Chapter 2

A Brief Review and Evaluation of Alternative Approaches to Case - Mix Measurement

Chapter 2

A Brief Review and Evaluation of Alternative Approaches to Case-Mix Measurement*

INTRODUCTION

Policymakers, hospital administrators, and health services researchers have long recognized the diversity of hospitals' outputs. Efforts to analyze hospital behavior and to establish effective and equitable reimbursement systems have been complicated by this diversity of hospital products, which include education and research as well as patient care.

For two decades, researchers have grappled with the measurement of hospitals' patient care output. While enormous progress has been made during that period, there remains no consensus about the most appropriate method of case-mix measurement. Failure to reach such a consensus is not surprising. Several substantial barriers have stood in the way, including the variety of purposes case-mix measures have been designed to serve and the significant data requirements associated with any but the most rudimentary measures. The variation in the purposes to be served has been a barrier, because it appears that the "optimal" measure may be different if it is to be used for reimbursement, quality assurance, management, or for some other purpose. Data requirements stem from the need, irrespective of the specific measure, to obtain detailed information about the patients for which the case-mix measure is to be defined. Even in the smallest hospitals, admissions total about 2,000 annually; in the largest institutions, more than 40,000 patients are admitted each year. Even in this age of computers, the national number of admissions—3l million*—is formidable.

This chapter presents a brief overview of the development of case-mix measures, from the early rudimentary techniques to the most recent advances. It is intended to provide a frame of reference against which Diagnosis Related Groups (DRGs) can be assessed. It is not intended to provide a detailed review of past or current approaches to case-mix measurement. Such detail can be found in the references cited throughout the chapter or, alternatively, in an excellent review article by Hornbrook (38).

"Total admissions to non-Federal short-term general hospitals in 1980.

EARLY APPROACHES

The earliest measures of patient care output were developed at the institutional level. That is, the initial measures represented one or more indices, developed for the hospital as a whole, that were designed to reflect a dimension of hospital performance assumed to be associated with case mix. These measures included average length of stay (LOS), surgery rates, relative volume of outpatient visits, number of births, and other similar indicators. Data to construct such measures were readily available from published sources, such as the American Hospital Association's annual survey of hospitals.

It was recognized that simple summary indices such as these were no better than crude surrogates for case mix. However, including such measures in an analysis of interhospital cost variation, for example, seemed preferable to excluding any output characteristics, and for a number of years

● This chapter is based on a paper prepared for OTA by Nancy L. Kelly, Diane E. Hamilton, and Ralph E. Berry of Policy Analysis, Inc.

there were no alternatives. But the empirical evidence indicated that these measures did not explain very much of the interhospital variation in costs (see, e.g., 44). Clearly, the low explanatory power resulted in part from the unidimensionality of the measures. The ratio of surgical operations to admissions, for example, fails to distinguish hospitals performing a great deal of complicated, high-risk surgery from those performing equal numbers of simpler, more common procedures. Similarly, a long LOS may be experienced by patients with acute and severe illnesses or by patients with chronic conditions awaiting discharge to a nursing home.

Often, additional surrogate measures were used along with those described above. The added measures described characteristics of the hospital, rather than the patient population, but they were considered to be highly correlated with case mix. The earliest surrogate was the absolute size of the hospital, measured in terms of numbers of beds or admissions. Case mix was assumed to be more complex in larger hospitals than in smaller facilities. Teaching status is another commonly used surrogate; it is widely believed that teaching hospitals treat a more severely ill mix of patients than nonteaching hospitals. Similarly, physician staff characteristics have occasionally been identified as useful surrogates. Measures of the scope of facilities and services have also been used as indicators of the case mix of patients treated, under the assumption that hospitals are equipped to treat a particular array of illnesses. While size, teaching status, physician staff composition, and scope of services undoubtedly have some predictive power with respect to case mix, the evidence suggests that this power is far from perfect (see, e.g., 4s).

A later development among the early approaches involved the linkage of demographic and economic characteristics of the hospital's "market area" (from sources such as the U.S. Bureau of Census) to the hospital-specific information. This was based on the further assumption that unmeasured dimensions of case mix, not contained in the institutional measures, could be obtained from the characteristics of the area in which the hospital was located. Examples of the characteristics thought to be important were the age distribution of the population (especially the proportion over age 65), median income and education levels, and the numbers of Medicaid recipients. Even such indirect measures as urban versus rural location and/or population density were considered. Associated with each of these characteristics was an underlying hypothesis about its effect on the case mix of area hospitals. For example, a high proportion of poor and/or elderly people in the surrounding area was thought to be more closely associated with severe illnesses than a high proportion of well-educated, moderate income households (70). Broad descriptors of the area, such as population density, have often been considered as surrogates for differences in lifestyle, which in turn, lead to differences in health status.

While it can be argued that they add important information to the limited hospital-specific indices, the market area characteristics also must be viewed as particularly crude surrogates for case mix. The chief drawback of these measures lies in their imprecision. Given current data sources, it is not possible to accurately identify, on a national scale, the precise market area from which the hospital draws its patients. This is especially difficult in urban areas, where many hospitals are clustered in a small geographic area, and where some of the hospitals are referral centers for a much larger region. Except for the relatively few States in which patient origin data are collected, counties or Standard Metropolitan Statistical Areas are typically the only geographic units for which data can be obtained. At best, these will be rough approximations of the true market areas for most of the Nation's hospitals.

It should be noted that the measures that have been labeled "early" are, in fact, still widely in use in research on hospital costs and behavior today. Although, as mentioned previously, considerable advances have been made in the area of case-mix measurement, none of the more recent developments is yet widely available, nor are the requisite input data. As the next section describes, the data requirements of the principal alternatives have placed new demands on traditional recordkeeping and data collection procedures. Gradually, however, alternative measures are likely to become more widely available as existing systems respond to these demands.

RECENT DEVELOPMENTS: DESCRIPTION AND EVALUATION OF NEW APPROACHES

Dimensions of Case Mix

Recognition that case-mix measurement was an important problem—and probably a necessary tool for developing solutions to such longstanding dilemmas as hospital cost inflation-aroused considerable interest as well as funding for a number of research and development efforts. As a result, several important advances have been made. Before they are described, however, a conceptual framework for viewing the problem of case-mix measurement will be presented as background.

Perhaps the first "advance" that motivated the new developments was the recognition that the patient, not the institution, was the appropriate unit of analysis. No matter what the hospital was equipped to do, where it was located, or who served on its physician staff, the types of patients it treated during the course of a given year were the true determinants of the hospital's patient care output. The "condition" of the patients in that population was believed to dictate treatment patterns and, consequently, resource use within the hospital. Further, it is useful to consider "condition" as having two key dimensions: the ***nature of theproldem*** underlying hospitalization (usually indicated by the diagnosis or diagnoses), and the ***relative severity*** of that problem.

A major difficulty facing evaluators of case-mix measures lies in choosing the appropriate frame of reference for such an evaluation. For instance, some measures, such as DRGs, were developed specifically (though not exclusively) to account for differences in resource use. Other approaches do not explicitly address resource use, although they may in fact be highly correlated with this indicator. An evaluation of how well a given measure explains variation in resource use, therefore, should in fairness recognize the purpose that the measure was designed to serve (though this may not alter the conclusions reached).

Related to this issue are the different perspectives of potential users of case-mix measures and their implications for the validity and acceptability of a given measure. For example, measures that seem intuitively reasonable, exhibit high explanatory power in statistical analysis, and can be constructed from readily available data sources may be perfectly acceptable to officials of reimbursement agencies, but they may have no meaning or legitimacy for clinicians. Conversely, measures that are acceptable to clinicians may be infeasible to employ in a large-scale, national program. It seems unlikely that a single measure will satisfy the requirements of all potential users, though some approaches will come closer than others. The likelihood that different conclusions will be reached by different groups of users is not necessarily a problem, however, since alternate approaches may productively be used in tandem.

Against this background, two broad categories of approaches will be described. The first consists of institutional measures, which (to distinguish them from the early, traditional approaches) are derived from data describing the diagnostic composition of the hospital's patients. Included in this category are the "ad hoc" grouping methods (45), the Resource Need Index (100), and information theory measures (17,33). The second category consists of patient-level measures. Of these, the most well known is DRGs. The other approaches reviewed in this chapter include two major efforts to measure severity of illness within a disease entity: Disease Staging **(24)** and the Severity of Illness Index (30,31). Finally, the discussion includes a new approach to defining patient categories and assessing treatment patterns, known as Patient Management Categories ***(PMCs) (107).*** These last three approaches should be viewed as potential alternatives to DRGs, although the Severity of Illness Index can also be viewed as a complement to them.

Hospital-Level Measures

The institutional-level measures referred to above all represent aggregations of patient data designed to capture the overall case-mix complexity of, and the resource implications for, a particular hospital. Of these, the most rudimentary can be termed the "ad hoc" grouping techniques.

These involve the collection of diagnostic information for patients from a sample of hospitals and aggregating those data in ways that appear to be analytically meaningful. For example, Lave and Lave (45) used this approach to define a set of characteristics for each hospital they studied. In their study, a set consisted of proportions of patients in broad diagnostic categories, such as cardiovascular diseases. The categories selected were assumed to distinguish groups of patients that have different resource requirements, although no data were available to directly evaluate that hypothesis.

A measure that takes resource use directly into account is the Resource Need Index (RNI), developed by the Commission on Professional and Hospital Activities (100). The construction of RNI is a two-step process. The first step involves the development of "relative need units" for each diagnosis (or diagnostic category). The units, which are based entirely on charge information, represent the ratio of average charges for a particular diagnosis to overall average charges. RNI is then constructed as the average number of relative need units for a given hospital. RNI thus makes a start at the simultaneous evaluation of diagnosis and resource use, though a possible drawback lies in its reliance on charges as the sole indicators of resource requirements. To the extent that differences between charges and costs differ by type of service, hospitals' charges may not reflect actual resource use very precisely.

Finally, information theory measures have also been derived for the institution as a whole. Studies by Evans and Walker (17) and Horn and Schumacher (33) have employed this approach to case-mix measurement. Information theory measures are based on the assumptions that rare conditions are complex and tend to be treated by a few specialized hospitals. In contrast, common conditions are assumed to be less complex and likely to be uniformly distributed across hospitals. The measure is derived mathematically from the proportions of cases in each diagnostic category. The highest scores for this measure will therefore go to the "most complex" hospitals—i.e., those that treat the uncommon diseases. The specific resource requirements of any of the diseases, including the rare ones, are not explicitly considered in the information theory approach. Instead, its validity in measuring case-mix complexity rests on the validity of underlying assumptions.

The only common threads in these divergent techniques are their use of diagnostic data and their assumptions about the appropriateness of aggregating the information to the hospital level. While this aggregation is undoubtedly convenient, it involves considerable simplification and, as a consequence, loss of information. Some, such as Klastorin and Watts (44), have considered the issue of hospital-level case-mix indices and have concluded that summary indices may not be comparable across hospitals. However, the thrust of most recent research in the area of case-mix measurement has been toward the development of patient-level measures. Though it is possible to aggregate these measures into a hospitalwide index (and indeed several such indices have already been constructed), such aggregation has not been a principal focus of the development process.

Patient-Level Measures

Diagnosis Related Groups

DRGs are undoubtedly the most well known of the patient classification systems that have been introduced during recent years. As this section indicates, however, several major alternatives now exist that differ from DRGs both conceptually and in practice. All of the prominent systems are described briefly in this chapter. A more complete description of DRG development methods can be found in appendix B.

The development of DRGs has been ongoing since the late 1960's, and it is appropriate to view the concept as one that is continuously evolving. The evolution of DRGs has involved both conceptual refinements and technical improvements, spurred by the availability of more and better quality input data and by feedback from a wide variety of observers and users of DRGs. It is likely that the evolution will continue as relevant data increase in availability and improve in quality and as the concept is subjected to more and more scrutiny.

The first version of DRGs to be widely disseminated was a set of 383 categories, described by

their developers in 1980 (19). Subsequently, in early 1982, a second and much revised set of 467 categories was issued (103). This revised set bore little resemblance to the "original" 383, as it was based on different definitional procedures and a different coding convention. Both sets had several common objectives. Both were designed to identify patients with similar expected resource use, measured by length of hospital stay. Both versions were defined so as to be medically meaningful to physicians, the key decisionmakers within the hospital with respect to patient care, though the operationalization of this objective varied significantly between the two. Finally, both sets of DRGs were deliberately based on data that are commonly available, and both sets sought to be limited to "manageable" numbers of groups.

In general, the broad outlines for the construction of both sets of DRGs were the same for each version. Actual patient stays in a sample of hospitals were the units of analysis. Each patient's principal diagnosis—i.e., the principal reason (after study) for his/her hospitalization—was coded using a detailed coding system that allowed for many thousands of possibilities. The first step, therefore, was to collapse the detailed diagnosis codes into meaningful, but broad, subcategories called "Major Diagnostic Categories (MDCS). " These MDCs were then further subdivided, using a combination of statistical analysis and medical judgment, according to other characteristics that accounted for differences in resource use within each MDC. (The remainder of this discussion will focus on the current set; a comparison of the two can be found in app. B.)

The major differences, however, appear to outweigh the similarities. Significantly modified procedures were used to develop the **467** DRGs. These differences included the involvement of a far greater number of participants, many of them clinicians, which accompanied a shift in the fundamental orientation of the development process. Whereas the development of the 383 DRGs had involved both statistical analysis and expert clinical judgment, the balance between the two components was relatively more even than it became in the revised method, in which the balance was shifted in favor of clinical judgment.

The current set of **467** DRGs was derived from 23 MDCs, most of which were defined around organ systems of the body (in conformance with the organization of medical practice). Subdividing MDCs into DRGs was performed by expert panels comprising physicians and others in the health industry. Their decisions were guided by several criteria that had been established. For example, the guidelines required that the initial partition (when possible) be based on the presence or absence of a surgical procedure performed in an operating room. Panels were also instructed to rank order surgical procedures according to resource intensity and to assign patients with multiple procedures to the procedure involving the greatest intensity. In addition to considering the type of surgery performed, the nature of coexisting conditions and complications were explicitly evaluated. "Substantial" conditions and complications were distinguished from those less significant. Surgery, coexisting conditions, and complications were all viewed as indicators of severity of illness. Finally, other variables were taken into account when the experts determined that they were relevant. Examples of other factors used for subdivision include death and "left hospital against medical advice. "

Though clinical judgments dominated the decisionmaking process, statistical analysis was used to aid those judgments. Patient-level data were made available by several organizations, principally the Professional Activity Study of the Commission for Professional and Hospital Activities. These data were viewed to be representative of national treatment patterns. Reduction in variance for LOS was examined for possible partitioning variables, but the fact that variance may have been significantly reduced by a particular variable did not guarantee that that variable would be included in the DRG definition. The need to group patients with clinically related diseases, above all, dictated which measures would be used.

The outcome of this process was a mutually exclusive and exhaustive set of 467 DRGs. * By de-

***In operation, there are 468 DRGs, the last for patients who have received an operating room procedure unrelated to their MDC.**

98-823 0 - 83 - 4

sign, DRGs can be determined from discharge abstract data, which are commonly available in computerized form. A computer algorithm is available to classify each patient into the appropriate DRG. As a consequence, use of this classification system poses few administrative burdens. As the following chapter will describe, little empirical research has been conducted to date on the current set of DRGs. Some recent evidence indicates that there still remains substantial within-DRG variation in resource use. For example, one analysis of a random sample of cases in 12 high-volume DRGs applicable to older patients in New Jersey hospitals found that 13.6 percent of discharges had a LOS more than two times the average of cases in that category (75). Analysis of Medicare discharges for 1979 also showed a wide range of within-DRG variation around the mean cost per discharge. For both the old 383 DRG and the new 467 DRG classification systems, it was found that DRGs had coefficients of variation ranging from about 0.2 to 1.5 in the Medicare population (93). A coefficient of variation of 0.2 can be interpreted as indicating that roughly two-thirds of patients in the DRG have costs within 20 percent above or below the mean cost of the DRG. A coefficient of 1.5 indicates that about two-thirds of patients in the DRG lie within Iso percent of the mean cost. The new DRGs do not appear to increase the homogeneity of the groups with respect to their actual consumption of resources. Finally, the extent to which DRGs are acceptable to clinicians is unclear.

New Jersey and several other States have used DRGs in hospital payment systems with varyingly degrees of success. Appendix C contains brief descriptions of several State per-case payment systems, some of which used DRGs.

Disease Staging

Apart from DRGs, the most prominent patient-level measure in the literature is Disease Staging. Both Disease Staging and the Severity of Illness Index (to be described subsequently) were developed to address the perceived need to measure the severity of the patient's illness as well as his or her diagnosis. Severity has been defined as the risk of death or permanent impairment resulting from the illness (38).

Staging consists of the specification of progressive levels of severity for disease in terms of the events and pathophysiological observations that characterize each stage (24). As described in Hornbrook (38), staging involves the segmentation of a disease entity into five primary stages, which are defined as follows:

- Stage O: No disease present, or diagnosis unknown.
- Stage 1: Disease is certain and no complications are present either local or systemic.
- Stage 2: Disease process is limited to an organ or system; significantly increased risk of complications.
- Stage 3: Significantly greater problem than stage 2: multiple site involvement; generalized systemic involvement; poor prognosis.
- Stage 4: Death or most severe stage possible (i.e., the final event of the illness).

Six, rather than five, primary stages are used for cancers, to maintain consistency with previous work. (In fact, staging was first used in oncology in clinical trials of new treatments to incorporate illness severity into experimental design and evaluation (38).) Substages have also been defined for the cancers.

Stage assignments can be made by a computer algorithm, based on data recorded on discharge abstracts. Computer-assigned stages, however, may represent underestimates of the stages that technicians would derive manually from medical records. Again, according to Hornbrook (38), underestimates may occur for two reasons: 1) the primary diagnostic coding systems used for discharge abstracts are not sufficiently precise, and 2) insufficient data are included on the discharge abstracts.

Staging is the product of physicians' judgments about the ***biological progression*** of a given disease. First and foremost, it was developed to be a clinically meaningful concept. The extent to which costs, charges, or LOS varied within stages was not considered during the development process. Although some limited evaluation has indicated that the stages of a disease are indeed systematically related to differences in those other measures (6,23), the relationship between the stages of a disease and resource consumption has

not yet been investigated in depth. Further, it has not been demonstrated that the individual stages are homogeneous with respect to resource use (23). In addition, the stages are not necessarily comparable across diseases, as each disease entity is staged separately. Consequently, stage 2 of a surgical condition may be much more serious than stage 2 of a medical condition and thus require more resources during treatment.

The primary advantages of Disease Staging lie in its apparent meaningfulness to clinicians, as well as in the direct way in which the stages capture the biological severity of illness within a given diagnosis. Staging has the added advantage of requiring only information commonly available on computerized discharge abstract data (although some precision is sacrificed when computerized methods are used). A significant drawback lies in the likelihood that certain diseases cannot be staged. (In general, medical conditions are more difficult to stage than surgical.) The fact that resource consumption was not an explicit consideration in developing the stages (and as a result may or may not be captured by them) may be a serious drawback if stages were used in a reimbursement context. Moreover, since stages are based on single diseases and on the prognosis for each patient, they ignore concurrent conditions and other patient characteristics which affect resource use, such as social, economic, and psychological factors **(38).**

Severity of Illness Index

Still more recent than staging is a measure that was developed to reflect the overall severity of illness of the patient, and not just the severity of the principal diagnosis. The Severity of Illness Index, developed at Johns Hopkins University (31), classifies patients into four severity levels. Unlike the staging procedure, this Severity of Illness Index is not disease-specific but instead was designed to apply to all conditions treated in the medicine, surgery, obstetrics/gynecology, and pediatrics departments of hospitals. Developed in conjunction with a panel of physicians and nurses, the index is built from seven criteria deemed to be the best indicators of overall illness severity. These include:

- stage of the principal diagnosis;
- complications of the principal condition;
- concurrent, interacting conditions that affect the course of hospital treatment;
- dependency on the hospital staff;
- extent of nonoperating room procedures;
- rate of response to therapy, or rate of recovery; and
- impairment remaining after therapy for the acute aspect of the hospitalization.

Data relevant to each of the above criteria are obtained from the patient's medical record. Abstraction of data is performed manually, by a trained rater. Based on the combined pattern of severity levels within each of the seven criteria, the rater makes a judgment about the overall severity of the patient's condition. The overall index can range from 1 (least severe) to 4 (most severe).

The Severity of Illness Index may be used as an adjunct to other patient classification systems, such as DRGs. In that context, refined categories of severity can be developed within categories of patients. Current research, however, suggests that a preferred use of the Severity of Illness Index would be in conjunction with a very broad classification system, such as the 23 MDCsdescribed earlier (30).

The major advantage of the Severity of Illness Index, particularly for payment purposes, would appear to be the extent to which it explains variation in resource use. In a comparative analysis involving six disease conditions, Horn and colleagues (36) found that the Severity of Illness Index produced groups that were more homogeneous than DRGs, Disease Staging, or PMCS (to be discussed subsequently). Homogeneity was assessed with respect to total charges, LOS, laboratory charges, routine charges, and often radiology charges. The index has also been shown to be a good predictor of resource use (35). This explanatory power may, in part, result from the method used to ascribe a severity level to a given patient. Although the seven criteria do not explicitly address resource use, some of the criteria (e.g., extent of nonoperating room procedures) are clearly correlated with it, and there may be

a tendency on the part of raters to take it into account when forming a judgment about overall severity (38).

A drawback of the Severity of Illness Index is its reliance on manual abstraction of data from the medical record and on the judgments of individual raters. Although recordkeeping systems could be modified in such a way that the data necessary to construct the Severity of Illness Index would be available on computerized discharge abstracts, in general, discharge abstract systems are currently inadequate for this purpose. Thus, assignment of severity levels by this method is relatively costly. Also, so long as subjective judgments are employed in assigning the index values, there are likely to be problems with the reliability and acceptability of the measure.

The acceptability of the Severity of Illness of Index to clinicians is currently unclear. Although the development of the index was in conjunction with physicians and nurses, there is as yet no indication of how meaningful the index is to clinicians around the country. At present, the chief advantage of this approach seems to be its success in accounting statistically for variation in resource use.

Patient Management Categories

A criticism that has been leveled at all case-mix measures based on discharge diagnosis is that the diagnosis at discharge is not the only relevant diagnostic information **(106).** Instead, it is argued that diagnosis at the time of admission determines the course of treatment that the physician will employ. In other words, not only the diagnosis of the patient, but also the reason for admission, will affect the ultimate LOS and total costs/charges. Reasons underlying admission can range from observation to chemotherapy to major surgery, all of which have vastly different implications for resource use. Young and colleagues (106) also have argued for the development of a measurement system that avoids building in ***actual*** treatment patterns, regardless of their appropriateness. They favor a method that is more normative—i.e., one that views patient characteristics and management without regard to current treatment patterns that may result from discretionary decisions, differences in the availability of particular facilities and services, inefficiencies, etc.

As a consequence, Young and colleagues (105) at Blue Cross of Western Pennsylvania have developed an alternative patient classification system, whereby patients are classified into PMCs * PMCsdiffer from most of the other systems, including DRGs, in that they are based primarily on the patient's clinical characteristics. The definitions of PMCsdo not hinge on how the patient was treated while in the hospital. PMCsdiffer in other ways as well, as will be discussed subsequently.

Like the other case-mix measures (all to varying degrees), PMCshave been developed in consultation with physicians so that they represent clinically meaningful entities. Both final diagnosis ***and*** reason for admission are considered simultaneously in defining the categories. PMCs have been designed to take levels of severity into account, again from a clinical perspective. For each PMC, physicians have specified components of care (i.e., diagnostic services, treatment procedures, and expected LOS) that, in their view, are required for the effective management of the typical patient. Thus, a "patient management path" has been associated with each PMC. These components of care then form the basis for the derivation of relative cost weights for each PMC. Weights are based on actual cost data from six participating hospitals that have been adjusted and allocated to the components of care. The identification of patient management paths and relative costs during the development process is another distinguishing feature of the PMC approach.

Currently, PMCs are still being defined, although it is anticipated that the process will be completed by the end of the summer of 1983. Computer software is also being developed that will enable the automated mapping of patients into PMCs. Although discharge abstract data typically do not include information on reason for admission, Hornbrook (38) reports that preliminary research indicates that valid mapping into

● **Much of the substance of this discussion was drawn from unpublished documents provided by Wanda W. Young (ICM,IOS).**

PMCS may be possible without collecting additional data.

The principal advantage of PMCs over either DRGs or staging would appear to derive from the joint consideration, during the development process, of payment ***and*** patient management. While the system was designed for use in a payment context, actual patterns of use (as has been noted) were not directly considered in defining the categories. However, the relative cost of each PMC ***is*** calculated as part of the development process. This can presumably be incorporated into a payment procedure. What is most unique to this system, however, is its recognition that patient management should be the focus of any system that seeks to encourage efficiency and the deliberate attempt on the part of the developers to produce a system that would simultaneously be meaningful to physicians and facilitate efficiency improvements in the management of patient care.

Because the system is not yet completed, it is premature to make comparisons between it and other alternatives. For the same reason, no empirical evaluations have yet been performed. Clearly, however, PMCs represent an interesting innovation in the area of case-mix measurement that has considerable potential.

SUMMARY

Until recent years, case mix has been measured by hospital-level surrogates for patient care output. These measures have been derived from readily available sources of data and generally represent crude volume and performance measures along with relevant characteristics of the hospital (e.g., teaching status) and location.

The early measures have been useful in explaining some of the observed interhospital variations, but it is apparent that these measures do not contain the amount of information necessary to accurately capture interhospital case mix differences. As a result, considerable effort has been devoted in the past decade to developing new and improved measures of patient care output. The most well known of these advances, DRGs, are the subject of this technical memorandum. Other advances, including both substitutes for and complements to DRGs, have also been reviewed briefly in this chapter.

The major advances in measuring health care output have been in the area of severity of illness measurement. Disease Staging and the Severity of Illness Index were both designed to provide a framework for classifying diseases according to the relative severity of the patient's condition. Both have required extensive developmental work and testing, which are still underway. Use of either measure would require more data than are generally available at the present time, though the staging approach can be employed using data that are normally included in hospitals' computerized records. Neither measure has reached the point where it is suitable for widespread implementation in a reimbursement context. However, the existence of such measures, at a minimum, serves as a reminder that the relative severity of patients' illnesses is important to consider when measuring case mix.

Patient Management Categories represent the newest of the major advances in case-mix measurement. It will still be some time before the system is fully operational and adequate testing can be performed. In the short term, however, this method of patient classification again represents a reminder that currently used methods may not be appropriately targeted.

This review of alternative methods of measuring hospital case mix has revealed that early methods are lacking in precision and that the new approaches (with the exception of DRGs) are not yet ready for widespread use. CIearly, any "refined" system that tackles patient-level case-mix information will require considerably more data than has been employed (or even available) in the past. Feasibility considerations, therefore, should include the relative administrative burdens associated with each measure as well as the stage

of development each measure has reached. Most importantly, the conceptual differences among the alternatives should be evaluated (as well as the statistical evidence) in order to assess the appropriateness of each for the purposes it is to serve.