
Pew Memorial Trust Policy Synthesis: 4

Identifying and Managing Inappropriate Hospital Utilization: A Policy Synthesis

Susan M. C. Payne

Utilization review, the assessment of the appropriateness and efficiency of hospital care through review of the medical record, and utilization management, deliberate action by payers or hospital administrators to influence providers of hospital services to increase the efficiency and effectiveness with which services are provided, are valuable but relatively unfamiliar strategies for containing hospital costs. The purpose of this synthesis is to increase awareness of the scope of and potential for these approaches among health services managers and administrators, third-party payers, policy analysts, and health services researchers. The synthesis will assist the reader to trace the conceptual context and the historical development of utilization review from unstructured methods using individual physicians' professional judgment to structured methods using explicit criteria; to establish the context of utilization review and clarify its uses; to understand the concepts and tools used in assessing the efficiency of hospital use; and to select, design, and evaluate utilization review and utilization management programs. The extent of inappropriate (medical unnecessary) hospital utilization and the factors associated with it are described. Implications for managers, providers, and third-party payers in targeting utilization review and in designing and evaluating utilization management programs are discussed.

Prepared under a grant from the Pew Memorial Trust for the Association for Health Services Research.

Susan M. C. Payne, Ph.D. is Research Assistant Professor in the Section of General Internal Medicine, Health Care Research Unit, Boston University School of Medicine. Address correspondence and requests for reprints to Susan M. C. Payne, Ph.D., Boston University School of Medicine, 720 Harrison Avenue, Suite 1102, Boston, MA 02118.

I. INTRODUCTION AND BACKGROUND

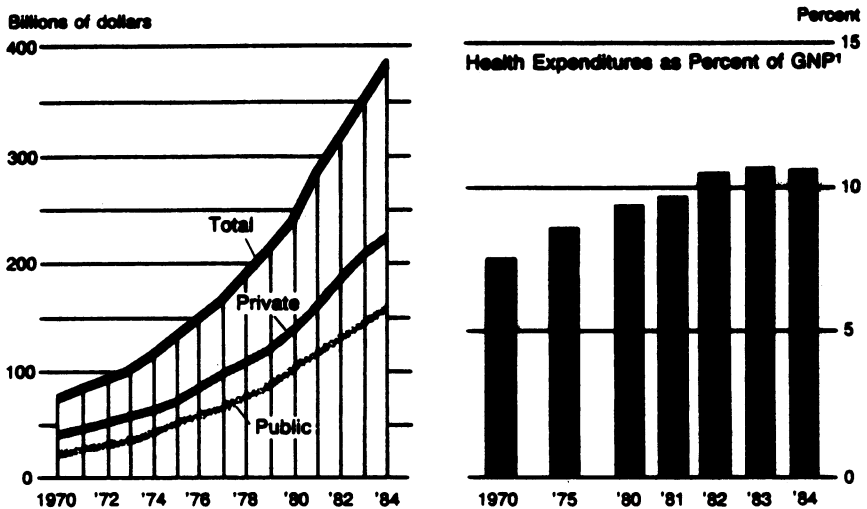
A. WHY IS INAPPROPRIATE HOSPITAL UTILIZATION IMPORTANT?

The Problem: Soaring Health Care Costs

From 1970 to 1985, expenditures on health and medical services in the United States rose from 7.6 percent of the gross national product to 10.7 percent, a rate of increase consistently greater than that of other goods and services. In 1984, \$157.9 billion, or 41 percent of the health care dollar, was spent on hospital services (U.S. Department of Commerce, 1986). Prices of hospital services have risen faster than those of any other health services (Figures 1 and 2).

Because of the sheer magnitude of hospital costs and their rapid increase, third-party payers and employers have focused their cost-containment efforts on containing or reducing hospital expenditures, pressuring hospitals to improve efficiency, reduce utilization, and compete against each other in price.

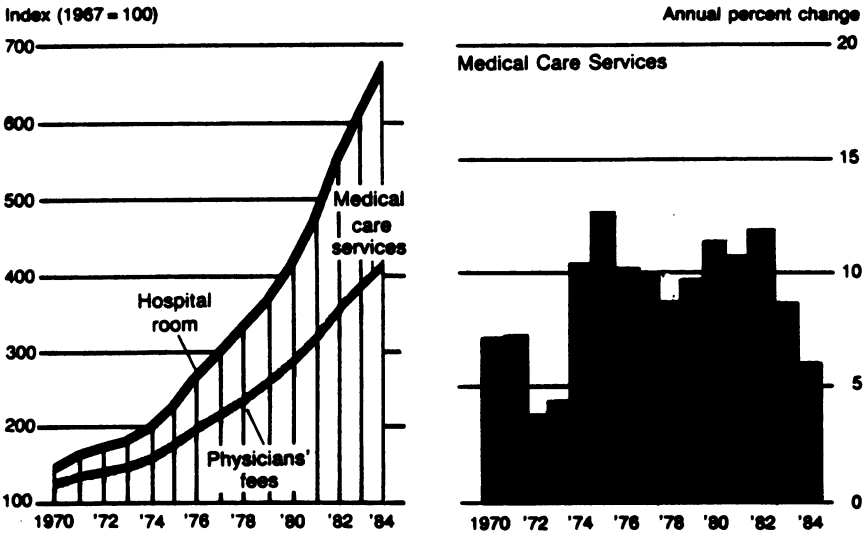
Figure 1: National Health Expenditures, 1970-1984



¹ Gross national product.

Source: U.S. Department of Commerce, Bureau of the Census, 1986.

Figure 2: Indexes of Medical Care Prices, 1970-1984



Source: U.S. Department of Commerce, Bureau of the Census, 1986.

Cost-Containment Strategies

Two approaches to cost-containment — prepaid, capitated systems such as HMOs, and fixed price reimbursement plans such as the Medicare DRG-based Prospective Payment System (PPS)— are by now familiar to many people. Few in the general public, however, know about another promising approach that is increasingly being adopted by large corporate and public payers — utilization management based on utilization review.

Utilization Review and Utilization Management

Utilization review is review of the patient's medical record through application of defined criteria and/or expert opinion. The purpose of utilization review is to assess the efficiency of the health care process and the appropriateness of decision making related to the site of care, its frequency, and its duration (adapted from Donabedian, 1982; and Demlo, 1983).

Utilization review can be conducted either before admission (pre-admission review), during the patient's stay (concurrent review), or

after discharge (retrospective review). Four dimensions of utilization can be reviewed:

- *What* care was provided? Did it meet the patient's medical needs?
- *When* was the care provided? Did the admission occur an appropriate length of time before the surgery, or were there unneeded preoperative days of care?
- *How much* care was provided, in terms of the duration and frequency of care? Was the hospital stay the appropriate length? Was the number of laboratory tests appropriate?
- *Where* was the care provided? Did the patient require hospital-level care, or could he or she have been treated in an alternative site such as an outpatient clinic, an ambulatory surgery center, or a nursing home?

Utilization review is closely related to quality assurance. Both techniques rely on review of the individual medical record (medical audit) and both evaluate the relationship between the patient's need for medical services and the services received. Utilization review is generally limited to the review of the medical care process, while quality assurance is broader, including the review of the structure, process, and outcome of care, the sequelae of treatment and, at times, the interpersonal relationship between provider and patient (Donabedian, 1973). Utilization review focuses on the *efficiency* with which care is provided (the cost per unit of service delivered, such as the cost of a day in the hospital or of a given laboratory test). Quality assurance, on the other hand, focuses on the *effectiveness* with which hospital services are provided (the benefit or impact of a given unit of service, such as the amount of improvement in ability to function which occurs after a day in the hospital), often with less emphasis on the costs of the services provided than is true with utilization review (Donabedian, 1985(2)).

Utilization review is an important component of *utilization management*, which is deliberate action by payers to influence hospitals or physicians or by hospital administrators to influence physicians, to increase the efficiency or the effectiveness with which hospital services are provided.¹

In the past, the payer had no direct relationship with the provider, serving simply to establish benefits provided to employees or services covered for beneficiaries. Decisions on delivery of care were made by the employee/beneficiary and the physician, with no involvement of the payer. Hospitals generally had quality assurance programs but

were not actively involved in cost-containment or utilization management.

Utilization management represents a radical change in that traditional relationship between third-party payers and providers, since in utilization management the payer (or the hospital) is an active participant, influencing the scheduling, location, and process of providing care. This proactive relationship offers opportunities for cost savings while it places new technical, logistical, and legal demands on the payer or the hospital for ongoing decision making and program implementation.

In contrast to alternative delivery systems such as HMOs, utilization management can be undertaken without establishing new organizational or contractual arrangements with providers. In contrast to payment system approaches such as the Medicare PPS (which require a large share of the market to be effective), utilization management can be undertaken by individual corporations, insurance companies, or hospitals. Blue Cross plans, commercial health insurers, state Medicaid programs, corporations purchasing hospital services for employees, and hospitals forced to improve their efficiency in order to remain competitive have all initiated utilization review and utilization management programs.

Utilization management can have several benefits:

- Reducing costs through reducing unnecessary hospital utilization. Recent studies of the appropriateness of hospital utilization (discussed in Section III) have found that 6–40 percent of admissions and 12–39 percent of days of care may be medically unneeded, which indicates the potential for reducing hospital costs by reducing inappropriate utilization.
- Improving the quality of care by reducing the chance of nosocomial infections (acquired in the hospital) or iatrogenic illness (illness caused by treatment, such as side effects of medication) (Mills, 1978; Steel et al., 1981; Steel, 1984).
- Maintaining quality of care by assuring that the hospital services provided are of a sufficient duration, frequency, and level of care to promote optimal health outcome.
- Preserving access to care by reserving hospital resources for those who need inpatient care.
- Defining and articulating standards of care (which can help resolve conflicts about acceptable or proper practice in malpractice cases).

The elements of utilization management programs are discussed in Section D.

B. UTILITY OF THE SYNTHESIS

Objectives

This synthesis will assist the reader (1) to trace the conceptual context and the historical development of utilization review, (2) to clarify the uses of utilization review, (3) to understand the concepts and tools used in assessing the efficiency of hospital use, and (4) to select, design, and evaluate utilization review and utilization management programs.

Scope of the Synthesis

The synthesis addresses the following questions:

1. What are the conceptual foundations of utilization review? How does utilization review fit into utilization management?
2. What are the characteristics of utilization review programs?
3. What are the characteristics of methods and instruments developed to identify appropriate and inappropriate care?
4. How much utilization is inappropriate? Where is inappropriate utilization likely to occur?
5. What are the implications of the findings for policymakers, third-party payers, and health delivery managers? How can the findings be used to target, design, and evaluate utilization management programs?
6. What additional information is needed by policymakers, third-party payers, and providers?

The synthesis is limited to inpatient hospitalization since methods for identifying inappropriate utilization are more fully developed for inpatient services than for ambulatory, long-term, or other types of care. The emphasis is on methods related to overutilization because they have been more extensively developed than methods related to underutilization. For the same reason, methods to assess the appropriateness of the timing of the services provided and the level of care (e.g., inpatient versus outpatient) are discussed in more detail than are methods to assess the medical need for services.

Because this approach to cost-containment is new and the literature on evaluating it is still rudimentary, the synthesis defines the

characteristics of utilization review methods and utilization management programs but does not formally evaluate their impact.

Intended Audience

The target audience is hospital managers functioning in a prospective payment environment, e.g., in hospitals treating Medicare patients or members of HMOs, or in Preferred Provider Organizations; managers in managed care systems such as HMOs or PPOs; utilization review directors; and decision makers for large third-party payers such as corporations, Medicaid, Medicare, Blue Cross plans, and commercial insurers.

C. CONCEPTUAL CONTEXT OF UTILIZATION REVIEW

The Relationship between the Patient's Need for Services and the Services Provided

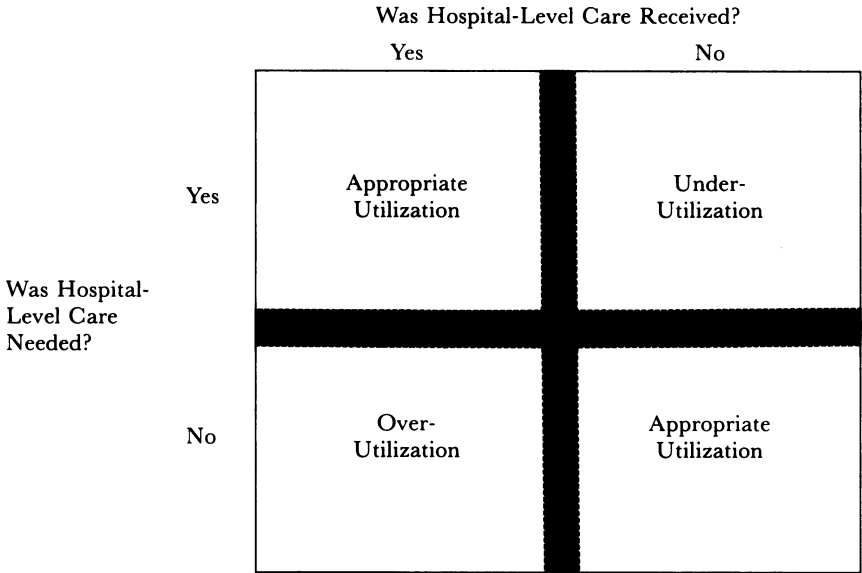
The goal of utilization review and utilization management programs is to identify and reduce unnecessary or inappropriate hospital use while maintaining access to needed utilization. Inappropriate hospital utilization can be defined as utilization which is not suitable to the patient's medical need.

There are two types of inappropriate utilization. *Overutilization* is care which is of no benefit to the patient (such as extra days at the end of a hospital stay, after the patient has recovered enough to go home) or care which could be provided in a lower-level, less costly setting (such as in a nursing home instead of a hospital or on an outpatient basis rather than as an inpatient). *Underutilization* is care which is not sufficient in type, length, location, or intensity to meet the patient's medical need. For example, underutilization would occur when a patient who still needs inpatient-level care is discharged or when a patient whose health status warrants treatment in an inpatient setting has a surgical procedure in an outpatient setting).

The relationship between the patient's need and the services provided is highlighted in Figure 3.

Patients who need services but do not receive them are *underserved*; patients who do not need services but receive them are *overserved*. Patients in the upper left quadrant in Figure 3 are *appropriately hospitalized* and patients in the lower right quadrant are *appropriately not* in the hospital.

Figure 3: Relationship between Need for Hospital Services and Services Received



Source: Restuccia and Holloway, in Restuccia and Tracey, 1987.

The boundaries in Figure 3 between the quadrants are hatched to reflect the uncertainties inherent in medical practice due to differences in patients' signs, symptoms, and responses to treatment (Eddy, 1984), and the existence of differences among practitioners or across geographic regions about acceptable medical practice (Wennberg and Gitelsohn, 1982). The implications of these uncertainties for utilization review are discussed below.

Relationship between Inappropriate Utilization and the Costs and Outcomes of Care

Both under- and overutilization have implications for the *cost* and *quality* of care.

Overutilization has a clear impact on health care costs: unnecessary use of the hospital instead of less expensive alternatives, such as outpatient clinics or ambulatory surgery centers, increases costs, as do unnecessarily long hospital stays. Overutilization has important quality-of-care implications, too, since it increases the chance of hospital-acquired infections, treatment-induced side effects, or other

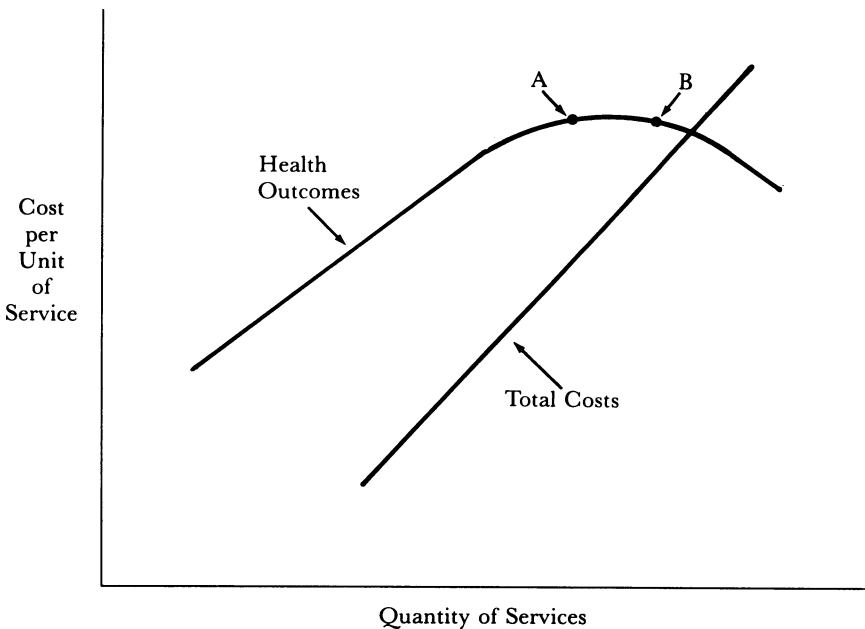
unintended consequences of hospital care such as excessive blood tests resulting in blood loss (Mills, 1978; Steel et al., 1981; Steel, 1984).

Underutilization is clearly related to quality of care: a patient who does not receive care at a high enough level (a patient, for example, who needs hospital-level care but who is treated on an outpatient basis) or one who receives care that is too infrequent or too short in duration to meet his or her medical needs is more likely to suffer relapse, readmission, decrease in functional status, or even death than is a patient receiving treatment sufficient to meet his or her medical need. Although underutilization may save money in the short run, it has long-run cost implications, since the consequences of undertreatment can have economic, personal, and social costs.

The relationship between the *quantity* of services provided (such as the length of a hospital stay or the number of laboratory tests provided), the *costs* of the services received, and the *health outcomes* received from the services can be depicted graphically (Figure 4).²

As the quantity of services increases, the total cost of the care rises.

Figure 4: Relationship between Hospital Costs, Quantity of Services Received, and Health Benefits



Up to a given point, A, the benefits to the patient in terms of health outcome increase. Between A and B, however, an increase in quantity produces no appreciable improvement in the outcome (as, for example, when a patient receives unnecessary but noninvasive laboratory tests). Beyond point B increases in quantity can produce a *worse* outcome. The area to the left of point A is the area of potential underutilization, and the area to the right of point A is the area of potential overutilization.

In economic terms, the goal of utilization management programs is to move along the health outcome curve from point B toward point A to reduce quantity (in order to reduce costs) without dropping below point A. In clinical terms, the goal is to reduce medically unnecessary use while maintaining access to appropriate, high-quality hospital services.³

Relationship Between Utilization Review and Changing Incentives in the Health Care System

Until recently, most of the incentives in the health care system serving the nonindigent population have encouraged the overutilization of hospital services:

- Economic incentives in the reimbursement system including fee-for-service payment, which pays the physician or hospital for each service rendered; cost-based retrospective payment, which reimburses for any costs incurred; and generous third-party insurance coverage
- Clinical incentives (the tendency to err on the side of providing too much rather than too little care)⁴
- Legal incentives (such as the practice of “defensive medicine,” which is the provision of unnecessary or marginally valuable services as a protection in case of malpractice suits).

Various cost-containment programs have recently been initiated to counteract incentives toward overutilization. In addition to utilization management programs, the Medicare Prospective Payment System, for example, provides economic incentives to hospitals to reduce length of stay by paying a fixed price for each type of patient. Systems such as HMOs, PPOs, and IPAs compensate physicians using a prospectively set fixed price per capita or per case instead of paying on a cost-plus, fee-for-service basis to afford them economic and/or organizational incentives to reduce hospital admissions and length of stay. These innovations and the resulting competition they have caused

among hospitals have reversed many of the previous incentives toward overutilization and have increased the risk of underutilization.

While detecting overutilization will continue to be a primary focus of utilization management and other cost-containment programs in the near future, detecting underutilization is expected to become more important as a counterbalance, assuring that cost-containment does not reduce acceptable access to and availability of care. Accordingly, utilization review/methods to detect underuse will need to be developed in the future.

A Caveat

As mentioned above, there is an element of uncertainty in the practice of medicine and there are individual and regional differences in practice patterns. Unlike the production of manufactured goods, the production of hospital services is a highly specialized process, tailored to the health status and the behavior of the patient and dependent on the physician's skills, knowledge, and practice pattern or style. Monitoring the effectiveness and efficiency of the medical care process is therefore a complex undertaking, not amenable to a rigid approach. It requires a recognition of the individuality of the patient, variations in physicians' practice patterns, and the nature of medicine as an art as well as a science. Utilization review systems which ignore that complexity risk being either unrealistic or narrow:

The difficulties in making judgments or decisions which are acceptable to the profession and to the public at large, and which are in some way "valid," are nowhere more apparent than in studies of medical care facility utilization. Clearly, the more or less arbitrary establishment of rigid criteria to suit all cases, if this were possible, would solve the problem of the variability of opinion between observers, but this might not be entirely appropriate to either the tradition of individuality among physicians or the relatively unscientific basis of a good deal of medical practice.

(Zimmer, 1967)

Utilization review must be based on fundamentals of medical practice, not on "cookbook" directives:

The practice of medicine is an art which can no more be standardized than can the art of writing English, yet certain rules of grammar are followed by all good writers, even though each maintains his individuality of expression. Similarly, certain fundamentals of good current medical practice are generally accepted even though each physician has his own methods of procedure.

(Lee and Jones, 1933)

The purpose of utilization review is to identify those fundamentals of medical practice while recognizing the complexity of medical practice.

D. PROGRAMMATIC CONTEXT OF UTILIZATION REVIEW: ELEMENTS OF UTILIZATION MANAGEMENT PROGRAMS

Although utilization management programs are diverse, it is possible to identify several elements common to many. Listed below, these elements will enable us to establish the context of utilization review and to highlight key issues related to the planning, implementation, and evaluation of utilization management programs.

Goals. Goals may be defined as general statements of the ends the program is designed to attain, such as reducing hospital expenditures, improving the efficiency with which services are provided, detecting and eliminating underutilization, or assuring patient, employee, or beneficiary satisfaction. The particular design of a program will depend on the goals to be attained and the priority given to each.

Objectives. Specific, feasible, clearly stated intermediate steps which, when accomplished, will help achieve the program goals are called objectives. For example, one objective for a program with the overall goal of reducing a corporation's hospital costs might be to reduce the average length of stay of employees with hernia operations to that of members of HMOs, or to institute a mandatory second opinion for selected surgical procedures.

Method of Identifying Inappropriate Utilization. A system or instrument is developed and used to determine which admissions, days of care, or other aspects of the hospital stay are unsuited to the patient's medical needs. Several systems are described in Section II.

Method of Implementation. Questions that must be answered in implementing the program include whether to conduct the program in-house or through a vendor or contractors; who in the organization will be responsible for implementing an in-house program or monitoring a vendor-provided program; the roles and responsibilities of the reviewers (e.g., review only, case management, health education, etc.); how to train, support, and monitor reviewers; whether to conduct reviews on a preadmission, concurrent, and/or retrospective basis; whether to conduct reviews on site in the hospital or remotely by telephone; how to deal with unions concerned with benefit cutbacks or with beneficiaries of public programs concerned about losing entitlements; and who to include in planning and monitoring the program.

Information. Information on hospital use is derived from patients' medical records or from insurance company, corporate, or Medicare/Medicaid claims files.

Data Cleaning, Processing, and Analysis. Most utilization management programs involve large numbers of reviews. They therefore require electronic data processing. This can be provided either in-house or purchased from a vendor.

Interventions. Deliberate actions may be taken by a payer, an agent of the payer (such as a utilization management vendor), or a provider to reduce inappropriate utilization.

Interventions can include (1) providing *feedback* to physicians or providers (e.g., information comparing appropriateness rates of different physicians or hospitals) in order to encourage more efficient and effective practice; (2) conducting *continuing education* seminars or courses on practice protocols; (3) *redesigning benefit plans*, (e.g., adding coverage for HMO or PPO membership, voluntary or mandatory *second surgical opinion programs*, *health promotion/disease prevention programs*, or *alternative types or sites of care* (such as alcohol, mental health, or rehabilitation counseling, halfway house treatment, or home health care services); or (4) using *sanctions* such as withholding payments from physicians or hospitals or restricting the hospital privileges of physicians.

Figure 5 illustrates the cost-containment programs most frequently used in several corporations or planned for introduction in the near future (Grossman and Magnus, 1986). The programs most frequently cited for introduction in the next two years are preadmission certification (listed by 12 percent of the respondents); utilization review (12 percent); increased deductibles or copayments (11 percent); and mandatory second surgical opinions (10 percent).

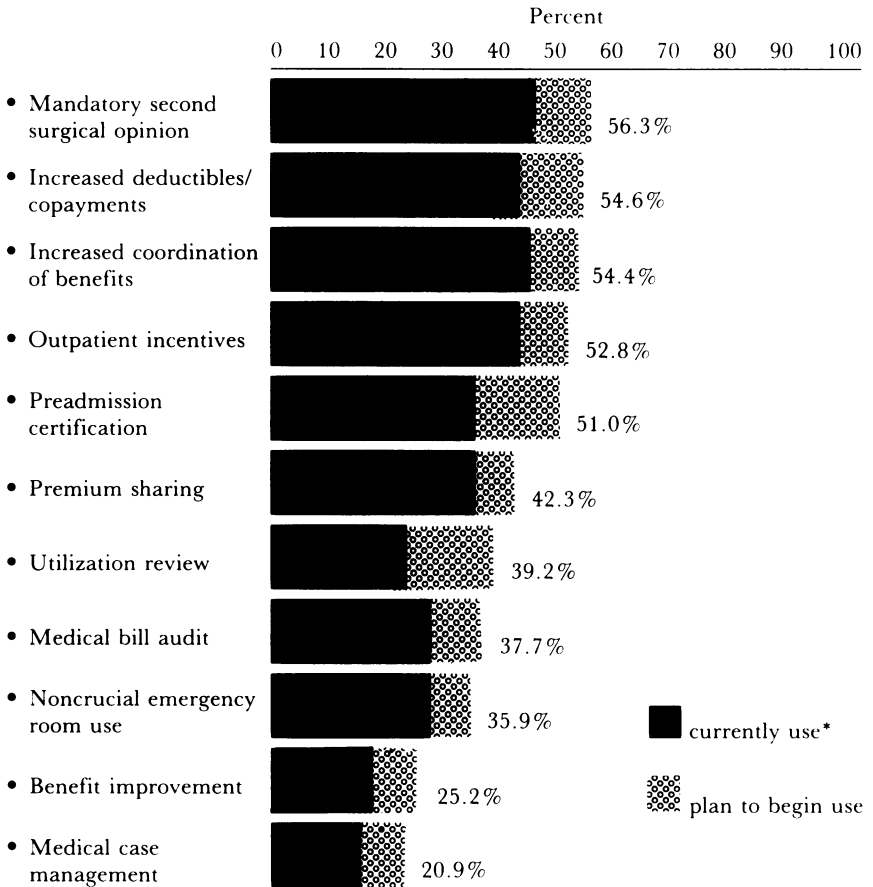
Evaluation. The program's degree of success in reaching its goals and objectives is assessed. Evaluation results can be used to guide program planning or to calculate the relative costs and benefits of the program or parts of the program.

II. CHARACTERISTICS OF METHODS USED TO IDENTIFY INAPPROPRIATE UTILIZATION

A. OVERVIEW OF UTILIZATION REVIEW METHODS AND INSTRUMENTS

This section describes the historical context of utilization review and several specific utilization review methods.

Figure 5: Health Care Cost-Containment Measures Used or Planned by Corporations



*Shaded area indicates those organizations that plan to begin this cost-containment method within the next two years. Nonshaded area indicates those organizations that currently use this cost-containment method.

Source: Grossman and Magnus, 1986. Used with permission.

The methods discussed have been categorized according to whether they use implicit criteria, explicit criteria, or a combination of both. In methods using implicit criteria, the "physicians review the entire patient record and make summary judgments of whether the

process of care is acceptable. In an explicit criteria audit, criteria for acceptable care are predetermined, and nonphysicians then review medical records to ascertain whether the care rendered has met those criteria” (Noren, 1982).

The explicit criteria methods are also divided — between *diagnosis-specific* methods (having different sets of criteria for patients with different diagnoses) and *diagnosis-independent* methods (having one set of criteria which applies to all patients).

Table 1 places the methods to be discussed along the implicit/explicit criteria continuum and indicates whether the explicit criteria methods are diagnosis-specific or diagnosis-independent.

For the purposes of this discussion, a broad continuum of possible methods is included, although in practice today only the diagnosis-independent criteria lists are widely used for utilization review. The full range of methods is discussed for two reasons.

First, in order to evaluate a particular review method, it is important to understand the approaches that could have been used instead. None of the methods is perfect, so the choice is always from among a number of less-than-perfect options. It is easier to assess the limitations

Table 1: Mapping of the Review Methods along Implicit/Explicit Dimension and Diagnosis-Specific/Diagnosis-Independent Dimension

	<i>Implicit Criteria Methods</i>	<i>Intermediate Methods</i>	<i>Explicit Criteria Methods</i>
Diagnosis-specific	Early quality assurance studies (e.g., Morehead's studies)	Guidelines	Protocols
			Optimal care criteria
			Clinical algorithms
			Tracers
		Length of stay profiles	Sentinel events
Diagnosis-independent			Criteria lists (e.g., Level of Care, ISD-A*, AEP†, and SMI‡)

*ISD-A: Intensity of Service, Severity of Illness, and Discharge Screens-Appropriateness.

†AEP: Appropriateness Evaluation Protocol.

‡SMI: Standardized Medreview Instrument.

of a specific method if one is familiar with drawbacks of the other options. For example, the limitations of diagnosis-independent methods, which assess primarily the timing and location of care, can be balanced against the difficulties in developing, testing, and updating diagnosis-specific methods which, while extremely cumbersome and complex, can be used to review several dimensions of care.

Second, utilization review methods grew out of quality assurance methods. To understand the family history and development of utilization review methods, some knowledge of their precursors from the field of quality assurance is useful.

Earliest Methods Designed for Individual Hospitals or Research Studies⁵

Although two early quality assurance studies were the first to address the problem of measuring the appropriateness of hospital care (Codman, *c.* 1916; and Lee and Jones, 1933), utilization review was not used widely until World War II, when internal hospital committees used it to monitor utilization and to free up beds during the wartime bed shortage.

Growing Pressure for Generalizable Methods

Several developments since the Second World War have moved utilization review from the individual hospital to the public realm. In the 1960s, the rapid increase in expenditures for hospital care led the federal government to pressure health insurance companies and hospitals to decrease length of stay. The legislation establishing Medicare and Medicaid in 1965 required participating hospitals and extended care facilities to have utilization review plans. The 1972 amendments to the Social Security Act provided for the establishment of Professional Standards Review Organizations (PSROs) to monitor the quality and appropriateness of care provided to Medicare and Medicaid beneficiaries.⁶ The requirement of the Joint Commission on the Accreditation of Hospitals in 1975—that hospitals perform a specified number of medical record audits in order to be accredited—also encouraged the wider use of utilization review. Thus, by the mid-1970s, both the hospital industry's own accrediting body and the federal government, the single largest purchaser of health care in the country,⁷ mandated utilization review programs. By 1976, 90 percent of the nation's hospitals had utilization review programs in place (Gertman et al., 1979).

These movements toward the wider use of utilization review in monitoring care and accrediting hospitals encouraged the development of methods which were standardized, transferable, reliable, valid, and widely accepted.

B. METHODS USING IMPLICIT CRITERIA

Description of the Methods

Using implicit criteria methods, the physician reviewer applies his or her clinical judgment to the quality and/or the appropriateness of the care provided. All aspects of appropriateness (location, timing, intensity, and volume) and of quality (except the interpersonal relationship between the physician and the patient) can be included in the review. Neither the information to use in making the judgment nor the process for weighting the various aspects of the information reviewed is spelled out by the researcher. The validity of these methods therefore depends entirely on the knowledge, skill, and judgment of the reviewer (Donabedian, 1982). For example, in several studies Morehead made extensive use of implicit methods to judge quality of care (Ehrlich, Morehead, and Trussell, 1962; Morehead, 1967; Morehead and Donaldson, 1974). In her study of the quality of care received by Teamsters in the New York City area (Morehead et al., 1964), reviewers were briefed on the purpose of the study and their general role in it, and were asked to use their clinical judgment in assessing the quality of the medical care provided. The entire medical record was available for their review, and no instructions were provided on how to weight various aspects of the review.

Advantages and Disadvantages of Methods Using Implicit Criteria

The Reviewer's Judgment: Critical in Using Implicit Criteria Methods. The dependence of implicit criteria methods on the judgment of the assessor is both a strength and a weakness, since these methods are only as reliable and valid as the judgment of the physicians making the assessments (Donabedian, 1982).

While Morehead and other proponents of implicit criteria methods consider them to be more valid than explicit criteria methods (i.e., coming closer to reflecting "true" appropriateness because the reviewer can take account of all relevant factors influencing clinical decisions and quality of care), Morehead has concluded that "not all physicians, even the most eminent, can perform this task in a constructive, analyti-

cal fashion . . ." and that the success of implicit criteria methods depends on selecting and training reviewers carefully, structuring the review process, and resolving differences between or among reviewers systematically (Morehead, 1976). This represents a move away from totally implicit, unstructured methods and toward the intermediate methods discussed further on in Section D.

Low Agreement Rates Across Reviewers. Implicit criteria methods have been characterized by low inter-rater reliability, that is, low levels of agreement among two or more reviewers examining the same records (Payne, B.C., 1977; Richardson, 1972; Gertman and Restuccia, 1981). "Whether reviewing entire records or abstracts of records, physicians agree with one another on judged quality of care at or near the level expected by chance alone" (Sanazaro, 1980).

Chance of Reviewer Bias. Low inter-rater reliability may be due to random differences, differences in practice patterns or judgment, or actual bias. Evidence of the potential for reviewer bias comes from studies in which opinions of admitting physicians were solicited regarding the appropriateness of admission timing and level of care. Physicians external to the hospitals studied found almost twice as many cases inappropriate as did physicians internal to the hospital reviewing the same cases (Fitzpatrick et al., 1962; Anderson and Sheatsley, 1967). Donabedian reports similar findings from nine studies he reviewed (1973).

The chance of bias in such an assessment highlights not only one of the major problems with the unstructured approach but also the importance of *who* is asked to make the determination of appropriateness.

Table 2 summarizes the advantages and disadvantages of implicit criteria utilization methods.

C. METHODS USING EXPLICIT CRITERIA

Explicit review methods provide specific criteria for the reviewers and spell out the review process in detail. Lembcke, one of the first developers of explicit criteria methods, has said that "ideally, explicit criteria should be objective, verifiable, uniform across different hospitals, physician specialties, and types of patients. . . . pertinent, and acceptable" (Lembcke, 1956).

Explicit criteria methods range along a continuum from diagnosis-specific to diagnosis-independent.

Table 2: Comparative Advantages and Disadvantages of Several Utilization Review Methods and Instruments

	<i>Methods</i>			
	<i>Implicit Criteria Methods</i>	<i>Length of Stay Profiles</i>	<i>Diagnostic-Specific Protocols</i>	<i>Diagnostic-Independent Criteria Lists</i>
<u>Method Scope</u>				
Levels of care	Yes	No	Yes	Yes
Timing of admission	Yes	No	Yes	Yes
Intensity/Volume of service	Yes (including underutilization)	No	Yes (including underutilization)	No
Length of stay	Yes	Targets patients for further review	Yes	Yes
Quality of care	Yes	No	Yes	No
Need for services	Yes	No	Yes	No
<u>Method Advantages</u>	All relevant information about patient care can be included in the review	Easy to identify cases for further review	All relevant information can be included in the review	Relatively few criteria compared with protocols
	Relatively inexpensive to develop	Can be applied by trained nonphysician reviewers	Can be applied by trained nonphysician reviewers	Standardized and transferable
			Can readily be applied to different types of institutions and providers	Can be applied by trained non-physician reviewers
				Validity of the decision does not depend on the accuracy of the diagnosis
				Easier to update than protocols

Table 2: Continued

<i>Methods</i>			
<i>Implicit Criteria Methods</i>	<i>Length of Stay Profiles</i>	<i>Diagnostic- Specific Protocols</i>	<i>Diagnostic- Independent Criteria Lists</i>
<p><u>Method Disadvantages</u></p> <p>Expensive to apply because physicians must perform all reviews</p> <p>Judgment is subjective</p>	<p>Large number of cases must be reviewed</p> <p>Errors in diagnoses may result in misclassification</p> <p>Inappropriate utilization missed in short and average-length stays</p> <p>Possibility of "DRG creep"</p> <p>Still must assess actual appropriateness using some other method</p>	<p>More expensive to develop and update</p> <p>Time-consuming to apply</p> <p>Errors in diagnoses result in misclassification</p>	<p>More expensive to develop and test than implicit criteria methods</p> <p>Use of reviewer overrides can introduce subjectivity</p>

<i>Methods</i>			
<i>Implicit Criteria Methods</i>	<i>Length of Stay Profiles</i>	<i>Diagnostic- Specific Protocols</i>	<i>Diagnostic- Independent Criteria Lists</i>
<u>Method Reliability</u>	Inter-rater reliability unacceptably low Reviewer bias can reduce reliability and viability	Presumably high, but not reported Reliability and validity depend on (a) the method used to group the patients into diagnostic groups and (b) the method actually used to assess appropriateness	ISD-A* and Level of Care: not reported AEP†: Inter-rater, overall, and specific agreement are high, as tested by developers and others SMI‡: Low reliability reported by external evaluator
<u>Method Validity</u>	By definition low due to low reliability of the method	Not reported for many systems Requires testing across geographic regions, different practice settings, etc.	Each instrument tested for face validity by medical committees during development phase AEP tested by developers for content and convergent validity with generally satisfactory results

*ISD-A: Intensity of Service, Severity of Illness, and Discharge Screens-Appropriateness.

†AEP: Appropriateness Evaluation Protocol.

‡SMI: Standardized Medreview Instrument.

Diagnosis-Specific Methods

In diagnosis-specific methods, distinctive guidelines are established for categories of patients with specified diagnoses or signs and symptoms. Because of the large number of diagnoses and possible treatments (e.g., the more than 8,000 diagnoses and 2,500 procedures listed in the ICD-9-CM codebooks used to code a patient's record information into a discharge summary); the wide range of patients' responses; and the need to incorporate variations in accepted practice patterns as well as advances in technology, diagnosis-specific methods require much more complex instruments and more structured review procedures than do the methods that rely on implicit criteria. (While the review process is also complex in methods using implicit criteria, the complexity there lies in the reviewers' assessment of the clinical decision-making process, not in the written review methodology itself.)

The challenge facing the developers of diagnosis-specific methods is (1) to identify each situation that the reviewer possibly might encounter, (2) to meet the "opposing objectives of specificity and applicability to the full range of practice" (Donabedian, 1982), and (3) to develop methods that are feasible to implement and readily adaptable to technological change.

Protocols. One way of meeting these challenges is to limit the criteria sets to a few specific diagnoses. For example, Lembcke developed a protocol in which several aspects of quality of care were reviewed—confirmation of the clinical diagnosis, justification of surgery, justification of giving or withholding chemotherapy, the extent or completeness of surgery, justification of the necessity for hospitalization, and the preventability of death—for a selected group of gynecological surgical procedures (Lembcke, 1956). Even though Lembcke's criteria applied to a limited number of procedures, he was not able to define absolute criteria, even for those few procedures, which could be applied in every situation. Instead, he allowed for *compliance ranges* within which differences in judgment or practice would be considered acceptable. For example, it would be considered acceptable if 80 percent of a physician's hysterectomies met the criteria for justifying that procedure and 90 percent of the operations were considered complete.

Rosenfeld described what for six diagnoses would constitute "good," "fair," and "poor" care regarding the history, the physical examination, laboratory and pathology procedures, x-ray examinations, screening procedures, consultation, diagnosis, medical therapy, surgical therapy, and anesthesia (Rosenfeld, 1957). He provided protocols for abstracting data from records, determining ratings for each aspect

of care, and aggregating the individual ratings into a single score. Donabedian notes, however, that the "guidelines for assessment, which have never been published, are too lengthy to reproduce and too complex to summarize easily" (Donabedian, 1982). And this is for only six diagnostic categories!

Even among explicit criteria sets there are differences in the degree to which individual judgment is exercised in the review. For example, compare "complication under control" and "afebrile, 99.4 degrees Fahrenheit or below" as two different levels of specificity provided in structured systems (Donabedian, 1982).

As researchers and policymakers move from broad-brush studies measuring rates of inappropriate hospital utilization toward focused attempts to change practice patterns, there is renewed interest in explicit criteria sets. For example, procedure-specific protocols are being used in a large study of variations in surgical rates conducted by the RAND/UCLA Health Services Utilization Study. Teams of physicians and health services researchers are developing guidelines to assess the appropriateness of six procedures: coronary angiography, coronary artery bypass graft surgery, cholecystectomy, diagnostic upper gastrointestinal endoscopy, colonoscopy and carotid endarterectomy (Solomon, Brook, Fink, et al., 1986).

Other Methods. Other diagnosis-specific explicit criteria methods include optimal care criteria (developed by B. C. Payne and colleagues and discussed in Sanazaro, 1980); clinical algorithms and criteria mapping (Sanazaro, 1980); tracers (Kessner et al., 1982); and sentinel events (Rutstein et al., 1976). (See Demlo, 1983, and Noren, 1982, for overviews of these methods.) Like the protocols discussed above, they were originally developed for use in assessing whether diagnoses and treatment are appropriate to the patient's need. As such, they are more readily adapted for use in assessing underutilization than overutilization.

Disadvantages of Explicit Criteria Methods. In spite of their level of detail, diagnosis-specific methods are not detailed enough to make possible definite conclusions regarding appropriateness. There are several reasons for this.

- *Operationally*, these reasons include the practical need to limit the scope of the procedures or diagnoses covered by criteria and the need to provide "compliance ranges" instead of absolute standards of compliance due to factors such as the potential errors in the coding information used to classify patients into diagnostic groupings (Demlo, Campbell, and Brown, 1978); individuali-

ties of patients; or vagueness and overlap in the ICD-9-CM categories (Donabedian, 1982; Hornbrook, 1982; and Iezzoni and Moskowitz, 1986).

- *Conceptual limitations* of diagnosis-specific methods include difficulties in assigning diagnoses (Eddy, 1984); difficulties in classifying patients not given specific diagnostic codes but given codes related to signs and symptoms; how to reduce the potential 400,000 combinations of principal diagnoses/age/treatment modes/complications/comorbidities (Pettengill and Vertrees, 1982) into a manageable number of diagnostic categories; and whether the basis of classifying patients should be the *principal* diagnosis (the condition established *after study* to have been chiefly responsible for occasioning the admission of the patient to the hospital), the *primary diagnosis* (the condition that was responsible for the consumption of the most resources during a hospital stay), or the *admitting* diagnosis (the condition recorded upon admission as the reason for admission to the hospital).⁸ The admitting and the principal diagnoses are frequently different (Burford and Averill, 1979).
- In addition, it is necessary to decide whether to specify *optimal criteria* (the “best” care possible), *minimal care* levels, or *average care*.

Diagnosis-specific methods also pose some practical problems:

- They are vulnerable to errors and uncertainty on two levels—in assigning patients to diagnostic categories and in assessing appropriateness itself. In contrast, diagnosis-independent methods (discussed below) are vulnerable only in the assessment of appropriateness.
- The need either to review a *large number of cases* in order to have enough cases in each diagnostic category to make comparisons across physicians or hospitals, to *limit the analysis* to a few diagnostic groups, or to *lump unrelated diagnoses* into larger groups, which may not be medically meaningful (Borchardt, 1981).

For all of these reasons, diagnosis-specific methods, like the diagnosis-independent methods described in the next section, function with greatest validity and acceptability as screening devices—not as the final, absolute indicators of quality or appropriateness (Donabedian, 1982).

Diagnosis-Independent Criteria Lists

Earliest Criteria List: Level of Care Only. The earliest diagnosis-independent criteria (Goldberg and Holloway, 1975; Holloway et al., 1975) assessed only the appropriateness of the level of care, that is, whether or not the services provided justified hospitalization (Donabedian, 1982). The medical need and the timing of the services were not assessed. The reviewer used professional judgment to assess whether or not the services which the patient received justified acute-level care, but the criteria did not specify which services were appropriate to hospital care.

In order to broaden the scope of criteria lists and to make the criteria more specific, developers of subsequent lists added criteria related to patient morbidity or severity of illness (which relate to signs, symptoms, and laboratory values, but not to specific diagnostic categories) and specified services that would justify hospital care. Thus, just as the developers of implicit criteria lists gradually added more structure to their methods, developers of diagnosis-independent criteria lists added some criteria related to specific conditions and services.

Lists in Current Use. Three criteria lists are in common use today:

1. *The Intensity of Service, Severity of Illness, and Discharge Screens (ISD)* review system, developed by InterQual, Inc. in 1978, and revised several times since (InterQual, 1987). In 1984, InterQual developed an Appropriateness (A) component that was added to the ISD review system, which is now known as the ISD-A review system. It covers inpatient hospitalization of adult medical, adult surgical, obstetric, and pediatric patients, including support services (e.g., physical therapy, occupational therapy, speech therapy, respiratory therapy, and total parenteral nutrition), and care in coronary care, intensive care (including neonatal and pediatric), progressive care, and rehabilitation units.
2. *The Appropriateness Evaluation Protocol (AEP)*, based on level-of-care criteria developed by Goldberg and Holloway (1975), barriers to appropriate utilization identified by Restuccia and Holloway (1976); and the ISD admissions criteria. It was developed at Boston University in the late 1970s and early 1980s, and was subsequently revised (Gertman and Restuccia, 1981; Gertman and Restuccia, undated; Restuccia, Payne et al., 1986b). The AEP assesses the appropriateness of the timing and level of care of adult medical, surgical, gynecology, and obstetric patients and noninfant pediatric patients, and the

medical need for selected medical and surgical procedures. AEP criteria are being developed for psychiatric cases.

3. *The Standardized Medreview Instrument (SMI)*, which was developed by Systemetrics, Inc. from 1980 to 1983. It builds on the ISD-A generic criteria list and the 13 system-specific lists, with the system-specific distinction used in the ISD-A eliminated and the criteria collapsed into a single list. The use of the reviewer override was incorporated from the AEP, as were modifications of the AEP reasons and responsibilities lists. Representatives from 50 PSROs were involved in the development process.

The SMI covers adult medical/surgical, psychiatric, intensive care unit/coronary care unit, burn, alcohol/chemical dependency, rehabilitation, and trauma/shock patients (Longest et al., 1984; Terrell, 1983; Systemetrics, 1984).

The AEP and the SMI were developed under a grant and a contract, respectively, from the federal Health Care Financing Administration. Both instruments are in the public domain. The ISD-A is proprietary.⁹

How the Instruments Are Used. With each instrument the admission, one or more days of care, and the patient's readiness for discharge are reviewed, and in each both the services provided and the patient's severity of illness and stability are reviewed in making the decisions. Each instrument consists of a set of generic criteria which are applied, regardless of the patient's diagnosis or symptoms, to determine whether inpatient care is justified. Criteria can relate to either the patient's severity of illness (e.g., pulse rate of less than 50 or more than 140 per minute; active bleeding) or to specific services (e.g., intravenous medications and/or fluid replacement, treatment in an ICU, vital sign monitoring every two hours or more). In addition to the generic criteria, the instruments may include criteria which apply to specific conditions or symptoms (for example, fracture or dislocation of the spine or bizarre or delusional behavior (InterQual, 1986)).

In each instrument the relevant information reviewed comes from the medical record, including the history and physical examination, progress notes, nurses' notes, consultation notes, and notes of other providers such as physical therapists, discharge planners, or social services personnel.

The instruments are designed to be used by nonphysician reviewers, with recourse to physician consultants in cases of ambiguity.

The *ISD-A* includes a generic criteria list, which is applied to all

patients, and system-specific criteria lists for 12 body systems and for psychiatric patients. There are approximately 69 generic criteria items for severity of illness, intensity of services provided, and readiness for discharge, and 20–60 criteria for each of the body systems. Lab-value cutoff points are specified (e.g., a hemoglobin level of below 7 grams or above 20 grams to justify inpatient hospitalization).

The ISD-A has a relatively complex set of decision rules to determine appropriateness. “Any patient admitted to the hospital must have met either one severity of illness (SI) *or* one intensity of service (IS) criterion on admission and must have met *both* an SI and an IS criterion by the first review following the completion of 24 hours in the hospital” (InterQual, 1986). Reviewers can exercise judgment in cases in which, although no one lab value justifies hospitalization, a number of lab values taken together are close enough to the cutoff points to justify hospitalization.

The *AEP* includes generic and system-specific criteria which were adapted from the ISD-A criteria, after modification by representatives from PSROs and physicians.

The decision rule is relatively simple. If only one of the 16 admission criteria is met, the admission is deemed appropriate; if only one of the 26 day-of-care criteria is met, the day is appropriate. Specific lab values are indicated as cutoff points. There is a provision for overrides by the reviewer or a physician consultant for cases in which the criteria do not capture a patient’s situation. The reviewer can override the criteria in either direction: if the criteria indicate an appropriate admission or day but the reviewer does not judge it to be so, the criteria can be overridden; or, if no criteria are met but the reviewer deems the admission or day appropriate, he or she can enter it as appropriate.

The instrument provides a list of reasons explaining the cause of inappropriate admissions and days, and a list of persons, institutions, or factors (patient/family, hospital/physician, or environmental factors) considered to be responsible for inappropriate days.

In the *SMI*, the 117 admission criteria are divided into four sections, listed in order from the most to the least objective: laboratory abnormalities; medical problems/diseases/complications of medical or surgical care; abnormal signs; and symptoms. If an admission meets one criterion, it is considered appropriate. A day must meet one of the 30 level-of-care criteria and one of the 26 continued-stay criteria to be considered appropriate.

Testing of the Instruments. To be useful, any utilization review instrument must be reliable and valid. Several of the aspects of reliability and validity which can be tested are listed in Table 3. They are

Table 3: Criteria for Assessing the Performance of Utilization Review Instruments

-
- **Reliability.** Consistency; a reliable instrument will give the same result when applied more than once under the same conditions (e.g., when the same reviewer reviews a medical record on two separate days, or when two reviewers review the same record [inter-rater reliability]). Three kinds of inter-rater reliability are usually measured:
 - Overall agreement. The percentage of cases on which both or all reviewers agree on an appropriateness decision.
 - Specific agreement. The agreement rate on only those cases that were judged inappropriate by at least one reviewer.
 - Kappa statistic. An adjustment for the amount of agreement occurring by chance alone (Cohen, 1960).
 - **Validity.** The ability of the instrument to reflect "true" appropriateness.
 - Face validity. The extent to which an instrument measures what it is designed to measure, as determined by expert review of the instrument.
 - Content validity. The extent to which the instrument "covers" the concept it is designed to measure (does it include all the relevant aspects of the concept, not leaving out any important aspects?)
 - Convergent validity. The extent to which decisions based on the instrument agree with those made by clinicians using expert judgment.
-

described in more detail in Cook and Campbell (1979) and Carmines and Zeller (1979).¹⁰

All three instruments were tested for face validity by the medical panels used in developing them.

No results of any formal evaluations of the reliability and validity of the ISD-A have been reported in the literature. The original SMI was tested by the developers (Longest et al., 1984), who reported satisfactory reliability results. No reliability results for the revised SMI have been reported, however, by the developers (Blumenfeld, 1985). The AEP has been extensively tested by the developers and others for inter-rater reliability (Gertman and Restuccia, 1981; Gertman and Restuccia, undated; Restuccia et al., 1986b; Rishpon et al., 1986; Siu et al., 1986) and for face, content, and convergent validity (Gertman and Restuccia, undated; Restuccia et al., 1986b), with satisfactory results reported.

The relative reliability and validity of the ISD-A, AEP, and SMI are being addressed in a study underway by Ira Strumwasser of the

Michigan Health Care Education and Research Foundation, Detroit, Michigan.

The results of the study indicate that the Standardized Medreview Instrument, when applied by RNs newly and extensively trained in applying the criteria, is unreliable in assessing both the necessity of admissions and the need for continued days of stay. The SMI appears to be so unreliable, as applied by RNs with three months of audit experience using the SMI, that its further use in utilization review activities is not recommended. . . . The AEP is an extremely reliable instrument for identifying non-acute medical admissions and days of care in acute care hospitals.

(Strumwasser, Paranjpe, et al., 1987)

Due to its low reliability, the SMI was not tested for validity. The validity of the AEP was tested by comparing the results using it with the professional judgments of three physicians reviewing the same records. Quantifying the validity of the AEP was complicated by the low inter-rater agreement among the physicians, a difficult methodological problem often encountered when using unguided physicians' opinions as a validating standard.

The authors emphasize the importance of using the AEP as a screening tool when it is used to make preadmission, continued stay, or reimbursement decisions for individual patients, with records of patients deemed inappropriately hospitalized referred to physician advisors for definitive review (Strumwasser, Paranjpe, et al., 1987).

Self-Development of Criteria Lists. An organization or group could develop its own list of criteria. Such development should not be undertaken lightly, however, since it is a complex and demanding task. For example, the Mecklenburg County (North Carolina) Health Care Cost Management Council has developed a review program:

The basis of the program is a set of diagnosis and procedure specific criteria written by panels of local practitioners. Over 350 area physicians have participated in developing the criteria. There are now 23 panels, organized by specialty and subspecialty. Each panel is given breakdowns of the hospitals' most frequent diagnoses in its specialty, and it writes guidelines for length of stay and level of services. . . . These guidelines are sent for comment to every area physician who has admitting privileges in a (county) hospital, and substantive comments are referred back to the panels and reviewed.

(*Medical Practice Newsletter*, February 1986)

A simpler method would be to begin with a preexisting criteria list, adapting it to local practice patterns. In either case, the new or modified instrument should be tested for reliability and validity.

Advantages and Disadvantages of Criteria Lists. Peter Borchardt lists several advantages of diagnosis-independent criteria lists, especially in comparison with length-of-stay profiles:

- They are not affected by errors in diagnostic coding and “DRG creep.”
- They avoid discussions with physicians on whether DRGs accurately describe their patient mix and the severity of illness of their patients.
- Problems with inappropriate admissions are specifically identified and not masked by low average lengths of stay.
- Case-mix adjustments are unnecessary.
- Even a small sample (such as 100 discharges) can represent each physician’s total hospital practice and identify variations in non-acute hospital-days when compared to other physicians’ practice patterns, since it is not necessary to collect large numbers of cases to make comparisons across diagnostic groups.

(Paraphrased from Borchardt, 1981)

Diagnosis-independent criteria lists have two major disadvantages:

- As with all utilization review methods based on specifying explicit criteria *a priori*, the lists are screening tools, not the definitive arbiter of appropriateness (Donabedian, 1982). They are best applied to aggregates of patients to find patterns in inappropriateness and to flag patients for further review by clinicians, but are not to be used alone to make decisions regarding admission, discharge, or reimbursement.
- Use of the reviewer override of the criteria in the AEP and SMI for cases to which the criteria do not apply introduces the possibility of reviewer bias, particularly with in-house reviewers (Restuccia et al., 1986). Thorough reviewer training and supervision and monitoring of the use of overrides is necessary to assure that the criteria are used effectively.

D. INTERMEDIATE METHODS

In intermediate methods, specific criteria are not defined but various aspects of the review process are specifically formulated. For example, researchers can define what information should be reviewed; whether the outcome or process of care is to be used in determining quality or appropriateness (Brook et al., 1977); the weights to be given to each

aspect of care reviewed (Butler and Quinlan, 1958); or what to look for and how to arrive at an overall assessment (Peterson et al., 1956).

Examples of studies using intermediate methods are Wenkert and Terris, 1960; Zimmer, 1967; Browning et al., 1969; and Berg et al., 1970.

Length-of-Stay Profiles

An intermediate method that deserves special mention because it has been widely used in hospitals and PSROs is the length-of-stay profile. In this method, ICD diagnostic codes with similar lengths of stay are grouped together into categories. (In fact, the DRG system was originally developed to group patients for length-of-stay profile review.) The average length of stay for each category and certain percentiles of the length of stay are determined. A target percentile (for example, the 50th percentile) is selected as the flag to indicate that review is warranted. The reviewer then inspects the records of each patient for whom the length of stay has exceeded the percentile for the appropriate category. It should be noted that the profile method only targets patients for further review. In order to assess appropriateness, implicit criteria or an explicit criteria method must be used.

Length-of-stay profiles have several limitations:

- They indicate that a problem may exist (such as unnecessarily long stays) but do not identify the source of the problem (Borchardt, 1981). Another method or instrument must be used to assess the appropriateness of the length of stay and the cause of the problem.
- They are more “gameable” than many other review approaches. The chance of “DRG creep,” in which the diagnosis or the patient’s complications or comorbidities are coded in the hospital in order to maximize the patient’s documented severity of illness—hence the hospital’s reimbursement under case-mix-based prospective payment systems (Simborg, 1981)—may distort the review process. If a patient is placed in a category with a higher severity level due to coding practices designed to maximize reimbursement, he or she will probably have a shorter length of stay than others who are properly in the category. The length-of-stay profile will not identify the patient as needing further review because he or she has a shorter length of stay than others in the category.

- There is no clear relationship between diagnosis, length of stay, and appropriateness. Patients with short stays or average stays can be inappropriately admitted or have stays that are unnecessarily long. Length-of-stay profile methods that target patients with unusually long stays will miss inappropriate utilization occurring among other patients.

E. SUMMARY

The first appropriateness review studies were conducted for a variety of purposes, including quality-of-care assessment, health systems planning, and facility planning. Reflecting those early purposes, the first studies of appropriateness relied almost exclusively on unstructured review methods and implicit criteria. They were not designed to be generally applicable or standardized. Because of changes in the purpose of utilization review, the early methods gradually gave way to generalizable methods which could be tested, validated, and standardized for use across hospitals and geographic regions. The early methods were unstructured, placing heavy reliance on professional judgment and implicit criteria; later methods were more structured, using explicit criteria and specifying the information to be reviewed, what should be looked for in reviewing the information, how to evaluate it, and how to weight the various findings to come to a judgment. Early methods were often diagnosis- and procedure-specific, but as time went on the difficulties of developing and updating such systems became more apparent.

The move from unstructured, subjective methods to structured, objective methods shifted responsibility for the reliability and validity of the methods from the individual reviewers to the developers of the instruments. "In the fully explicit and specified approach, the nature of the criteria and the method used in their application determine the conclusion, whereas with the implicit, unstructured approach, the conclusion depends on the knowledge, skill, and judgment of the assessor" (Donabedian, 1982). Accordingly, developers of later methods have been more rigorously challenged in terms of both the reliability and the validity of the instruments and methods.

In resolving the dilemma of how to be both comprehensive and explicit, the earliest developers of explicit criteria methods established diagnosis-specific methods. The problems inherent in that approach, however, caused later developers, in varying degrees, to develop diagnosis-independent methods. At the extreme, that approach is limited to reviewing only location of care, so later methods have incorporated at least some condition- and service-specific criteria.

Finally, the methods are differentiated into those that are *proscriptive*, stating what care should be provided to patients (e.g., what care should be provided to patients in the upper left quadrant in Figure 3) or *prescriptive*, stating what care is not required by patients. Proscriptive methods are more appropriate to quality assurance studies and to detecting underutilization, and prescriptive methods are more appropriate to detecting overutilization.

Continued testing of the instruments for reliability and validity are needed as they evolve and as practice patterns change. It should be noted that, although all three criteria lists are in use by PROs and third-party payers, only the AEP has been extensively tested for reliability and validity. Results of the systematic testing of the ISD-A, if it has been conducted, have not been reported. The satisfactory results regarding the reliability of the SMI reported by the developers apply to the original instrument only and were not replicated in the external evaluation by Strumwasser.

Finally, continued testing of the instruments for reliability and validity are needed as they evolve and as practice patterns change; and further comparative evaluations of the instruments are needed.

III. AMOUNT OF UTILIZATION CONSIDERED INAPPROPRIATE—AND WHERE IT IS LIKELY TO OCCUR

A. CAUTIONS

Before examining the findings from different studies on the extent of inappropriate utilization and the factors associated with it, it is important to remember several points:

- As discussed above, review methods and instruments vary greatly in reliability and validity.
- Characteristics of the reviewer are critically important to the levels of inappropriateness detected in unstructured methods relying on implicit criteria.
- Apparent differences in rates of inappropriate days may be due to differences in methods of sampling the patients or the days of care or to differences in ways of calculating rates of inappropriate utilization.

Rates of inappropriate days can be presented either for all days reviewed or for appropriate admissions only (the

“adjusted-day” rate, which tends to be lower than the unadjusted rate). Studies in which the day before discharge was sampled will show higher rates than studies in which random days throughout the stay were sampled.

- Relationships established in the earlier studies using univariate statistics may not stand up when multivariate methods are used.

Earlier studies used univariate statistical analysis (for example, to identify the relationship between inappropriate admissions and payer, or inappropriate admissions and patient’s age) to identify factors associated with inappropriate utilization. Later studies used multivariate techniques, in which interactions among several variables could be accounted for. For example, in one study Medicaid patients were found to have lower rates of inappropriate utilization than Medicare patients (Borchardt, 1981). But Medicaid patients are younger on the average than Medicare patients. Without controlling for the interaction between payer and age, it is not possible to know whether the differences in rates of inappropriate utilization were due to age or payer, or to a combination of those factors.

- There may have been changes over time in the amount of inappropriate utilization or the factors associated with it which need to be considered when comparing studies conducted several years apart.
- Results from studies of specific types of patients may not apply to other groups or may not be generalizable to the population at large. For example, findings from studies of Medicare patients, who are primarily elderly, may not apply to Medicaid patients, who are primarily lower-income children and women of child-bearing age, or to employed persons, and so forth.
- There are differences in the variables included in the various studies. The fact that a variable is not found to be significantly related to inappropriateness in a particular study may simply mean that the study did not include the variable.
- Different aspects of appropriateness may have been assessed. For example, many criteria-based instruments assess the timing of the admission and the length of stay of surgical cases; others assess only the appropriateness of the location of care. Studies using instruments which assess the location of care as well as the

timing can result in higher rates of inappropriateness than studies which assess only location.

B. HOW SERIOUS A PROBLEM IS INAPPROPRIATE UTILIZATION?

Despite differences in methods and changes over time, one consistent finding from the studies reviewed is the existence of a substantial amount of inappropriate utilization. Table 4 presents information on inappropriateness levels from six early studies. The rates range from a low of 9 percent of the days reviewed (Zimmer, 1974) to a high of 18 percent (Rosser, 1976) in these studies.

Higher rates were found by Rosenfeld et al. (1957), Wenkert and Terris (1960), and Van Dyke and colleagues (1963), who found 42

Table 4: Study Populations and Rates of Inappropriate Hospital Bed Use in Early Utilization Review Studies

<i>Study</i>	<i>Hospital Study Population</i>	<i>Inappropriate Use (%)</i>
Querido (1963)	General hospital services, 20 Amsterdam hospitals	17
Browning and Crump (1969)	Medical/Surgical services, Rochester, NY hospitals	14
Gertman and Bucher (1971)	Medical services, Baltimore City Hospital, Baltimore, MD	12
Zimmer (1974)	All clinical services, Strong Medical Hospital, Rochester, NY* (medical service alone)	9 (12)
Rosser (1976)	Medical service, Teaching Hospital, London, England†	18‡
Restuccia and Holloway (1976)	Medical/Surgical services, Herrick Memorial Hospital, Berkeley, CA‡	11

Source: Gertman and Restuccia, 1981.

* Major medical school hospital.

† Community hospital.

‡ Based on total available bed-days, not appropriateness of actually utilized days; correction would make result = 20 percent.

percent, 70 percent, and 41 percent, respectively, of the patients they reviewed on given dates to have been hospitalized inappropriately. It should be noted, however, that in each of the three last-mentioned studies, only "long-stay" patients were included, and they were defined as patients who had been in the hospital at least 30 days; "short-stay" patients were those who had stayed for as many as 29 days!

Table 5 presents information from later studies. Rates of inappropriate admissions range from a low of 6 percent (Longest et al., using the SMI) to a high of 19 percent (Restuccia and Gertman, 1984, using the AEP). The lowest rates in Table 5 are from studies sampling random days of the stay and reporting adjusted days (27, 20, and 4 percent, for the Urban Hospital, National AEP, and National SMI studies, respectively); the highest rates are unadjusted rates from studies in which the day before discharge was sampled (37 and 39 percent — from the Delmarva PSRO and Feedback AEP studies, respectively).

C. WHERE INAPPROPRIATE UTILIZATION IS LIKELY TO OCCUR AND FACTORS ASSOCIATED WITH IT

The medical care process can be thought of as involving the *patient and the family or support system*, the *physician*, the *hospital*, and the *environment* (other facilities, the reimbursement system, environmental conditions, etc.) (Donabedian, 1973). Factors related to inappropriate utilization may be related to any one of those persons or groups; to any combination of them; or to characteristics of the hospital stay itself, such as the length of stay or the part of the stay reviewed.

Patient and Family

Conceivably, the patient or the patient's family can contribute to unnecessary utilization by pressuring the physician to admit (Anderson and Sheatsley, 1967) or by delaying discharge. In addition, patient characteristics such as lack of family support, lack of a healthy postdischarge environment, disorientation, or frailty may influence the physician to hospitalize a patient who needs lower-level care instead (Glass et al., 1977; Donabedian, 1973). Glass and colleagues found that the patients' age, the need to change residence after discharge, disorientation, and the duration of the disorientation were associated with an increased chance of "social" or non-medically indicated days.

In the Delmarva PSRO study (Borchardt, 1981) and the National SMI study using univariate analysis, Medicaid patients were found to have lower levels of inappropriate days than Medicare patients. This may be due to more conservative medical practice patterns in care of

Table 5: Study Populations, Rates of Inappropriate Admissions, and Rates of Inappropriate Days of Care in Recent Utilization Review Studies

<i>Study</i>	<i>Study Population, Review Instrument, Sampling Method Used to Select Cases for Review</i>	<i>Inappropriate Admissions (%)</i>	<i>Inappropriate Days of Care (%)</i>
Delmarva PSRO	Medicare and Medicaid medical/surgical discharges; 2711 admissions and 2523 days of care AEP:DBD*	14†	37†
SMI	All payers. 5732 admissions and 24,051 days of care SMI:RD	6	12 (4 adjusted‡)
Delmarva HDI	Medicare medical/surgical discharges; 2085 admissions and 1995 days of care AEP:DBD	10	20
AEP National	Adult medical/surgical discharges from 4 PSROs; 1232 admissions and days of care AEP:CD	19	27 (20 adjusted)
Feedback	12,071 adult medical/surgical and 10,054 days of care from 8 New England hospitals AEP:DBD	12†	39†
RAND	1132 adult non-Medicare medical/surgical admissions and days of care from 6 cities AEP + AEP Ambulatory Surgery Procedure List	40 (17 due to surgeries that could have been performed in outpatient facilities)	34 (with cost sharing) 35 (without cost sharing)
Massachusetts	4928 adult medical/surgical admissions and discharges in 4 PSROs AEP:CD	19§	20 (adjusted‡)

Continued

Table 5: Continued

<i>Study</i>	<i>Study Population, Review Instrument, Sampling Method Used to Select Cases for Review</i>	<i>Inappropriate Admissions (%)</i>	<i>Inappropriate Days of Care (%)</i>
<i>Urban</i>	<i>297 adult medical/surgical discharges from 1 urban teaching hospital AEP:CD</i>	<i>12</i>	<i>32 (27 adjusted†)</i>

Sources: Delmarva PSRO: Borchardt, 1981; SMI: Longest et al., 1984; Delmarva HDI: Health Data Institute, 1984; AEP National: Restuccia et al., 1984; Feedback: Restuccia, Payne, et al., 1986b; RAND: Siu et al., 1986; Massachusetts: Restuccia, Kreger, et al., 1986a; Urban: Restuccia, Payne, et al., 1987a.

*Key to sampling methods used to select days of care for review:

RD = random day of stay.

DBD = day before discharge.

CD = calendar dates selected; all or some of the patients in the hospital on those dates were reviewed.

† Only the rates for the control group or baseline are reported.

‡ Adjusted rates are the rates of inappropriate days for the patients who were appropriately admitted only. Patients who were inappropriately admitted are excluded.

§ Subjective review of admissions; AEP admissions criteria were not developed at the time of data collection.

the elderly, or perhaps to the younger age of Medicaid patients compared with Medicare patients—which itself may be a proxy for differences in severity of illness (discussed further on). The relationship was not found in the multivariate analysis conducted for the SMI study, however. (No multivariate analysis was conducted to control for interactions in the Delmarva study.)

In the four studies reviewed here which specifically assessed the responsibility of the patient/family for inappropriate days, only 2–5 percent of the inappropriate days (Table 6) were attributed to those actors (Borchardt, 1981; Restuccia et al., 1986b; Restuccia, Kreger, et al., 1986a; Restuccia, Payne, et al., 1987a).¹¹

This finding is substantiated by the RAND Health Insurance Experiment, in which the relationship between cost sharing by the patient and inappropriate utilization, as measured by the AEP, was studied. No significant difference in rates of inappropriate admissions or days of care was detected between patients who shared and those who did not share in the costs of hospital care (Siu et al., 1986). More important, cost sharing was found to reduce the rate of *appropriate*

Table 6: Group, Person, or Factor Deemed Responsible for Inappropriate Days of Care

<i>Study</i>	<i>Rates of Inappropriate Days</i>		
	<i>Physician/Hospital</i>	<i>Patient/Family</i>	<i>Environment</i>
Delmarva PSRO*	82	5	13
Feedback	86	2	12
Massachusetts	72	4	24
Urban	82	4	14

Sources: Delmarva PSRO: Borchardt, 1981; Feedback: Restuccia, Payne, et al., 1986b; Massachusetts: Restuccia, Kreger, et al., 1986a; Urban: Restuccia, Payne, et al., 1987a.

*Baseline period only.

hospitalization, which could adversely affect patients who need hospital care.

In summary, none of the several patient characteristics tested in the studies reviewed—including age, sex, and payer—have consistently been found to be significantly related to inappropriate utilization. The effect of a possible interaction among age, payer, and severity of illness, however, does warrant further study. The patient/family appears to be relatively unimportant in influencing the total level of inappropriate days of care.

Physician and Hospital

Analysis of the factors underlying inappropriate utilization in several AEP studies indicates that approximately three-fourths of the inappropriate days identified are the responsibility of the physician or hospital (Table 6). The most common reasons for inappropriate admissions (premature admission or institutional care not required) and inappropriate days (no further institutional care required or lower-level care required) relate to admitting and discharge practices under the control of the physician and/or the hospital (Table 7).

Although none of the studies reviewed here found variations in inappropriate utilization by *type* of physician, it is increasingly evident, based on related studies, that the *individual* physician probably plays an important role in inappropriate utilization. Evidence of this comes from several sources.

1. Quantitative evaluation of the effects of informational feedback on rates of inappropriate utilization in seven New England hospitals indicated such a wide range of rates of inappropriate utilization among physicians within hospitals, and such a wide range of *changes* in

Table 7: Reasons for Inappropriate Utilization

Study	Admissions: Reasons			
	Premature Admission	Institutional		Other
		Care Not Required	Lower-Level Care Required	
SMI	NR	76	24	NR*
AEP National	43	51		6
Feedback	44	56		NR
Massachusetts†	43	51		6
Urban	40	37	12	11

Study	Days of Care: Reasons		
	Continued Stay Necessary	No Further Institutional	
		Care Required	Lower-Level Institutional Care Required
SMI	9	67	24
AEP National	8	92	
Feedback	NR	78	22
Massachusetts	7	54	39
Urban	8	54	38

*NR = not reported.

† Subjective review of admissions was used because AEP admissions criteria were not developed at the time of data collection.

those rates among physicians within hospitals in response to the feedback that the investigators concluded that the appropriate unit of analysis was the physician and not the hospital (Restuccia et al., 1986b).

2. Recent work on variations in hospital utilization rates across geographic regions and small geographic areas suggests but does not confirm the physician's influence on inappropriate utilization. Several studies have identified extraordinary differences in population-based rates for many common surgical procedures which do not appear to be related to patients' socio-demographic-economic characteristics, to pre-existing health status, or to postdischarge outcome (Chassin, 1983). Many observers have concluded that the differences are due to differences in practice patterns across geographic areas (Wennberg, 1984; McClure, 1982) which, in turn, may be related to differences "among physicians in the evaluation of patients' diagnoses or meeting patient need" (Wennberg, 1982); the surgeon-to-patient ratio (Wilson and Tedeschi, 1984); to other factors. McMahan and Newbold (1986) have concluded that differences in physicians' practice patterns account for more of the variation in length of stay than does severity of illness. (It

should be emphasized that the link between variations in utilization rates and variations in rates of *inappropriate* utilization has not been established, although findings from the National AEP and SMI studies suggest that that may be the case; nor has the role of the physician in influencing rates across geographic regions been established. Detailed studies of physician characteristics are difficult to undertake because of the sensitive nature of information on individual physicians' practice patterns.)

Hospital characteristics were found to be related to inappropriate admission rates in the national SMI study, with lower occupancy rates, nonteaching status, and fewer beds related to higher rates of inappropriate admissions (Longest et al., 1984). Once interactions among variables were controlled for, however, only the *number of beds* was significantly related to inappropriate admission rates.

Hospital size and teaching status were found to be related to inappropriate admissions in the national AEP study, but not in any consistent pattern across the four regions studied. For instance, teaching hospitals had the highest rates of inappropriate admissions in the east-rural and west-urban PSROs in comparison with large and small community hospitals, but the lowest rates in the east-urban and west-rural PSROs (Restuccia et al., 1984).

Only occupancy rate was significantly related to rate of inappropriate *days* in the univariate analysis in the national SMI study.

In summary, no hospital characteristic has been found to be consistently associated with higher rates of inappropriate utilization, although bed-size and occupancy rates appear to warrant further study.

Other Factors

Factors related to the broader institutional setting, such as the form of organization or reimbursement (e.g., prepaid group practices versus fee-for-service payment and individual practice) (Luft, 1978) and the proportion of physicians in the local area practicing in organized settings (Nobrega et al., 1982) have been linked with variations in *rates* of hospital admissions. These findings suggest possible differences in *inappropriate* utilization, but without further study they do not provide definite evidence of a relationship between institutional setting and inappropriate utilization.

The SMI study found that a rural environment, a lower number of nursing home beds per capita, fewer physicians per capita, and a higher percentage of the population with incomes below the poverty

level were associated with a higher level of inappropriate admissions in univariate analysis.

Multivariate analysis, however, showed that only *geographic region* (along with number of beds) was significantly related to inappropriate admissions. Rates of inappropriate admissions were high in the South and low in the Northeast, which parallels overall admissions rates (which are higher in the South and lower in the Northeast) (Figure 6). Rates of inappropriate days were highest in the Northeast (which has the longest average length of stay) and lowest in the West (which has the shortest length of stay) (Longest et al., 1984). In contrast, no relationship between admission rates and rates of inappropriate admissions by geographic region was found in the RAND Health Insurance Experiment (Siu et al., 1986).

Between 12 and 15 percent of inappropriate days of care were attributed to *environmental factors* in three of the AEP studies which examined responsibilities for inappropriate days of care (Restuccia, Kreger, et al., 1986a; Restuccia, Payne, et al., 1987a; Restuccia et al., 1986b) (Table 6).

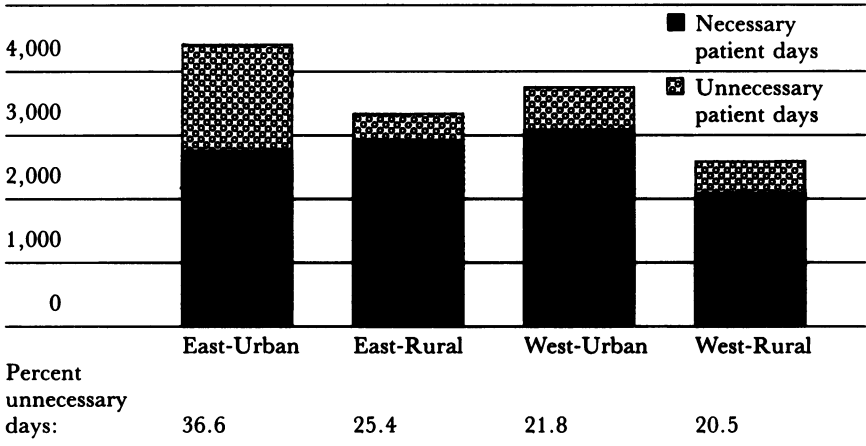
Characteristics of the Hospital Stay

The factor most consistently associated with inappropriate days of care is the *part of the stay* reviewed. All studies reviewed found the highest rates in days in the latter third of the stay (Longest et al., 1984; Restuccia et al., 1984; Restuccia, Kreger, et al., 1986a; Restuccia, Payne, et al., 1987a). This relationship was found under both univariate and multivariate analysis in the SMI study.

The findings are not as clear-cut, however, regarding the absolute *length of the stay*. Several earlier studies found a strong relationship between longer stays and higher levels of inappropriate days of care (e.g., Zimmer, 1974), but inappropriate days were identified in short stays as well (Zimmer, 1974; Fitzpatrick et al., 1962; Gertman and Bucher, 1969). It is important to remember that in earlier studies the definition of "short" stays could be as long as 29 days, which would be considered an extremely long inpatient stay today.

In the SMI study, inappropriately admitted patients had a shorter than average length of stay, and a high percentage of inappropriate admissions was found among stays of 3-5 days (45 percent of the inappropriate days identified in the study occurred in stays of 3-5 days) (Longest et al., 1984). One AEP study found that the probability of a day's being inappropriate fell for patients in medical services and rose for patients in surgical services as stays went from short (1-10

Figure 6: Estimated Necessary and Unnecessary Medicare Patient Days (per 1000 Medicare Population by PSRO Area)



Source: Restuccia, Gertman, Dayno, et al., 1984. Used with permission.

days) through medium-length (11-21 days) to long (22 days and longer) (Restuccia, Kreger et al., 1986a). Days in stays of 1-7 days were more likely to be inappropriate than days in longer stays in the SMI study.

Stays in which a higher proportion of the stay was devoted to diagnostic workup and nonemergency admissions were more likely to have been inappropriately admitted (Longest et al., 1984). These relationships were found in both the univariate and multivariate analyses.

Thursday through Saturday admissions were several times more likely to be inappropriate than other admissions in the Urban Hospital study (Restuccia, Payne, et al., 1987a). This pattern is not consistent, however (Restuccia et al., 1984), and probably depends on the particular admitting patterns in the hospital and how closely weekend admissions are monitored.

Diagnostic Category

The Medicare Prospective Payment System has focused attention on differences in length of stay across diagnostic groups of patients. Because of that and because none of the characteristics of patients, hospitals, or physicians discussed above seem obviously useful for targeting utilization review, researchers are beginning to explore whether

some diagnostic categories have higher rates of inappropriate admissions or days than others, indicating that these could be targeted for review. Although the studies conducted to date are preliminary and more research is needed, the results are promising.

Table 8 and Table 9 present information from three studies on rates of inappropriate admissions and days by diagnostic category, indicating a wide variation across diagnostic categories in rates for both admissions and days. Note that comparing the rates across categories is complicated by the studies' use of different systems of grouping diagnoses (Health Data Institute and Payne used Major Diagnostic Categories (MDCs) and the SMI study used groupings of ICD-9 codes for principal diagnoses); inclusion of different types of patients (Medicare patients in the first two studies, all payer groups in the latter); and use of different utilization review instruments (the AEP in the first two studies and the SMI in the latter).

Diseases and disorders of the blood and blood forming organs, myeloproliferative disorders, and digestive disorders have relatively high rates of inappropriate admissions in each of the studies in which they are included. Diseases and disorders of the eye, infectious and parasitic diseases, pregnancy, and trauma are consistently low in inappropriate admissions.

For days of care, diseases and disorders of the musculoskeletal and connective tissue, and trauma have relatively high rates in the studies in which they are included. Diseases and disorders of the eye, pregnancy, myeloproliferative disorders, infectious and parasitic diseases, and diseases and disorders of the respiratory system are consistently low.

It is important to remember that MDCs, which are extremely general categories, can disguise variations in rates of inappropriateness. More detailed breakdowns are advisable to target utilization review more efficiently. For example, in one study, MDC 5: Circulatory System, is about average among the MDCs studied in inappropriate admissions (with 13.6 percent inappropriate admissions) (Payne, 1986). However, this MDC includes several DRGs that have relatively high rates of inappropriate admissions, for example, DRG 107: coronary bypass with and without catheterization (55.3 percent inappropriate admissions); DRG 110: major reconstructive vascular procedures, age greater than 69 and/or complications or comorbidities ("cc's") (27.7 percent); DRG 113: amputation for circulatory system disorders except upper limb and toe (23.3 percent); and DRG 125: circulatory disorders except acute myocardial infarction with cardiac catheterization (25.0 percent). In the majority of these cases the inappropriateness was due to premature admission (that is, too many days

Table 8: Rates of Inappropriate Admissions and Days of Care by Major Diagnostic Category (MDC) or Major ICDA Class of Principal Diagnosis

<i>Study</i>	<i>Admissions</i>			<i>Days of Care</i>			
	<i>HDI</i>	<i>Payne</i>	<i>SMI</i>	<i>HDI</i>	<i>Payne</i>	<i>SMI</i>	
MDC or Class of Principal Diagnosis*							
8.	Musculoskeletal and connective tissue	11	25.8	9.6	41	71.2	14.6
16.	Blood and blood-forming organs	21	24.8	—	28	40.0	—
17.	Myeloproliferative disease	23	34.8	—	11	35.3	—
6.	Digestive	13	17.3	13.3	20	42.5	9.5
10.	Endocrine, nutritional, and metabolic	10	23.0	—	13	53.1	—
3.	Ear, nose, and throat	7	30.0	—	16	73.3	—
	Mental†	11	—	10.6	32	—	8.9
7.	Hepatobiliary system and pancreas	6	19.6	—	31	44.0	—
9.	Skin, subcutaneous system, and breast	20	20.0	—	21	42.3	—
11.	Kidney and urinary tract	4	18.4	—	20	39.8	—
1.	Nervous system	13	13.7	3.7	21	51.6	4.5
12.	Male reproductive	11	21.7	—	10	53.3	—
4.	Respiratory	9	13.3	4.3	14	36.8	6.9
13.	Female reproductive	11	7.9	—	16	24.0	—
5.	Circulatory	9	13.6	1.1	18	46.2	5.9
2.	Eye	2	3.2	—	1	27.8	—
18.	Infectious and parasitic diseases	0	0.0	—	14	33.3	—
14.	Pregnancy	0	—	2.4	0	—	6.9
	Symptoms†	—	—	9.6	—	—	13.3
	Trauma†	3	—	5.6	27	—	12.2
	Neoplasms†	—	—	4.1	—	—	8.6

Sources: HDI: Health Data Institute, April 1984; Payne: Payne, 1986; SMI: Longest et al., 1984.

*MDCs were used in the HDI and Payne studies. ICDA classes were used in the SMI study. Numbered entries are MDCs.

† Entries not numbered are major classes of ICDA codes of principal diagnoses.

Table 9: Major Diagnostic Categories (MDCs) and Major ICDA Classes of Principal Diagnosis, Ranked in Terms of Inappropriate Admissions

	<i>MDC or Major Class of Principal Diagnosis*</i>	<i>Study†</i>		
		<i>HDI</i>	<i>SMI</i>	<i>Payne</i>
16.	Blood and blood-forming organs	1	—	1
17.	Myeloproliferative disorders	1	—	1
8.	Musculoskeletal and connective tissue	2	1	1
6.	Digestive system	1	1	2
10.	Endocrine, nutritional, and metabolic	2	—	1
3.	Ear, nose, and throat	2	—	1
	Mental‡	2	1	—
7.	Hepatobiliary system and pancreas	2	—	2
9.	Skin, subcutaneous tissue, and breast	1	—	2
11.	Kidney and urinary	2	—	2
1.	Nervous system	1	3	2
12.	Male reproductive	1	—	2
4.	Respiratory	2	2	2
13.	Female reproductive	2	—	3
5.	Circulatory system	3	3	2
2.	Eye	3	—	3
18.	Infectious and parasitic diseases	3	—	3
14.	Pregnancy	3	3	—

Sources: HDI: Health Data Institute, April 1984; Payne: Payne, 1986; SMI: Longest et al., 1984.

*MDCs were used in the HDI and Payne studies. ICDA classes were used in the SMI study. Numbered entries are MDCs.

†“1” indicates the categories in each study with the highest rates of inappropriate admissions; “3” indicates the categories with the lowest rates.

‡Entries not numbered are major classes of ICDA codes of principal diagnoses.

before the procedure) and not to the fact that the patient did not require inpatient-level care.

IV. IMPLICATIONS OF THE FINDINGS

A. POTENTIAL FOR SAVINGS

Large potential savings can be realized from effective utilization management programs. Even using 4.7 percent, the lowest rate of avoidable inappropriate admissions,¹² an estimated 1.8 million inpatient admissions in 1983 could have been eliminated. At an average cost that year of \$2,789 per admission, this represents a maximum potential savings of \$5.1 billion (Longest et al., 1984; U.S. Department of Commerce, 1986). With one exception, the studies reviewed here did not examine the need for surgery or the location of surgeries (e.g., outpatient versus inpatient), so the potential savings are probably even greater. (Estimating the potential savings which could be realized through effective utilization review is highly speculative because some of the patients would need to be treated in nonhospital settings. Also, it must be noted that, in order to realize the optimum savings on a societal level, a drop in inappropriate utilization will need to be followed by a corresponding decrease in the number of hospital beds in order to reduce the fixed costs associated with maintaining those beds.)

B. IMPLICATIONS FOR TARGETING UTILIZATION REVIEW

The next two sections discuss the implications of the findings for policymakers, payers, and health delivery managers as they target utilization review and as they design and evaluate utilization management programs to realize savings. The implications are presented on a global basis, but different actors face different incentives and will want to target their utilization management programs differently.

Who Should Be Targeted?

Based on the studies reviewed here, the prime target for intervention should be the *provider*, since 72-86 percent of the inappropriate days of care, in the studies assessing responsibility, were attributed to the physician or the hospital (Restuccia, Payne, et al., 1987a; Borchardt, 1981; Restuccia et al., 1986b). Since there appears to be considerable variation in utilization rates and, at least in one study, in rates of inappropriate utilization by the individual physician, the information should be categorized by physician.

Interventions to reduce inappropriate utilization at the provider level can include monitoring the appropriateness of the admission and readiness for discharge; establishing guidelines for inpatient admissions and the location of surgical procedures; and providing education and feedback to hospitals or physicians on appropriateness rates.

The second most influential factor in inappropriate days is *environmental* factors, such as the unavailability of skilled nursing home care or home health care. Of the inappropriate days in the Urban and Feedback studies, 12-14 percent, respectively, were attributed to environmental factors.

Patients who are inappropriately hospitalized because they need lower-level care cannot simply be discharged, however. Effective and timely discharge planning can anticipate patients' needs and initiate postdischarge arrangements early in the stay so that an appropriate facility will be available when the medical need for acute care ends (Inui et al., 1981).

If the delayed discharge is due not to inadequate discharge planning but to shortages of lower-level facilities, the ultimate solution must be systemwide action by society to provide enough beds for long-term and rehabilitative care, home health care, or housekeeping services for convalescents.

These studies indicate that the patient and the patient's family are relatively insignificant contributors to inappropriate days, since a maximum of only 4 percent of the inappropriate days were attributed to the patient/family (Restuccia, Payne, et al., 1987a). This suggests that, unless demand-side cost-containment measures (such as coinsurance and deductibles) are unacceptably Draconian, which could jeopardize access to care, they will probably have less impact on overall hospital utilization than supply-side measures directed toward providers (although the former can save the payer money by shifting costs to the employer or beneficiary). This point is directly substantiated by the findings from the RAND Health Insurance Experiment discussed previously (Siu et al., 1986).

What Types of Utilization Should Be Targeted?

The studies reviewed here suggest that utilization management programs and utilization review should be directed toward certain diagnostic categories, days in the last third of the stay, nonemergency admissions, and stays with a high proportion of days devoted to diagnostic workups. Both short and long stays should be reviewed. The implications of these findings for program design are discussed below.

The studies do not indicate that any particular type of hospital, physician, or patient should be targeted, since no clear-cut differences in rates of inappropriate utilization have been found by type of hospital, physician, or patient. Two of the studies suggest that higher admission rates may be related to higher rates of inappropriate admissions or days of care across geographic regions (Longest et al., 1984; Restuccia et al., 1984). One study suggests that longer average length of stay may be associated with higher rates of inappropriate days of care (Longest et al., 1984). (It should be noted that these relationships between rates of utilization and rates of inappropriate utilization were not found in the RAND study of six cities (Siu et al., 1986).) While not conclusive, these findings suggest that payers or purchasers of hospital services which operate across several small areas (for example, counties) or geographic regions may want to determine if the areas with higher admission rates and longer lengths of stay have higher rates of inappropriate use.

Based on the preliminary studies described here, it seems most promising to target utilization review to specific diagnostic groups with higher rates of inappropriate admissions or days of care. Either the diagnostic groups identified above can be targeted, or the reviewing organization can identify the categories with high rates in its own population. Utilization review programs designed to reduce admissions, such as Medicare PRO reviews, HMOs, or corporate payers, will want to target the categories that are high in inappropriate admissions. Programs designed to reduce length of stay (such as hospital utilization review programs) will want to target the categories that have high rates of inappropriate days of care.

Programs should be flexible in terms of the types of utilization to target, since the characteristics associated with inappropriate utilization will probably change over time as the economic incentives for hospitals and physicians change and as utilization management programs become effective in reducing inappropriate utilization in the areas targeted.

C. IMPLICATIONS FOR DESIGNING AND IMPLEMENTING UTILIZATION MANAGEMENT PROGRAMS

Make or Buy Decision

It is clear from the discussion so far that initiating a utilization management program is a relatively complex undertaking and that designing, implementing, and evaluating a program demands time, expertise,

and resources. For this reason, the first decision to make in initiating a utilization management program is whether to have an internal program or to purchase utilization review and/or utilization management services from an outside vendor. The advantages and disadvantages of each option will be briefly summarized here. Gary Horvat provides a detailed discussion of each option and a checklist of questions to ask potential vendors (Horvat, 1986).

An in-house program offers more direct control over operations, especially in contrast to offsite vendor services; the possibility of greater flexibility in modifying the program; the possibility of more direct interactions between the program staff and other parts of the organization; and the savings that accrue from not paying the vendor's profit margin. On the other hand, purchasing utilization management services offers the possibility of a quicker startup time and access to specialists in the field who have experience in designing and implementing utilization management programs. The need to make institutional commitments to utilization review staff is avoided, and the organization retains flexibility in moving from one vendor to another if performance is unsatisfactory. In addition, the opportunity exists to benefit from economies of scale (in development and implementation costs) arising from the vendor's experience with other clients.

Participants

Regardless of the option chosen, all relevant actors should be included in a steering or advisory committee to the in-house program implementers or the vendors. The group should be involved in planning the program and in monitoring its performance and impact on costs, quality of care, and employee, patient, or beneficiary satisfaction.

For example, a corporate program would include corporate executives, legal counsel, benefit managers, union representatives, employee representatives, representatives from local hospitals and, perhaps, physicians. Utilization reviewers, data processors, and data analysts should be included, either from the internal staff or, if the program is provided by a vendor, from the vendor organization.

The next sections outline questions that in-house developers of utilization management programs should consider.

Selecting the System

Selection of the specific methods or instrument to use in identifying inappropriate utilization depends on the overall goals and scope of the program and on the timetable for implementation.

Utilization criteria can be adapted to reflect the higher utilization levels found under fee-for-service payment or the lower levels found in managed care systems and HMOs. Those planning a utilization management program should consciously decide which level of utilization is acceptable to their membership or beneficiaries and the utilization management program's goals (Milstein, Oehm, and Alpert, 1987). For example, the "state-of-the-art" AEP is designed to reflect practice patterns in an HMO setting instead of fee-for-service, which is reflected in the original AEP.

For large-scale utilization review programs designed to detect overutilization, use of diagnosis-independent, explicit criteria methods such as the ISD, AEP, or SMI is easier and less expensive than either the application of diagnosis-specific protocols or the development of criteria from scratch. While the explicit criteria methods are specifically designed to assess the timing and location of services provided, they are not designed to assess the appropriateness of services provided to the patient's medical need. For that, the diagnosis-specific methods discussed (protocols, algorithms, tracers, and other criteria-based methods that evaluate quality of care) could be adapted.

The methods that group patients into diagnostic categories prior to review (length-of-stay profiles; diagnosis-specific methods) face two problems: how to group the patients (by diagnosis? principal or primary diagnosis? presenting signs and symptoms? age, sex, or other patient characteristics?) and how to assess appropriateness. In contrast, the explicit criteria methods avoid the problem of how to group patients into diagnostic categories, although these latter methods are somewhat limited in that they are not as suitable as diagnosis-specific methods for assessing the need for service or quality of care.

Need for Flexibility

Whatever method is chosen, uncertainties in the medical care process, individualities among patients, possible inaccuracies in the medical record, and differences in practice patterns across physicians, hospitals, or geographic areas, make it advisable to be flexible in applying the instruments. For example, if decisions are to be made on the hospitalization or discharge of individual patients based on the review (in contrast to accumulating statistics on a large number of patients to use in monitoring providers), the instruments should be used as a screen to identify cases for further review by physician consultants. Diagnostic-specific instruments should incorporate "compliance zones" specifying acceptable ranges of practice. The numbers of cases sampled

for each physician, diagnostic category, or hospital should be large enough to make the statistics meaningful.

Concurrent or Retrospective Reviews?

To maximize the chance of preventing individual cases of unnecessary use, utilization review should be preadmission or concurrent.

Concurrent review poses certain logistical problems due to difficulties in obtaining the record while the patient is in the hospital. More importantly, it is harder to target review to DRGs or to the last third of the stay using concurrent review, since it is not possible to know either the principal diagnosis (upon which the DRG classification is based) or the length of the stay before the patient has been discharged.

One solution is to assign provisional DRGs based on the admitting diagnosis instead of the principal diagnosis (which is the condition, established following discharge, that is chiefly responsible for occasioning the admission). Lengths of stay could then be projected based on diagnostic category. Alternatively, MDCs, which are more general and less likely to change than the DRG category to which the patient is ultimately assigned, could be used for targeting reviews.

For detecting patterns in inappropriate use among physicians or hospitals, retrospective review is sufficient and preferred, since it is logistically easier to obtain access to the large numbers of medical records needed to detect patterns after discharge than while the patient is in the hospital. When combined with timely, provider-specific feedback, retrospective review has been shown to be effective in reducing inappropriate utilization (Restuccia, 1982; Borchardt, 1981; Studnicki, 1984).

Analytical Methods

The benefits of using multivariate statistical methods over univariate methods in order to understand the factors that contribute to inappropriate utilization have been emphasized here. However, to target interventions, comprehensive understanding is not needed and univariate analysis is probably adequate.

Evaluation

In evaluating utilization review and utilization management programs, it is important to measure changes only in *inappropriate* utilization, not changes in total utilization such as average length of stay or admission rates (Gertman and Restuccia, 1981). There are two reasons for this. First, an effective preadmission program will probably result

in an *increase* in length of stay, since patients who can be treated on an outpatient or ambulatory basis will be treated outside the hospital. Such patients have shorter stays on the average than patients who, because of health risk factors or type of treatment, must be treated as inpatients. Removing the shorter stays of the patients who are treated outside the hospital, and leaving the more seriously ill patients in the inpatient pool, will increase average length of stay. An examination only of changes in overall length of stay, without separating appropriate and inappropriate use, may show falsely that the program is having an effect on utilization contrary to the effect intended. More importantly, it is necessary to be sure that the program is not achieving a reduction in utilization by restricting access to *necessary* care (Siu et al., 1986), and this can be determined only through distinguishing between appropriate and inappropriate use.

The ideal evaluation consists of comparing changes in inappropriate use and/or costs in a group covered by the program (the "treatment group") with changes in a comparable (not covered) control group. Without such a control group, it is impossible to know whether any changes noted in the treatment group were due to the program itself or to some other factor influencing utilization. Meaningful evaluation requires careful planning before the program is implemented to identify a control group and to collect baseline statistics (O'Donnell, 1987).

In the absence of a control group (as when all employees are covered under a utilization management program), it is tempting to use routinely collected, computerized hospital discharge data as a pseudo control group. Such data should be used only to derive information on trends in overall utilization which may be influencing the treatment group (such as a general decrease in length of stay in the general population). It cannot be used as a substitute for information on a true control group, since the treatment group and the general population will differ in age, sex, socioeconomic characteristics, health status, and benefits.

V. ADDITIONAL INFORMATION NEEDED BY POLICYMAKERS AND HEALTH DELIVERY MANAGERS

The findings reviewed here suggest several areas for further study:

- Further comparisons of the reliability and validity of the three most commonly used explicit criteria lists, the ISD, AEP, and SMI, are needed.

- Additional large-scale studies are needed, using valid and reliable instruments, randomly selected cases, multivariate analyses, and a full range of variables to explore further the complex relationships found in the studies discussed here. Special attention should be paid to studying patterns in inappropriate utilization by physicians. Studies conducted after the implementation of prospective payment systems will be needed to determine how utilization review programs should be targeted under those systems.
- The reasons why certain diagnostic categories have extremely high rates of inappropriate utilization can be studied in order to determine what factors contribute to high rates. Possible factors include characteristics of the patients, their clinical conditions (especially differences in severity of illness), practice patterns, the patients' postdischarge needs, or the relative profitability and convenience of the procedure to the provider.
- Formal evaluation of various utilization management programs is advisable, to determine the most effective and efficient models.
- More extensive work is needed on detecting underutilization, which is expected to become a more important issue as new incentives in the payment system produce their impact and utilization management programs take effect.

ACKNOWLEDGMENTS

I am grateful to Stephen Crane of the Health Policy Institute and Joseph Restuccia of the Health Care Research Unit/School of Management, both of Boston University, who generously shared their insights and experience with me as I developed and refined the synthesis. Alice Hersh of the Association for Health Services Research; Katherine Morrison of the Wisconsin Department of Health; and Peter O'Donnell of Affirm, Inc. tested the ideas against their practical experience in the field. I want to thank Regina Dahlgren and Andrea Myers of InterQual, and Christy Moynihan of SysteMetrics, Inc. for reviewing the descriptions of the ISD-A and SMI review systems, respectively. My father, Loyal Payne, generously assisted in editing the manuscript, and Camara Phillips skillfully supervised the typing of the tables. I want to thank the Pew Memorial Trust and the Association for Health Services Research for conceiving the synthesis series and the Trust for its financial support.

NOTES

1. In practice utilization management focuses on the *quality* and *quantity* of the inputs into the hospital product (e.g., the number of laboratory tests provided or the length of stay) and not on the cost of inputs (such as wage rates or supply costs).
2. For the sake of simplicity, the outcomes are measured in dollars, the effect of quality on costs and outcomes is ignored, and the cost of only a single hospitalization (with no attention to possible readmissions resulting from underutilization) is considered.
3. This is discussed in more detail in Donabedian, 1980; and Fuchs, 1986.
4. Discussed in Donabedian, 1980; Scheff, 1963; and Scheff, 1964.
5. I have drawn heavily from works by Avedis Donabedian (1973; 1980; April 1980; and 1982) in this section for examples and for the general framework of the discussion. Much of the material on the history of utilization review is from Wilson and Neuhauser, 1982.
6. In 1983, PSROs were replaced by Peer Review Organizations (PROs), which have a similar responsibility for reviewing the hospitalization paid for by Medicare.
7. In 1985, 29.1 percent and 8.9 percent of hospital charges were covered by Medicare and Medicaid, respectively (Waldo, Levit, and Lazenby, 1986).
8. For example, DRGs, a hospital case-mix system designed to group patients into diagnostic categories with similar lengths of stay and charges, group patients based on the principal diagnosis, which is determined after the patient is discharged. Another case-mix system, Patient Management Categories (PMCs), groups patients based on the admitting diagnosis, which is the basis on which the physician, at least initially, forms the treatment plan for the patient (Young, 1979).
9. Information on the instruments is available from the following sources:
 - ISD. InterQual, Inc., 1165 North Clark Street, Chicago, IL, 60610. (312) 751-2327.
 - AEP. Joseph Restuccia, Health Care Research Unit, Boston University School of Medicine, Suite 1102, 720 Harrison Avenue, Boston, MA 02118. (617) 638-8188.
 - SMI. SysteMetrics, Inc., 104 West Anapamu Street, Santa Barbara, CA 93101. (805) 963-6589.
10. The fact that these tests are introduced in this section, and not in the discussion of implicit methods, is not meant to suggest that the earlier methods should not have been tested for reliability and validity, but to recognize that generally they were not tested as rigorously or comprehensively as the later methods.
11. Responsibilities for inappropriate admissions are not assessed with the AEP. The SMI includes an analysis of the reasons and responsibilities for inappropriate admissions and days, but it is not possible from the reported results of the SMI study to separate days attributable to physicians from those attributable to patients.
12. This rates excludes the 18 percent of the inappropriate admissions detected which were due to a lack of alternative sources of care.

REFERENCES

- Anderson, O. W., and P. B. Sheatsley. *Hospital Use—A Survey of Patient and Physician Decisions*. Research Series 24. Chicago: Center for Health Administration Studies, University of Chicago, 1967.
- Berg, R. L., et al. Assessing the health care needs of the aged. *Health Services Research* 5:36-59, Spring 1970.
- Blumenfeld, W. S. Critique of the Development and Evaluation of the Standardized Medreview Instrument (SMI). Submitted to S. A. Terrell, Health Care Financing Administration, Baltimore, MD, 1985.
- Borchardt, P. J. Nonacute profiles: Evaluation of physicians' nonacute utilization of hospital resources. *Quality Review Bulletin* 21-26, November 1981.
- Brook, R. H., et al. Assessing the quality of medical care using outcome measures: An overview of the method. *Medical Care* 15(9):1-165 Supplement. September 1977.
- Browning, F. E., and L. Crump. Report to the Patient Care Planning Council on a Bed Utilization Study. Rochester, NY: Council on Social Agencies, 1969.
- Burford, R., and R. F. Averill. The relationship between diagnostic information available at admission and discharge for patients in one PSRO setting. *Medical Care* 17(4):369-81, April 1979.
- Butler, J. J., and J. W. Quinlan. Internal audit in the department of medicine of a community hospital: Two years' experience. *Journal of the American Medical Association* 167:567, 1958.
- Carmine, E. G., and R. A. Zeller. *Reliability and Validity Assessment*. Beverly Hills, CA: Sage Publications, 1979.
- Chassin, M. R. *Variations in Hospital Length of Stay: Their Relationship to Health Outcomes*. Washington, D.C.: Office of Technology Assessment, 1983.
- Codman, E. A. *A Study in Hospital Efficiency*. Boston: Thomas Todd Co, undated, c. 1916.
- Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37, 1960.
- Cook, T. D., and D. T. Campbell. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Company, 1979.
- Demlo, L. K., P. M. Campbell, and S. S. Brown. Reliability of information abstracted from patients' medical records. *Medical Care* 16(12):995-1005, December 1978.
- Demlo, L. K. Assuring quality of health care: An overview. *Evaluation and the Health Professions* 6(2):161-97, June 1983.
- Donabedian, A. *Aspects of Medical Care Administration: Specifying Requirements for Health Care*. Cambridge, MA: Harvard University Press, 1973.
- Donabedian, A. *Explorations in Quality Assessment and Monitoring: The Definition of Quality and Approaches to Its Assessment (Volume I)*. Ann Arbor, MI: Health Administration Press, 1980(1).
- Donabedian, A. *Explorations in Quality Assessment and Monitoring: The Methods and Findings of Quality Assessment and Monitoring: An Illustrated Analysis*. Ann Arbor, MI: Health Administration Press, 1985(2).
- Donabedian, A. Criteria, norms, and standards of quality: What do they mean? *American Journal of Public Health* 71(4):409-12, April 1980.

- Donabedian, A. *Explorations in Quality Assessment and Monitoring: The Criteria and Standards of Quality (Volume II)*. Ann Arbor, MI: Health Administration Press, 1982.
- Eddy, D. M. Variations in physician practice: The role of uncertainty. *Health Affairs* 3(2):74-89, Summer 1984.
- Ehrlich, J., M. A. Morehead, and R. E. Trussell. *The Quantity, Quality and Costs of Medical and Hospital Care Secured by a Sample of Teamster Families in the New York Area*. New York: Columbia University School of Public Health and Administrative Medicine, 1962.
- Fitzpatrick, T. B., D. C. Riedel, and B. C. Payne. Appropriateness of Admission and Length of Stay. In W. J. McNerney and Study Staff, *Hospital and Medical Economics, Volume I*. Chicago: Hospital Research and Educational Trust, 1962, pp. 471-94.
- Fuchs, V. Has cost containment gone too far? *Milbank Memorial Fund Quarterly* 64(3):479-88, 1986.
- Gertman, P., and B. M. Bucher. Inappropriate Hospital Bed Days and Their Relationship to Length of Stay Parameters. Presented at the Medical Care Section, 99th Annual Meeting of the American Public Health Association, Minneapolis, MN, October 11, 1969.
- Gertman, P. M., et al. Utilization review in the United States: Results from a 1976-1977 national survey of hospitals. *Medical Care* 17(8):1-148 Supplement, August 1979.
- Gertman, P., and J. Restuccia. The appropriateness evaluation protocol: A technique for assessing unnecessary days of hospital care. *Medical Care* 19(8):855-71, August 1981.
- Gertman, P., and J. Restuccia. Methods to Determine Inappropriate Use of Hospital Service: The Appropriateness Evaluation Protocol (AEP). Final Report: Grant No. 18-P-97513/1-02, Boston, MA: Health Care Research Unit, Boston University Medical Center, undated.
- Glass, R. I., et al. The 4 score: An index for predicting a patient's non-medical hospital days. *American Journal of Public Health* 67(8):751-55, 1977.
- Goldberg, G. A., and D. C. Holloway. Emphasizing 'level of care' over 'length of stay' in hospital utilization review. *Medical Care* 13(6):474-85, June 1975.
- Grossman, M. E., and M. Magnus. Benefits subscriber survey. *Personnel Journal* Costa Mesa, CA: *Personnel Journal*, P.O. Box 2440, Costa Mesa, CA 92628, 1986.
- Health Data Institute. Analysis of DRG Patterns and Appropriateness of Utilization in Delmarva. Newton, MA: Health Data Institute, 1984.
- Holloway, D. C., et al. Deciding on the Appropriate Level of Care for Patients. Proceedings, Measures of Quality of Care Forum, National Cooperative Services Center for Hospital Management Engineering, Richmond, VA, 1975, pp. 169-90.
- Hornbrook, M. C. Hospital case mix: Its definition, measurement, and use: The conceptual framework (Part I). *Medical Care Review* 39(1):1-43, Spring 1982. Review of alternative measures (Part 2). *Medical Care Review* 39(2):73-123, Summer 1982.
- Horvat, G. *Evaluating Utilization Review Systems*. Sylvania, OH: United States Health Management Solutions, Inc., P.O. Box 8037, 1986.

- Iezzoni, L. I., and M. A. Moskowitz. Clinical overlap among medical diagnosis-related groups. *Journal of the American Medical Association* 255(7):927-29, February 21, 1986.
- Inui, T. S., et al. Identifying hospital patients who need early discharge planning for special dispositions: A comparison of alternative techniques. *Medical Care* 19(9):922-29, September 1981.
- InterQual, Inc. *The ISD-A Review System with Adult Criteria; The ISD-A Review System with Pediatric Criteria*. Chicago, IL: InterQual, Inc., 1165 North Clark St., Chicago, IL. 60610, 1987.
- Kessner, D. M., C. E. Kalk, and J. Singer. Assessing Health Quality—The Case for Tracers. In N. O. Graham. *Quality Assurance in Hospitals*. Rockville, MD: Aspen Systems Corporation, 1982.
- Lee, R. I., and L. W. Jones (assisted by B. Jones). *The Fundamentals of Good Medical Care*. Chicago: The University of Chicago Press, 1933.
- Lembcke, P. A. Medical auditing by scientific methods: Illustrated by major female pelvic surgery. *Journal of the American Medical Association* 162:646-55, October 13, 1956.
- Longest, B. B., Jr., et al. Measuring Inappropriate Utilization: What Educators, Students, and Administrators Must Know in a Prospective Payment Environment. Paper presented at the Annual Meeting of the Association of University Programs in Health Administration, April 28, 1984.
- Luft, H. How do health maintenance organizations achieve their savings? *The New England Journal of Medicine* 298(24):1336-43, 1978.
- McClure, W. Toward development and application of a qualitative theory of hospital utilization. *Inquiry* 19(2):117-35, Summer 1982.
- McMahon, L. F., Jr., and R. Newbold. Variation in resource use within diagnosis-related groups: The effect of severity of illness and physician practice. *Medical Care* 24(5):388-96, May 1986.
- The move to managed care. *Medical Practice Newsletter*. 6(2):1-4. New Haven, CT: The Practice Letter, Inc., 227 Everit St., New Haven, CT 06530, February 1986.
- Mills, D. H. Medical insurance feasibility study: A technical summary. *Western Journal of Medicine* 128(4):360-65, April 1978.
- Milstein, A., M. Oehm, and G. Alpert. Gauging the performance of utilization review. *Business and Health* 10-12, February 1987.
- Morehead, M. A. The medical audit as an operational tool. *American Journal of Public Health* 57:1643-56, September 1967.
- Morehead, M. A., et al. *A Study of the Quality of Hospital Care Secured by a Sample of Teamster Family Members in New York City*. New York: Columbia University School of Public Health and Administrative Medicine, 1964.
- Morehead, M. A., and R. Donaldson. Quality of clinical management of disease in comprehensive neighborhood health centers. *Medical Care* 12:301-15, April 1974.
- Morehead, M. A. P.S.R.O.—Problems and possibilities. *Man and Medicine* 1:113-23, Winter 1976.
- Nobrega, F. T., et al. Hospital use in a fee-for-service system. *Journal of the American Medical Association* 247(6):806-10, February 12, 1982.
- Noren, J. Establishing Effectiveness and Efficiency of Current Health Care. In J. W. Williamson and Associates. *Teaching Quality Assurance and Cost*

- Containment in Health Care: A Faculty Guide*. San Francisco, CA: Jossey-Bass, Inc., 1982, pp. 136-63.
- O'Donnell, P. S. Managing health costs under a fee-for-service plan. *Business and Health* 38, March 1987.
- Payne, B. C. Research in Quality Assessment and Utilization Review in Hospital and Ambulatory Settings. In P. Y. Ertel and M. G. Aldridge (eds.). *Medical Peer Review: Theory and Practice*. St. Louis: Mosby, 1977.
- Payne, S. M. C. Exploration of the Relationship between Inappropriate Hospital Utilization and Case-Mix-Adjusted Length of Stay. Unpublished Doctoral Dissertation. Boston University, University Professors' Program, 1986.
- Peterson, O. L., et al. An analytical study of North Carolina general practice, 1953-1954. *Journal of Medical Education* 31(2):1-165, December 1956.
- Pettengill, J., and J. Vertrees. Reliability and validity in hospital case-mix measurement. *Health Care Financing Review* 4(2):101-28, December 1982.
- Querido, A. Analysis of Length of Stay. *The Efficiency of Medical Care*. Leiden, N. V.: H. E. Stenfort Kroese, 1963, pp. 127-134.
- Restuccia, J. D. The effect of concurrent feedback in reducing inappropriate hospital utilization. *Medical Care* 20(1):46-62, January 1982.
- Restuccia, J. D., et al. A comparative analysis of appropriateness of hospital use. *Health Affairs* 130-38, Summer 1984.
- Restuccia, J. D., and D. C. Holloway. Barriers to appropriate utilization of an acute facility. *Medical Care* 14(7):559-73, July 1976.
- Restuccia, J. D., B. E. Kreger, et al. The appropriateness of hospital use in Massachusetts. *Health Care Financing Review* 8(1):47-54, Fall 1986a.
- Restuccia, J. D., et al. Assessing the appropriateness of hospital utilization to improve efficiency and competitive position. *Health Care Management Review* 12(3):17-28, Summer 1987b.
- Restuccia, J. D., S. M. C. Payne, et al. Reducing Inappropriate Use of Inpatient Medical/Surgical and Pediatric Services. Report on HCFA Contract No. 18-C-98317/1-02. Boston, MA: Health Care Research Unit, University Hospital, 1986b.
- Restuccia, J. D., and L. Tracey. The Definition and Measurement of Underutilization. Working paper. Boston University School of Management/Health Policy Institute, May 1987b.
- Richardson, F. M. Peer review of medical care. *Medical Care* 10(1):29-39 January-February 1972.
- Rishpon, S., S. Lubacsh, and L. M. Epstein. Reliability of a method of determining the necessity for hospital days in Israel. *Medical Care* 24(3):279-82, March 1986.
- Rosenfeld, L. S. Quality of medical care in hospitals. *American Journal of Public Health* 47:856-65, July 1957.
- Rosenfeld, L. S., F. Goldman, and L. A. Kaprio. Reasons for prolonged hospital stay. *Journal of Chronic Diseases* 14:789-800, December 1957.
- Rosser, R. The reliability and application of clinical judgment in evaluating the use of hospital beds. *Medical Care* 14(1):39-47, January 1976.
- Rutstein, D. D., et al. Measuring the quality of medical care: A clinical method. *New England Journal of Medicine* 294(11):582-88, 1976.
- Sanazaro, P. J. Quality assessment and quality assurance in medical care. *Annual Review of Public Health* 1:37-68, 1980.

- Scheff, T. J. Decision rules, types of error and their consequences in medical diagnosis. *Behavioral Science* 8:97-105, 1963.
- Scheff, T. J. Preferred errors in diagnosis. *Medical Care* 2:166-72, 1964.
- Simborg, D. W. DRG creep: A new hospital-acquired disease. *The New England Journal of Medicine* 304(26):1602, 1981.
- Siu, A. L., et al. Inappropriate use of hospitals in a randomized trial of health plans. *New England Journal of Medicine* 315(20):1259-66, November 13, 1986.
- Solomon, D. H., et al. Indications for Selected Medical and Surgical Procedures—A Literature Review and Ratings of Appropriateness: Cholecystectomy. RAND Corporation, Santa Monica, CA, May 1986.
- Steel, K. Iatrogenic disease on a medical service. *Journal of the American Geriatrics Society* 32(6):445-49, 1984.
- Steel, K., et al. Iatrogenic illness on a general medical service at a university hospital. *New England Journal of Medicine* 304:638-42, March 12, 1981.
- Strumwasser, I. S., N. V. Paranjpe, and the staff of the Michigan Health Care Education and Research Foundation, Inc. Executive Summary: Estimates of Non-Acute Hospitalization: A Comparative Analysis of the Appropriateness Evaluation Protocol and Standardized Medreview Instrument. *Final Report*. Health Care Financing Administration Grant No. 18-C-98582/5-01 and 2. MHCER Foundation, Inc., Detroit, MI, 1987.
- Studnicki, J., and C. E. Stevens. The impact of a cybernetic control system on inappropriate admissions. *Quality Review Bulletin* 304-11, October 1984.
- SysteMetrics, Inc. *The Valid and Reliable Measurement of Nonacute Hospital Utilization in a Nationally Representative Sample: Final Report*. Deliverable No. 6. Santa Barbara, CA: SysteMetrics, Inc., 1984.
- Terrell, S. A., C. Moynihan, and C. Barnes. The Measurement of Nonacute Hospital Utilization in a National Sample: The Admission Model. Presented at American Public Health Association Meeting, Dallas, TX, 1983.
- U.S. Department of Commerce, Bureau of the Census. *Statistical Abstract of the United States, 1986*. 106th Edition. Washington, DC: Bureau of the Census, 1986.
- Van Dyke, F., V. Brown, and A. Thom. *Long Stay Hospital Care*. New York: Columbia University School of Public Health and Administrative Medicine, 1963.
- Waldo, D., K. Levit, and H. Lazenby. National health expenditures, 1985. *Health Care Financing Review* 8(1):1-22, Fall 1986.
- Wenkert, W., and M. Terris. Methods and findings in a local chronic illness study. *American Journal of Public Health* 50:1288-97, September 1960.
- Wennberg, J. Dealing with medical practice variations: A proposal for action. *Health Affairs* 6-32, Summer 1984.
- Wennberg, J., and A. Gittelsohn. Variations in medical care among small areas. *Scientific American* 246(4):120-34, April 1982.
- Wilson, F. A., and D. Neuhauser. *Health Services in the United States*. Cambridge, MA: Ballinger Publishing Company, 1982.
- Wilson, P., and P. Tedeschi. Community correlates of hospital use. *Health Services Research* 19(3):333-55, August 1984.

- Young, W. W. Measuring the Cost of Care Using Generalized Patient Management Paths. Progress Report for Grant No. 18-P-97063/3-01, Submitted to Health Care Financing Administration, November 1979.
- Zimmer, J. G. An evaluation of observer variability in a hospital bed utilization study. *Medical Care* 5:221-33, July-August 1967.
- Zimmer, J. Length of stay and hospital bed misutilization. *Medical Care* 12(5):453-62, May 1974.