, and preparing and distributing case summaries to minimize "nonreviewer" sources of variability (309). Indeed, one observer cites studies indicating that although physicians untrained in abstracting and evaluation have interrater reliability scores approaching 50 percent (the same as pure chance), training physicians in peer review and training abstracters to extract explicit infor-

^aOf course, medical records have a number of limitations, as described in the preceding section and below.

mation from records is "reliable and rapid" and results in substantial reliability gains (479). More rigorous studies report that complete agreement among reviewers occurred in 70 to 80 percent of the judgments (99,309,518); less rigorous studies usually obtain higher rates. Findings regarding intrarater reliability have been somewhat more divergent, but generally show slightly higher consistency (e.g., 85 percent in Brook's study).

Richardson concluded that 16 to 28 judges would be required to obtain a reliability of 95 percent for expert evaluation of a given case (518). Brook noted, however, that "unsatisfactory" judgments by two judges indicated that the record involved had a comparable probability of reflecting unsatisfactory care, although only some 20 percent of unsatisfactory cases would be detected (99). Thus, identical judgments by several reviewers may be adequate for detecting unsatisfactory care with a high degree of specificity, but the sensitivity of implicit review methods for identifying particular cases of inadequate care is more questionable.⁷

In explicit evaluations of the process of care, the criteria used in the evaluation are specified in more or less detail, and the reviewer need only determine whether items meeting the criteria are present in the medical record. Consequently, in studies using explicit criteria, high reliability tends to be reported if the reliability issue is addressed at all. A finding well above 90-percent concordance between abstracts by different reviewers, or between staff auditors and project directors, is typical (309,569). More general or nebulous criteria items tend to result in lower reliability (86), and failure to note items present in the record (false negatives) seems more prevalent than crediting items not present (569). Use of physician auditors is not essential for achieving high reliability in explicit evaluations; however, reliability may be significantly enhanced by using reviewers who are familiar with medical terminology and

⁷Specificity and sensitivity are statistical measures relating to accuracy. In this case, specificity represents the proportion of actual cases of satisfactory care that are identified as "satisfactory" (true negative rate). Sensitivity reflects the proportion of actual cases of unsatisfactory care that are identified as "unsatisfactory" (true positive rate). Generally, increasing sensitivity in a measurement system reduces specificity, and vice versa.

reading medical records (e.g., nurses or graduate students), by providing training sessions, and by conducting periodic reliability checks (309). More limited data on the consistency of a physician's recordkeeping across cases, and thus on the reliability that the provider will consistently record specific process items, yield results that are not quite as impressive but are encouraging (309). Few data on intrarater reliability in explicit evaluations of the process of care are available, probably because interrater reliability is reported to be so high.

The high reliability of evaluations using explicit criteria suggest that steps to make implicit reviews of the process of care more explicit may increase reliability. For example, physician reviewers might be asked to comment explicitly on the basis of their judgments (309). Alternatively, guided criteria for implicit judgments might be developed, such as a checklist to guide reviewers' evaluations of patients' records (284). Even with use of the checklist, however, interrater reliability remained within the range typical for implicit evaluations.

A combination method reported by Palmer used both explicit and implicit approaches (475). This method involved using a small number of straightforward explicit criteria, with which 100-percent compliance was expected. Medical records not in full compliance with these explicit criteria were submitted for implicit judgments on whether the care provided was satisfactory or unsatisfactory. Screening with simple explicit criteria ensures that selection of cases for possible poor-quality care has high reliability.

Explicit evaluations thus have substantial advantages in reliability compared with implicit evaluations, particularly when appropriate steps are taken to promote it. Even though some activities may also enhance the reliability of implicit reviews, these evaluations have substantially less impressive reliability results, especially for the accurate detection of a high proportion of the cases with poor-quality care.

Validity of the Indicator

Despite problems with reliability, review of the process of care by medical experts using implicit criteria is intuitively valid to average consumers,

provided that the medical experts involved have acceptable qualifications. The use of explicit criteria have been criticized as invalid because such criteria do not reflect adequately patient heterogeneity and the complexities inherent in clinical practice (309). The use of medical experts theoretically permits clinical insights and consideration of all relevant factors contributing to the management decisions for a specific patient. The severity of a patient's illness, appropriate management of concurrent conditions, and other important elements may be difficult to assess properly with explicit criteria. For this reason, Donabedian concludes that current methods for assessing physician performance using explicit criteria are not substitutes for this comprehensiveness: "For though 'peer review' of the entire record of performance (whether of process alone, or process and outcome combined) is open to error and abuse, as we all recognize, there is nothing we now have that can handle better the entirety of practice in all its rich variety and detail" (184).

Obviously, setting criteria or standards for evaluating the process of care is critical for the authority of an explicit evaluation. Various methods have been used to obtain guidelines applicable to evaluations of care for hypertension. These methods involve variations on either deriving criteria from standards published in the internal medicine literature or developing criteria through some kind of clinical consensus process. One method, for example, involved submitting lengthy questionnaires to two panels of clinicians, generalists, and specialists, and adopting the criteria approved by two-thirds of each group (99).

Other researchers have either developed minimal standards for the various aspects of care (468), relied on criteria developed by an internal physician committee (86,174), used national clinical standards (417), or used items and scoring systems developed in previous process evaluations (567,576). The resulting criteria consequently may reflect guidelines produced or influenced by national or other formal medical organizations, academic physicians, specialists or generalists, or local practitioners. In the most extensive study of the subject, Hulka and colleagues compared results obtained through different criteria-selection mechanisms (309). Even though the lists of cri-

teria and physicians' adherence to them varied, all criteria sets tended to Produce parallel results.

Even if relative physician performance using various criteria sets may be similar, criteria lists must be limited not simply to critical items but to critical items likely to be recorded. Patients' medical records emphasize key positive findings, especially objective ones, such as test results. Counseling, communication issues, and other important interpersonal aspects of care are relatively inaccessible to record-based evaluations (475). Further, as Donabedian has noted, critics have argued that the medical record rather than the care itself is being assessed (185). In a review of studies of the validity of the medical record, Hulka and colleagues reported arguments that the legal record of care should be good enough for peer review; they also reported findings that one-third of internists kept records inadequate for review and a study noting poor concordance between written and tape records for information more detailed than a patient's chief complaint and diagnosis (309). In an analysis of the relationship between physician entries and independent records of care, however, Lyons and Payne found that all physician records were complete enough for abstracting and that correlations in adherence scores between the two sets of records were generally significant (400).

Thus, in setting criteria, some tension exists between using a fairly detailed list of evaluative criteria (achieving completeness but emphasizing technical aspects of care and including items more likely to be nonessential, redundant, or unrecorded) and using a shorter, less specific list of criteria (useful for determining if some minimal standards of care have been met) (185). Furthermore, the use of explicit criteria may undesirably reduce physicians' flexibility in approaching the care of a wide range of patients in a wide range of clinical situations or undesirably reduce physicians' incorporation of new clinical knowledge into their practices (185).

These problems and tensions in setting evaluative criteria are well illustrated in a series of studies designed to show a correlation between physician performance in hypertension case-finding and management, as measured by various criteria sets, and patient outcomes. The goal

of these studies has been to demonstrate that adherence to criteria lists derived by the methods described above, and presumably reflecting established medical practices, has been associated with favorable patient outcomes. The studies have typically used explicit process measures with the control of diastolic blood pressure as the outcome measure. Several studies have found little or no correlation between process and outcome, even with correction for initial diastolic blood pressure (as an indicator of disease severity) (176,309,339,465). On the basis of similar results, Romm and Hulka concluded that the setting and promoting of standards for the process of care do not guarantee adequate patient outcomes and that peer review groups should recognize the limitations of both process and outcome measures (530).

Some process-outcome correlations have suffered from poor research designs or statistical analyses, for example, failure to control for patients' initial status in their correlations (411). More fundamentally, many of the evaluative criteria have questionable validity because they are often related to matters such as identifying nonessential causes of hypertension or serious late-stage complications and therefore would not be expected to have a significant impact on overall outcomes. One observer suggests a two-stage approach for the acceptance of specific medical practices as assessment standards: 1) constructing criteria sets based on clinical research concerning diagnostic accuracy and therapeutic effectiveness, and 2) applying the criteria to evaluate physician performance (411).

The key point is that processes believed to have a significant impact on the outcome of care should form the basis for valid assessment criteria (68). Other items, however embedded in customary or established medical practices, should not (476).

For hypertension, examples of processes believed to have a significant impact on outcomes include adherence to a regimen of antihypertensive medications and behavioral and dietary modifications. Patients' knowledge of their disease and adherence to a physician's recommendations for its management appear to depend on the ability of providers to communicate the rationale and benefits of therapy (558). Unfortunately for assessments using medical records, these

items all involve key interpersonal components, including patient education and motivation as well as physician discretion.

Methods for measuring these interpersonal aspects of care lack sophistication, but some studies have used rather innovative approaches to address these measurement difficulties and generally have found process-outcome correlations. One study, for example, included a measure of patient compliance with therapy based on the patient's verbal reports about taking prescribed medications, following dietary guidelines, observing recommended changes in activities and habits, and keeping medical appointments (250). Compliance with therapy accounted for a greater portion of the variance in clinical outcomes than the type of therapy, and compliance was also strongly associated with both patient knowledge and perceptions of care. Assessing compliance with the medication regimen by counting the number of pills remaining in patients' prescription bottles, another study found significantly higher rates of blood pressure control among more compliant patients and among patients receiving a more vigorous medication regimen (286). Although patient compliance clearly depends on many factors—psychological, economic, demographic, and other—some of which lie beyond the influence of the physician, these studies indicate that a patient's compliance with therapy has a significant impact on the outcome of care and may be related to the physician's talents in educating, motivating, encouraging continuity of care, and other interpersonal matters.

Just as process and outcome measures may yield divergent results when used to judge the same cases, implicit and explicit process measures may yield results with some divergence (99,309). Implicit ratings for a case tend to be higher than ratings for the same case based on adherence to explicit criteria. Judges using implicit criteria were influenced by favorable outcomes, and they justified their conclusions with items different from the items on the explicit criteria lists; specifically, these judges mentioned procedures related to followup care, criticisms of the physician for performing too many procedures or failing to respond adequately to additional risk factors or comorbid conditions, patient characteristics, and

other processes difficult to specify on explicit criteria lists.

Research efforts have led to significant progress in identifying ways to increase the validity of process measures, but a number of difficult issues have not yet been fully resolved. In the validity of process measures, as in the reliability of process measures, trade-offs exist along the spectrum from implicit measures to explicit measures. Because implicit measures allow a patient's medical record to be reviewed in its entirety, they do not break down in the evaluation of cases that are not well suited to a specific set of explicit criteria. Much of the research on process assessment has focused on enhancing the validity of explicit process measures by refining methods for developing and using explicit criteria. Another approach to enhancing the validity of explicit process measures is to combine them with implicit peer review methods. An example is the use of a physician practice audit system that includes a review of each medical record using explicit criteria, which can be performed by nursing personnel, plus a more subjective review performed by a physician (417).⁸

Other validity-related difficulties in assessments of the process of care are common to both implicit and explicit process measures. Both types of measures are limited by the quality of medical records, and neither is well suited at present to evaluating interpersonal and other aspects of care not likely to be found in a patient's medical charts. It is important to note, however, that evaluations of the process of care are the only means of acquiring relatively direct information on whether a physician is following the best clinical practices; outcome assessments cannot be used for this purpose. This fact alone is a very important validity consideration.

In addition to all of the factors related to the validity of implicit reviews and the use of explicit criteria in the evaluation of care for hypertension, another major issue is the extent to which an evaluation of the process of care for hypertension reflects the quality of care a physician is likely

⁸Another example is "targeting" implicit review to cases judged unacceptable by simple explicit criteria (475), as noted above.



Photo credit: Harvard Community Health Plan

Combining explicit criteria, such as the monitoring of patients' blood pressure, with experts' implicit judgments improves the validity of using process measures to evaluate physicians' care for hypertension.

to provide for other conditions. Clearly, evaluating care for a single diagnosis appears insufficient to assess a provider's medical abilities generally. Kessner has suggested that the careful selection of a limited set of conditions for evaluation, called "tracers," could provide a framework for evaluating the routine diagnostic, therapeutic, and followup care provided by a health system to the different population groups that it serves (351). Although Kessner was optimistic about the workability of the tracer framework, most subsequent studies purporting to use "tracers" have simply applied the term as a label to the one or several conditions for which care was being evaluated.

There has been little real progress in developing a systematic method to evaluating quality of a physician's care comprehensively with only a limited number of indicator conditions. One study has confirmed the limitations of the generalizability of current explicit performance measures in evaluating internists' management of five hospital diagnoses and six office diagnoses, including hypertension (552). That study found that substandard performance by an internist in managing at least one office condition was associated with a significantly higher proportion of substand-

ard treatment of other office conditions. Substandard office performance by an internist, however, was unrelated to the internist's performance in the hospital, and substandard performance for a hospital condition or superior performance in any condition had no predictive value for substandard or superior performance in other areas. The investigators concluded that the lack of clustering of high or low performance across diagnoses implied that each major diagnostic category in an internist's practice must be assessed independently.

Since a physician's performance in treating one condition does not appear generalizable to the physician's treatment of other diagnoses, an alternative approach is to evaluate a physician's performance across all or most conditions the physician must treat. Borgiel, et al., have developed detailed unweighed explicit criteria sets for 180 conditions most commonly treated by Canadian family physicians (84,85,86). Expanding evaluation to a wide range of diagnoses eliminates the problem of generalizability. But validity issues relating to whether the quality of care for hypertension (or any other condition) can be assessed effectively through outcome or process measures remain.

Feasibility of Using the Indicator

Regarding the feasibility of using evaluations of the process of care to assess quality, the main issue centers on how expert review is incorporated into the evaluation—in developing an evaluative framework (explicit), in the individual reviews (implicit), or in some combination of these stages. Many of the same feasibility obstacles for outcome assessments (plus distinct validity problems) posed by using medical records apply as well.

Implicit judgments by medical experts regarding the process of care would have several important advantages in a widespread program of quality assessment to provide information to consumers. Such judgments might be more acceptable to providers as a fair means of assessing the many complex details of individual cases than assessments in which medical professionals do not participate directly (410), a desirable goal since professional support appears to promote the suc-

cess of an evaluation program (475). Similarly, judgments by clinicians might help promote public confidence in the assessment of a physician's care for a specific condition, since consumers appear to rely heavily on expert opinion in their decisions regarding medical treatment. Further, implicit reviews of the process of care obviate the need for developing and revising criteria lists.

A major disadvantage of implicit assessments of the process of care is their relatively intensive use of expert professional resources. Participation in evaluations would have to become a routine part of the physicians' duties (475), a situation that would reduce their activities in other clinical areas. If such formal responsibilities are not incorporated, record review will probably involve significant delays and inconsistencies (99). These ongoing commitments can be expensive financially as well. The guided implicit review method described by Hulka, et al., required about 15 minutes per case (309) and could be costly (410). Moreover, given the reliability concerns already noted, at least two or three physicians must review each case (and even then high accuracy rates are not guaranteed). Thus, to evaluate a substantial portion of the medical community on a regular basis, an implicit review program would require a major investment of funds and professional time. Additional costs would be incurred for such activities as administration, case abstracting, and training.

In contrast, explicit methods of assessing the process of care have much lower requirements for physician time, since expert participation is limited to developing and revising criteria and reviewing the reliability of the data collection. If training programs are provided, actual record review can be performed by nurses, medical students, and others familiar with the medical environment. The significantly lower expense and higher reliability of explicit reviews may account for their much more frequent use in studies of the quality of ambulatory care. Further, once the criteria and scoring method have been determined, the quantitative data resulting from the analysis can be summarized in a straightforward format to consumers.

These advantages of explicit assessments must be weighed against the validity limitations of such

assessments; as noted previously, adherence to criteria lists may not be a fully valid representation of the quality of care provided in specific cases. One likely effect of a policy decision to use explicit criteria to assess the process of medical care would be increased attention to the details of process being measured, possibly at the expense of other aspects of care that might be much more relevant to the clinical outcomes and well-being of a particular patient. Such distortion could be minimized by using only a short list of relatively simple criteria clearly tied to patient outcome, but an assessment based on such a list would probably be capable of determining only whether minimal care was provided.

Borgiel and his colleagues have used explicit criteria to assess the performance of family practitioners (84,85,86). Trained reviewers apply explicit criteria for 180 conditions to review 40 medical records chosen at random. A software program for a portable computer facilitates the abstraction procedure (418). The assessment also includes an interview of the participating physician and a survey of 60 current patients. Each assessment costs about \$500 for the patient record audit and \$500 for the patient survey (Canadian dollars), costs borne by the physician being assessed. Borgiel and his colleagues recently completed an assessment of 120 family practitioners in southern Ontario. Although participation in

```

Chart Format ( ID: 1 Sex: Male Age: 42 Years )
>13. If general assessment done were the following recorded
on that visit?
a) weight ( < )
b) blood pressure ( < )
c) Pap smear if patient on birth control pill or has IUD
    has been done by another physician regularly
   ( < )
d) urinalysis ( < )
e) hemoglobin/hematocrit (doctor or lab) ( < )
f) rectal if male > 45 years ( < )
g) cardiovascular system (3 of 9) ( < )
h) chest exam (1 of 5) ( < )
i) abdominal exam (1 of 6) ( < )

Diagnosis      Date      D  CF      Medication      Date Rx      R  CF
1 Bronchitis  11/11/84  1  100%      1  11/11/84  1  100%
2 1st visit    11/11/84  1  100%      2  11/11/84  1  100%
3 1st visit    11/11/84  1  100%      3  11/11/84  1  100%
4 1st visit    11/11/84  1  100%      4  11/11/84  1  100%

Press END key to exit from Chart window

```

Photo credit, College of Family Physicians of Canada

To assess a physician's performance using explicit criteria, trained reviewers use a specialized software program to abstract information from patients' medical records.

the assessment was voluntary, a response rate of over 80 percent was achieved through the use of a recruiting network of clinicians. At present, results are used primarily for educating the assessed physician and for certification decisions by the Canadian College of Family Practice rather than for public information.

Although Borgiel's review of Canadian family practitioners relies exclusively on an explicit method (84,86), other approaches attempt to combine implicit and explicit features with a goal of achieving some of the benefits of each. The targeted method used by Palmer focuses implicit review on cases likely to be unsatisfactory. This method promotes the validity of conclusions about poor quality while reducing expert time—

provided, of course, that a high proportion of cases meet the minimal explicit criteria (475). Another example is the explicit/implicit practice audit used in a Minneapolis/St. Paul health maintenance organization (417). Following a phase of feedback and revision to improve the use of the assessment program, this audit system has become regarded as acceptable to most clinicians and is strongly endorsed by the health maintenance organization's management for providing measures of process useful in improving quality of care. The practice audit is expensive, however: the audit requires three nurses and a physician to spend 6 hours on site at the clinic, plus additional time writing the report.

CONCLUSIONS AND POLICY IMPLICATIONS

The most reasonable method for assessing physicians' performance in providing care for a particular condition is to integrate measures of the outcome of care with implicit and explicit measures of the process of care. Depending on the specific method used, a combination of approaches would capture some of the advantages and minimize some of the disadvantages of each generic approach to some extent (see table 7-4). The use of a combined method would be most likely to achieve the goal of promoting reliable and valid judgments as efficiently as possible. Use of combined methods is becoming more common for internal purposes by utilization and quality control peer review organizations (PROS) and by large health care organizations (226), a trend indicating their feasibility.

An effective combined approach could have a range of features, depending on which features of each generic approach to quality assessment are adopted. Cases identified as problematic by the application of specific process or outcome criteria, for example, could be reviewed by physicians, thus providing a check on the validity of the judgment suggested by explicit criteria in a given case (475). Alternatively, physicians using implicit process criteria could review a fraction of the cases randomly selected from a given pro-

vider; in the process, reviewers could check whether the results of the explicit evaluation are valid and possibly detect cases of inadequate care that met explicit standards (417). At least at present, some component of peer evaluation appears necessary for supporting the validity of judgments about the adequacy of complex, evolving clinical practices and varied patient characteristics.

Assessment methods that combine the use of explicit criteria and implicit review by medical experts tend to be more expensive than assessment methods based on explicit criteria alone, but combined approaches that target the use of medical experts should cost substantially less than comprehensive peer review systems. The implicit review component of a combined method should be directed primarily toward addressing the weaknesses of the other components of the assessment, such as adjusting for relevant clinical features of the particular case. As assessment methods become more sophisticated, the role of physician review could be refined accordingly, to promote the efficient use of resources in the assessment process.

Although a combined approach to evaluating care for a specific condition appears most promising, many significant obstacles remain for the

Table 7-4.—Comparison of Generic Approaches to the Evaluation of Physicians' Performance: Care for Hypertension^a

Generic approach	Reliability		Validity		Feasibility	
	+	-	+	-	+	-
Outcome assessment	Blood pressure measurements are accurate	Repeated measurements over time are required; must depend on recording in patient records	Face validity apparent to consumers	"Inadequate sophistication of case-mix adjustment methods Provides no direct information on whether provider is using accepted medical practices	Likely to be recorded in patient records; easy to abstract	● Long followup period likely to be required
Implicit process assessment		Higher intra- and inter-rater variations because of method's dependence on internal standards (several or more reviewers required)	Face validity for consumers and providers ● Theoretically allows comprehensive consideration of all relevant elements in record	Standards may vary or be applied inappropriately	Practice can be audited in a day	● High costs Relatively intensive use of medical professionals
Explicit process assessment	● Specified criteria make measurements easier to replicate		Criteria for judgment are explicit and based on expert standards	Criteria may not fully reflect relevant elements in individual cases Tension between minimal and detailed criteria sets	● Lower cost Less intensive use of physicians ● Practice can be audited in a day	

asterisks are used to designate particularly strong (or weak) features.

SOURCE: Office of Technology Assessment, 1988.

development of any system to assess and disseminate information to consumers about the quality of care provided by individual providers for particular conditions. Some of these problems appear to be organizational and administrative in nature. Additionally, the problems in reliability, validity, and feasibility described in this chapter suggest that important research and implementation tasks remaining before such assessment systems could be realized effectively. Though the chapter has focused on hypertension as a case study, these issues are also relevant to providing information to consumers about the quality of care for other conditions.

Techniques exist to provide such assessments of physician performance. The work of Borgiel and colleagues, McCoy and colleagues, Palmer, and many others indicates that practice assessment systems can be implemented on a continuing basis (84,417,475). Present programs to assess the quality of physicians' care for specific conditions are not designed to provide public information. Instead, they appear to have other worthwhile purposes. Internal quality assessment systems within health care organizations provide feedback and education to providers, to promote quality assurance within a delivery system (58). PROS have been charged with evaluating the quality of ambulatory care of federally funded medical providers; the efforts of PROS are likely to emphasize screening out poor physicians rather than providing consumers information (see ch. 6).

In this institutional context, key administrative and policy issues would have to be resolved before a program could be implemented to provide systematic information about the performance of individual physicians. Since health care providers and organizations do not currently provide such information reliably and in formats useful to consumers, some type of incentive mechanism—either public (e.g., new regulations or enabling statutes) or private (e.g., directives from third-party payers)—would be essential for making the relevant data about patient care available for review. Incentives could be more or less compulsory, ranging from recommendations and voluntary guidelines to requirements that physicians undergo a practice audit as a prerequisite for payment of services, certification, recertification, or

licensure. Legal liabilities surrounding peer review and quality assessment would also require analysis (111). As the extant programs indicate, costs for any general audit system would be considerable; they could be borne by the Federal Government or spread among State governments, insurers, other payers, and providers. The issues just cited are only some of the relatively unexamined topics relevant to the successful implementation of a general system of providing physician assessments for consumers.

Moreover, the effects of requirements to disclose information on the dynamics of systems designed for internal quality assurance in health care organizations should be considered. The primary purpose of those systems is to provide effective feedback to improve the quality of work of physicians in the health care organization. But awareness that findings will be made public in a competitive environment could create incentives to minimize the discovery of substandard practices.

In addition to the organizational obstacles cited, many important technical obstacles remain in making this information optimally reliable, valid, and feasible to obtain. Many of these obstacles could be addressed through support from one or more of the research offices in the U.S. Department of Health and Human Services, from the private sector, or from cooperative efforts. To the extent that these difficulties remain unresolved, any assessment method adopted should include features to compensate for the assessment's deficiencies; in this regard, the flexibility provided by combined methods for evaluating the quality of care is especially advantageous.

For evaluations of care using patient outcomes, additional refinements of case mix and severity-of-illness adjustments are needed to make the measures more responsive to the quality of the physician. Additional investigations of methods to increase retention of patients for followup over time and decrease costs of the longer term followup required for adequate outcome assessment of care for a chronic condition might also be useful. For evaluations of care using process measures, investigations of ways to improve the reliability and efficiency of peer review, and in particular applied research on how many physicians and how much of their time is required for

a reasonably accurate practice assessment, would permit better use of implicit review methods. Perhaps more importantly, it would be useful to support more sophisticated studies on integrating expert judgments effectively into techniques that also rely in part on adherence to criteria, observed outcomes, or other less costly methods. As a number of investigators have demonstrated, combining features of the different approaches to assessing care can be a very effective way to minimize the weaknesses of individual approaches. A key goal of such studies should be to develop optimal methods in the assessment process for involving physicians, a limited and costly resource.

Another important area for further investigation is determining what relationships exist between the quality of a physician's care for one condition and the quality of the physician's care for other conditions—the issue of generalizability. The few studies that exist provide a sense that each condition is different, but whether assessments can focus on a limited number of diagnoses or must measure the quality of care across the entire spectrum of a physician's practice is obviously a crucial logistical question. Although measures do not appear generalizable at present, more sophisticated analyses might detect underlying patterns or correlations in physician treatment behaviors.

Another key area for further work is the development of better techniques for extracting relevant information from medical records. This is essentially an issue of data quality. Evaluations of care using patients' medical records can assess only items that should be present reliably in the charts, and ambulatory records have much more uneven quality than hospital information systems (479). Increasing computerization of patient data bases is a positive development in this regard. Some larger health care organizations and group practices are relying on such systems, and some quality assessments within hospitals involve manipulation of computerized patient data (446,547). The claims that physicians and hospitals submit to third-party payers could also provide computerized information, especially if entries concerning patients' diagnoses and clinical status were improved. Although a major segment of ambulatory practitioners has not yet adopted computerized office data systems, the creation of some kind

of national standards for computerized patient records could be an effective approach to improving reliable access to relevant information on the care process. More generally, uniform standards for data collection and reporting could be developed for all ambulatory records. Such measures would have to consider balancing increased time and cost of more detailed records with the benefits to quality assessment and other activities possible through more reliable or complete data.

Even with such improvements, many critical aspects of medical practice will remain difficult or impossible to capture in a provider's written record. Thus, increasing sophistication in measuring interpersonal aspects of care and physician influence on patient compliance with a therapeutic regimen could result in substantial improvements in the validity of process measures. These deficiencies can be addressed at least in part through patients' assessments of care (see ch. 11), and a physician assessment system featuring medical record reviews complemented by patient surveys could be a powerful approach to developing information on both the technical and interpersonal aspects of care provided. Borgiel and his colleagues currently use this combination in their practice assessments (84,86). Other creative approaches to measuring interpersonal aspects of care, as well as the other physician services not well reflected in the medical record, might also be useful.

Much research has already been devoted to setting standards for evaluating physicians' performance, but the development, evaluation, use, and timely revision of criteria and standards remain a central issue in any assessment that involves explicit criteria. In part, the development of criteria and standards requires clinical studies: much uncertainty remains about what clinical practices and procedures are most strongly associated with medical effectiveness. Ideally, only effective processes should form the basis for criteria developed for evaluations of care (411). In the care of hypertension and some other conditions, the processes that are effective have been relatively well established, and many useful criteria sets have been developed over the last 15 years. Some type of national clearinghouse, perhaps administered

through professional medical organizations, might both promote the effective use of these criteria sets and coordinate their refinement with guidelines for the content of medical records.

Improving the quality of ambulatory care assessments will also require further attention to more practical matters related to feasibility. Some of these concerns—such as promoting efficient use of medical experts—have already been mentioned. Many other approaches could also lead to lower assessment costs; examples include improved training methods, improved coordination with other quality-related projects and with organizations and activities designed to promote medical quality, and innovative approaches such as self-audits (417). Another key area is the adaptation of computer technologies to assist in the collection of assessment information. For example, office audits can be expedited using software programs to enter data on adherence to criteria (419). Conceivably, these methods could be coordinated with computerized data base record systems to make assessments more fully automated.

Two other crucial considerations related to the feasibility of using evaluations of physicians' management of specific conditions to evaluate quality deserve final mention. One is the need for further deliberation on whether attention to all of the research items detailed above is worthwhile, or whether less ideal or entirely different approaches would be better alternatives for provid-

ing consumer information or for increasing the likelihood that patients will receive high-quality care for specific conditions, such as hypertension. Although considerable experience with assessment methods in both research and practical settings indicates that these methods—especially combined approaches—have considerable promise, the discussion in this chapter suggests that serious technical, organizational, and economic obstacles remain before a functional system could be implemented nationally to provide useful information to consumers about individual physician performance for certain conditions. In this regard, it is important to recognize that almost no research has been directed specifically toward the question of providing information to consumers about the quality of the processes of care they receive for the treatment of hypertension or any other condition.

The other crucial consideration, running through out this chapter, is that evaluations of the process of care clearly require the leadership and assistance of the medical profession. Historically, professional medical associations have played the paramount role in evaluating physicians' performance; at present, they are continuing to expand their activities in promoting high-quality care. Independently of its own assessment activities, or in coordination with them, the Federal Government can support the medical profession's efforts.