Measurement of Severity of Illness and the Medicare Prospective Payment System:

State of the Art and Future Directions

LAURENCE F. McMAHON, JR., MD, MPH, JOHN E. BILLI, MD

THE ADVENT of the Diagnosis Related Group (DRG)-based Medicare Prospective Payment System (PPS) has occasioned concern about the equity of the system across the nation's hospitals.[1-3] The DRG prospective payment system is predicated on the assumption that an appropriate case-mix system will largely account for differences in the costs of caring for patients at the hospital level. As Secretary Schweiker of Health and Human Services noted, "Since patients have different diagnoses, require different treatments, are of different ages, and differ in other ways, it is important to develop a payment system that explicitly adjusts for these differences. Prospective payment systems which do not recognize differences in case-mix will severely harm the tertiary care hospitals which treat more complex illnesses, as well as rural hospitals, which have a volatile case-mix. The lack of a case-mix adjuster would also make the severely ill patient a financial liability to all hospitals and encourage some hospitals to admit only less severely ill patients."[4] As Schweiker mentioned, an equitable case-based prospective payment system depends on its ability to differentiate among clinically distinct patient types.

Some are concerned that a significant amount of the variation in the DRG-specific resource profiles (costs) is a function of differences in severity of patients' illnesses. Further, they argue that a severity of illness measure is necessary explicitly to account for the intra-DRG variation in resource use *between* hospitals.[5] Discussions of severity of illness as it relates to the DRG-based PPS have evolved into a rather one-dimensional argument, namely 1) that a significant amount of intra-DRG resource variability is secondary to differences in patients' severity of illness, 2) that the differences in clinical severity of illness are directly related to resource consumption, 3) that severity of illness profiles of hospitals differ significantly between hospitals, and 4) that a severity of illness system is necessary if PPS is to prove equitable. Empiric evidence to support any of the above contentions is either tenuous or nonexistent. Yet, without convincing evidence to the contrary, it is clear that these arguments will continue to be made.

This article explores the various reasons for intra-DRG variation in resource use, of which severity of illness differences are one component. We then provide a theoretical framework for evaluation of any severity of illness system, and examine the empiric evidence that supports the existing severity of illness systems. Finally, we propose future directions for the development of severity of illness measurement systems. An explicit assumption in our discussion is that the DRG-based PPS, although it may be modified, is unlikely to be replaced in the next ten years.

VARIATION IN RESOURCE USE WITHIN DIAGNOSIS-RELATED GROUPS— THE SEVERITY ISSUE

Because it is not possible to group patients so that every member of a group has the same resource use, variation in hospital resource use within a DRG is *expected*. Variation in resource use becomes a concern only 1) if it is predictable, i.e., associated with identifiable patient attributes such as a low hematocrit in a patient with a bleeding ulcer, or 2) if high-severity/high-cost patients are concentrated in certain hospitals, such as teaching and inner city hospitals.[6] This may allow a hospital to profile its cost of caring for patients, identify DRG "losers," and transfer them to other institutions and/or discourage their admission.[7]

OUTLIERS

Before evaluating severity of illness or the resource variation of patients within a DRG, an explicit definition of which patients actually constitute the DRG must be specified. Some patients' clinical conditions will always preclude their simple categorization. For example, a patient admitted for a cataract operation who has a myocardial infarction and then experiences a pulmonary embolus during the hospitalization does not "fit" in the group of patients who undergo uneventful cataract surgery. To include such a patient in the uncomplicated group and *then* attempt to evaluate the entire group's severity of ill-

Received from the Departments of Internal Medicine and Health Services Management and Policy, University of Michigan, Ann Arbor, Michigan.

Address correspondence and reprint requests to Dr. McMahon: 3116 Taubman Center, Department of Internal Medicine, University of Michigan Medical Center, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-0376.

ness (or resource use) confounds the discussion. Those patients whose resource use deviates *significantly* from the DRG to which they are assigned should be clearly identified as outliers. Because there are many reasons why outlier patients may exceed the resource consumption estimates for their assigned groups, outliers require case-by-case review.[8] Only after the outlier patients have been removed can the analysis of the reasons for variation *within* a DRG be undertaken.

REASONS FOR INTRA-DRG RESOURCE VARIATION

There are numerous reasons for variation in hospital resource use within a given DRG. These include 1) data errors; 2) physician practice variation; 3) inadequate number of patients within the DRG; 4) limitations of the data elements collected, the Uniform Hospital Discharge Data Set (UHDDS), and the diagnosis coding system, International Classification of Disease, 9th revision, Clinical Modification (ICD-9-CM); and 5) clinically significant patient differences related to severity of illness.[6] These reasons for variation within a DRG can be categorized operationally into two groups; 1) those for which severity of illness is not an issue, 2) and those for which severity of illness is potentially a confounding factor.

The first three reasons for variation within a DRG are largely unrelated to differences in severity of illness. The first, errors in recording data, has little to do with variations in severity of illness. The second, physician practice variation, identifies physicians who use different amounts of resources to treat patients with the same diagnosis and severity of illness. The third, inadequate numbers (too few patients in a hospital to ensure a statistically stable group), is also not primarily a severity problem. There are so few patients in these small groups one could never construct an intra-group severity stratification. It is in the last two areas of intra-DRG resource variation, coding limitations and clinically significant patient differences, that one would expect a severity of illness adjustment to be the most beneficial in providing more equitable payment.

Before discussing severity of illness measures, we will review each potential reason for intra-DRG resource variation. This discussion presumes that statistically unusual cases, the outliers, have already been removed, as noted above.

Data Errors

One reason for observed intra-DRG resource variation is an error, in the coding of the clinical data and/or in the tabulation of the resource data. For example, at one major teaching hospital, a patient assigned to DRG 75, Major Chest Procedures, had a total bill of $200 (personal communication). Because this DRG requires that the patient have a major surgical procedure, either the billing data or the clinical data are in error. Such errors need to be corrected before one attempts to explain residual intra-DRG resource variation.

Physician Practice Variation

Many studies have evaluated the ability of differences in physician practice to explain variations in resource use. The resources examined to date include the laboratory[9-11] and the use of the hospital itself.[12-16] Recently, Horn et al. and McMahon and Newbold have demonstrated that variations in physician practice explain 20 to 40% of the variance in charges and 16 to 42% of the variance in lengths of stay within a DRG.[17, 18] Physician practice variation, therefore, must be adjusted for before evaluating severity of illness to explain resource use.

Inadequate Numbers

As noted by the developers of the DRGs, "there must be a manageable number of classes."[19] In the latest version of the DRG system, if fewer than three patients per year of a given type would have been seen at a typical 300-bed community hospital, the group was not formed.[20] A few exceptions to this rule were allowed, where a few cases were easily identifiable, concentrated in specific hospitals, and costly. Such special DRGs were defined for both heart transplant and kidney transplant patients.

Limitations of UHDDS and ICD-9-CM

Any patient classification system depends on the quality of the data upon which it is based. The DRG system is limited to the data on the UHDDS abstract, with diagnoses and procedures coded in ICD-9-CM.[21, 22] The ability of the DRG system to describe "subclasses of patients to form homogenous diagnostic categories"[19] is constrained by the limitations in the breadth of clinical data abstracted on UHDDS and by the lack of clinical specificity in segments of the ICD-9-CM system.[23, 24] In this area the methodology used in DRG construction faces its most serious problem in terms of differentiating levels of severity of illness.

For example, UHDDS abstracts provide only basic demographic, diagnostic, and procedural data. The abstracted record does not contain laboratory or physiologic data, which might distinguish among different levels of severity of illness, and thus better predict resource requirements for subsets of patients within a DRG.

The International Classification of Disease, 9th revision, Clinical Modification coding system is

based on the parent ICD-9 system, developed largely for epidemiologic purposes. Because of the origin of this system, its clinical specificity is meager in many areas. Examples of the inability of certain ICD-9-CM diagnostic and procedural codes adequately to reflect differences in patients' clinical states are common. For example, *all* patients who suffer a cerebrovascular accident are described by three codes; 434.0 Cerebral Thrombosis, 434.1 Cerebral Embolism, or 434.9 Cerebral Artery Occlusion, Unspecified. It is clear that key variables that predict the severity of the stroke (e.g., mental status, presence or absence of aphasia, paralysis, and the like) are lacking in these descriptions.[23] The data limitations of UHDDS and ICD-9-CM's lack of clinical specificity severely limit the ability of the DRGs or any system relying on these data to reflect differences in severity of illness.

Clinically Significant Patient Differences: Severity of Illness Measurement

Before discussing the application of severity of illness measures, the various components of severity must be explicitly identified. As Gertman and Lowenstein have noted, "Severity is what sociologists term a folk wisdom word, like satisfaction or happiness, operationally indefinable in a way that is perfectly acceptable to all parties."[25] Nonetheless, we must attempt to classify proposed severity measures to determine which aspect(s) of severity of illness they purport to measure. For the purposes of classification we propose to identify three foci for any severity of illness system: the physiologic focus, the psychologic focus, and the economic focus. Any severity of illness measure may attempt to measure one, or more than one, dimension.

In evaluating a proposed severity of illness measure it is necessary *explicitly* to identify its focus or foci and to review the data used in its validation to ensure that the focus and subsequent validation are congruent. For example, if a proposed severity of illness measure purports to measure the physiologic dimension it must use physiologic data in its validation, not length of stay or cost data.

The physiologic focus of severity of illness attempts to stratify patients on clinical grounds, in terms of their increased likelihood of morbidity and/or death. This is the definition that most physicians use when they describe a patient as being severely ill. Systems designed to measure physiologically defined severity of illness are *not* developed to meaure resource consumption and cannot be validated using resource consumption as an outcome; any subsequent resource analysis is a secondary analysis. The psychologic definition of severity of illness includes elements that have to do with the patient's and family's emotional responses to the illness. Severity of illness from the economic perspective refers to the patient's resource use.[6] It is this latter aspect of severity that has received the most attention from payers and providers alike. What many hospital administrators and some physicians mean when they say "adjust for severity" is "adjust for resource intensity or cost."

An example of two different patients will highlight the differences in the definitions of severity of illness as they relate to these differing perspectives. Consider two patients, each scheduled for a thoracotomy for possible lung cancer. The first is found at surgery to have cancer with local spread, so primary resection is not performed. The second patient is found to have an inflammatory mass, which requires lobectomy for treatment. From the physiologic perspective the patient with terminal cancer is more severely ill as he is closer to death and is likely to have more physiologic derangement. Likewise, from the psychologic perspective such a patient is more severely ill, not only because of his more serious disease and the physiologic responses to cancer, but also because of the patient's and family's emotional needs in facing a terminal illness. On the other hand, the patient who has the wide resection of a benign lung lesion is likely to be viewed as being of higher severity from an economic viewpoint, since such a patient may require more hospital resources than the basic, supportive care of the post-simple-thoractomy cancer patient. While this type of comparison may appear clear from a clinical perspective, it is not unusual to find in studies of severity of illness measures that the authors have evaluated the severities of illness of patients based on their hospital bills, as is discussed later.

SEVERITY OF ILLNESS MEASURES— STATE OF THE ART

To evaluate the principal existing severity of illness measures, one must identify whether they purport to measure the physiologic, psychologic, and/or economic dimension(s); whether they use explicit or implicit data, whether the data necessary for the scale's application are readily available, and whether the system has been appropriately validated (see Table 1). Given the financial success thus far of DRG-based prospective payment, our subsequent discussion assumes that DRGs are unlikely to be replaced by an entirely new patient classification system. Therefore, Patient Management Categories, developed by Young et al., which provide a different approach to disease classification and are largely incompatible with the DRGs, are not explicitly included in this analysis.[26, 27]

TABLE 1

Severity Systems' Foci and Subsequent Validation*

	Severity Focus		
	Physiologic	Psychologic	Economic
Severity of Illness Index	D	D	D,V
Disease Staging	D†		V
APACHE[44]	D, V		
MEDISGRPS®[49]	D		D, V

*D = designed to evaluate this focus; V = validated for this specific focus (dimension).

†Physiologic validation limited to a significant mortality regression coefficient in a study of mortality rates for only three surgical procedures.[41]

Severity of Illness Index

The Severity of Illness Index of Horn et al. is a modification of the AS-SCORE developed in the late 1970s for quality assurance purposes.[28] The Severity of Illness Index (SII) uses the system's implicit criteria to assign to each case an overall severity rating from 1 to 4. This final 1-to-4 score represents, within general guidelines, an implicit integration of seven subcategory scores: stage of principal diagnosis, complications, interaction, dependency, non-operating-room procedures, and response to therapy, including both the rate of response and residual disease.[29] The SII is not disease-specific but generic, and includes the physiologic, psychologic, and economic severity orientations. Finally, this four-point scale can be appended to any existing case-mix system, such as DRGs, or it can stand alone.

Although the concept of a multidimensional scale such as the SII is, in theory, appealing, in application it has several shortcomings. First, because a clinical validation has never been published there is uncertainty about what, in fact, is being measured (see Table 1). Studies to date have used reduction in variance in hospital charges or lengths of stay as the dependent variable.[5, 17, 30-33] A principal question therefore is: Does this scale differentiate among patients of different clinical states, including the physiologic and psychologic perspectives of severity of illness? Validation of this aspect of the scale would require the use of morbidity, mortality, and psychologic measures rather than hospital charges or length of stay as the dependent variable. If a scale is validated only by its ability to predict resource use, no comment can be made about its ability to stratify patients along the physiologic or psychologic aspects of severity of illness. Additionally, such variables as dependency and response to therapy appear to be correlated closely with length of stay, and the procedure variable correlated to chargeable items, thus raising the possibility of covariation among some independent and dependent variables used in validation studies of the SII. Finally, a recent study has demonstrated a significant association between the SII and the Adverse Patient Occurrence Index, suggesting that the SII may be measuring not severity of illness but rather complications of treatment.[34]

Further shortcomings of the SII suggested by others include its subjectivity and the minimal independent assessment of the scale.[4, 35] Although it is compatible with the DRG system, a four-level scale combined with the more than 470 existing DRGs would form over 1,880 groups. This would render many groups unevaluable due to inadequate sample size. Thus, application of the SII would require selection of the DRGs to be modified. Furthermore, this system requires *manual* chart review, so the potential for fraud would necessitate an extensive monitoring system if it were to be adopted by Medicare for PPS. Finally, despite assurances to the contrary from the developers, a recent independent study demonstrated that the index had poor interrater reliability.[29, 34]

Taken together, these shortcomings make it unlikely that the SII system, in its present form, would provide an improvement to the Prospective Payment System in the near future. A computerized version of the Severity of Illness Index called the Computerized Severity Index is currently under development. While it is reportedly designed to address some of the concerns noted above, there is no published report to date to allow reviewers to evaluate the system adequately.[36, 37]

Disease Staging

The Disease Staging (DS) system was developed by Gonella and colleagues in the 1970s for quality review purposes.[38] This system was designed by physician panels to segregate diagnoses, based upon medical criteria, to partition patients in terms of their probabilities of increasing morbidity and death. Thus, this scale is designed to measure the physiologic component of severity of illness. The scale ranges from 1 to 4, with a variable number of subclasses depending upon the diagnosis. For example, diabetes has four major stages ranging from stage 1, diabetes mellitus without other problems, to stage 4, death, and five subclasses, for a total of nine levels.[39, 40]

Work on the DS system has yet to validate its physiologic partitioning. It does not attempt to measure the psychologic dimensions of severity, and it would be difficult to predict mortality without some modification as this is already part of the scale. Rather than examining its ability to predict morbidity, validation studies to date have used resource use as the dependent variable. Thus, although the DS system was designed to measure the physiologic component of severity of illness, most published studies have evaluated only this scale's ability to

account for variance in the economic severity dimension.

One study designed to evaluate hospital-specific mortality did include DS values as an independent variable. In a multiple regression with death as the dependent variable, the DS independent variable did have a positive and significant coefficient in three of four surgical groups evaluated.[41] The association of a positive coefficient in this one study of four surgical groups represents the only published analysis of the physiologic component of DS to date. Thus, like the SII, the DS system has yet to receive full clinical validation.

Disease Staging's ability to explain intra-DRG resource variation has been rather limited.[42] Moreover, if just the major stages (1 – 4) were used to modify the existing DRGs by the DS categories, too many groups would be constructed. Thus, as with the SII, some degree of selection would be necessary to arrive at a manageable number of groups. Additionally, because DS and DRGs are distinctly different in their methods of assigning patient abstracts to a group, they are largely incompatible systems.[43] At the operational level, DS uses available abstracted UHDDS data, making chart review or the collection of additional data unnecessary and hospital manipulation of the scores more difficult. In summary, it is not yet clear that this clinically unvalidated severity measure would add significantly to the existing DRGs.

APACHE

The development of the Acute Physiology and Chronic Health Evaluation (APACHE) severity of illness scale began in 1978. It was designed to categorize the physiologic severity of illness of intensive care unit (ICU) patients and to evaluate their care.[44, 45] To date it is the *only* severity of illness scale that has been clinically validated in terms of its ability to predict patient death[46, 47] (see Table 1). It is designed to measure the physiologic component of severity of illness, that is, the probabilities of morbidity and death, and it appears to do this well in the ICU setting. The current APACHE II scale uses laboratory data, physiologic values (such as blood pressure, pulse), age, chronic health points, and the Glascow Coma Score to develop the composite score used to predict ICU mortality.[45]

The adaptation of APACHE to a general hospital population would present four principal problems: 1) it requires data not routinely collected in the current UHDDS hospital abstract form; 2) it was developed for ICU patients, who have higher probabilities of having abnormal laboratory and physiologic values (the independent variables) and mortality (the dependent variable); 3) it has yet to be applied to or validated in a general hospital population; and 4) it was not designed to capture the economic or psychologic dimensions of severity. Its favorable features are: 1) it is largely objective; 2) it has been shown that high scores correlate with mortality in the ICU setting; and 3) because it is represented as a continuous variable, identifying clinically significant differences in severity can be made on a DRG-by-DRG basis, thus preserving the integrity of the DRG approach.

MEDISGRPS®

The development of the MEDISGRPS® severity classification system began in 1981. The purpose of the MEDISGRPS system was to "describe physician- and hospital-specific performance in terms of patient outcomes (effectiveness) and resource use (efficiency)."[48] MEDISGRPS thus attempts to measure both the physiologic and the economic dimensions of severity of illness.

In order to accomplish this task, physicians reviewed the information transmitted during the review of cases by house officers the day after admission (at morning report) in an attempt to identify the clinically important information. They found that the reason for admission, physical examination findings, the results of laboratory tests, x-ray results, and the like were important in predicting outcomes and resource use. The developers identified a list of over 500 Key Clinical Findings (KCFs) for different presenting complaints in order to identify patients with the potential for organ failure. The KCFs for each presenting complaint were then organized to develop a severity stratification. The severity scale for the KCFs ranges from 0 for patients with minimal findings, indicating a low potential for organ failure, to 3 for patients with critical findings indicating the presence of organ failure. The admission severity scores utilize the worst KCF values during the first 48 hours of admission. The average patient has five to ten KCFs assigned. The individual KCF scores are then integrated via the system's algorithms to arrive at a final 0 to 4 admission severity score.

This system has many attractive features. First, it makes use of data that have clinical "face validity," i.e., the approach makes sense clinically. Second, it is directed at the physiologic dimension of severity of illness. Third, many measures are objective (e.g., vital signs and laboratory data), as in the APACHE system, although much of the "objective" information is subject to interpretation (e.g., physical findings, x-ray findings).

The system has four principal drawbacks: first, because it, like the SII, is a proprietary system, most of the evaluations of it have come from the developers. Second, the system's ability to segregate mortality risk (physiologic outcome) has been analyzed using the *admitting* diagnosis. Obviously, when two patients have the same admitting diagnosis, chest pain, but one has reflux esophagitis and the other has an acute myocardial infarction, they are not at the same risk of dying. A severity system that identifies these simplistic stratifications (i.e., the mortality difference between a myocardial infarction and reflux esophagitis) would be of little value as a severity of illness modifier. The principal question for a severity of illness measure is: can it differentiate intra-diagnosis rather than inter-diagnosis severity differences? Third, in some studies published to date, it is unclear whether very unusual patients, the outliers, have been excluded from the analysis.[49] As noted above, the outliers can be readily identified on statistical grounds. It is not clear that a severity system whose major achievement is to identify these very unusual cases would be worth its incremental cost. Fourth, it was not designed to measure the psychologic severity dimensions. Finally, in a recent study, Iezzoni and colleagues found that, with outliers removed, the addition of the MEDISGRPS severity score provided modest improvement in the explanation of DRG-specific cost.[50]

In summary, MEDISGRPS, like APACHE, contains attractive features. However, unlike APACHE, it requires extensive chart review and includes subjective interpretations. Furthermore, prior to its adoption as a severity adjustor for prospective payment it would need to be validated by independent investigators. The requirement that charts be reviewed, while providing a level of detail unavailable with just UHDDS-level data, makes the system difficult to adapt to a national payment system, as it would require extensive monitering.

COMPARATIVE STUDIES

Relatively few studies have compared the existing severity of illness measures using a common database. Two recent studies attempted to address this important issue. Calore and Iezzoni compared Disease Staging and Patient Management Categories (PMCs) in cases of pneumonia and prostatic disease.[51] They conclude that neither of these two systems did as well as DRGs in explaining variation in costs in the tracer conditions, and that while within some of the individual DRGs there is some increased cost variance that can be explained with these severity systems, their overall effect is likely to be modest.

In the most ambitious comparative study to date, Thomas, Ashcraft, and Zimmerman have compared PMCs, APACHE II, MEDISGRPS, Clinical Staging (the descriptive version of Disease Staging requiring individual chart review), and Disease Staging.[52] They analyzed these systems using data from four major teaching hospitals and two community hospitals and evaluated the severity of illness systems in selected medical and surgical DRGs. Thomas et al. evaluated the severity of illness systems within groups of adjacent DRGs along the following dimensions: construct validity, content validity, predictive validity, interrater reliability, potential for manipulation, and cost of implementation and operation. They found that no one severity of illness system was clearly better than the others along each dimension. Although this study assessed the severity of illness measures along a multitude of dimensions, the bulk of the analysis and the major recommendations deal with the ability of the severity of illness measures to explain differences in hospital costs. As with previous validations, this analysis allows economic, rather than clinical, severity considerations to predominate.

Studies comparing severity of illness measures are essential if we are to understand the relationship between a patient's illness and the associated resource consumption and outcome. Current efforts to develop and evaluate severity of illness measures have been skewed by the use of DRGs for payment and the resultant desire to use severity of illness systems to modify the payment system. Although hospitals are concerned with equitable payment rates, Jenks and Dobson have emphasized that the Medicare Prospective Payment System has many areas where payment can be affected (urban/rural adjustments, indirect teaching adjustments, etc.) and that there is no evidence that any severity of illness system would affect overall payment levels *between* hospitals.[53]

While the debate over the effects different types of patients have on hospital payment is important, it has disrupted the analysis of severity of illness systems. In the analysis of severity illness measures the principal focus has been on their abilities to explain variations in hospital costs—the economic focus. Relatively little attention (except for APACHE II and perhaps MEDISGRPS) has been directed toward the analysis of any system's ability to stratify a hospital patient's severity of illness along the physiologic and/or psychologic dimension. When the latter two dimensions are not examined, it is not clear whether a "severity of illness" measure is not, in fact, merely a "resource intensity" measure—a difference that is more than semantic.

SEVERITY OF ILLNESS MODIFICATION OF DRGs

Over the next five to ten years, DRGs are likely to continue to be an integral part of the system of payment for hospital care. In 1984 the number of hospital admissions for those over age 65 decreased for the first time.[54] Additionally, the increase in health expenditures was the smallest in 19 years and the percentage of the gross national product devoted to health care decreased to 10.6%.[55] While the Prospective Payment System may not be responsible directly for all of these decreases in lengths of stay and expenditures, it is a strong contributing factor. Given the federal deficit, it is unlikely that a system effecting such changes will be scrapped unless pervasive adverse health effects are documented and causally linked to prospective payment.

For the near future, therefore, any change in the PPS will probably be incremental. Systems to severity-adjust DRGs will need 1) to be objective, 2) to be validated in terms of their physiologic predictive value, and 3) as much as possible, to use existing data. Of the four severity systems outlined above, the only one clearly to meet these three criteria at the present time is the APACHE system. However, before such a system could be adopted, it would need to be tested in terms of its ability to stratify general hospital patients according to their risks of morbidity and mortality. Thus, although APACHE II has been shown to be a valid predictor of intensive care mortality, it may do less well in predicting mortality in a general hospital setting, in which patients have fewer physiologic derangements. The other systems, the Severity of Illness Index, Disease Staging, and MEDISGRPS, have yet to be validated clinically as severity measures (see Table 1). Furthermore, because the SII uses subjective data it is manipulable and thus less suitable for a national payment system.

CONCLUSION

The use of DRGs for payment has focused attention on two of their limitations: 1) the current ICD-9-CM disease classification system and its lack of clinical specificity, and 2) the lack of availability of a clinically validated severity of illness adjuster for general hospitalized patients. If a short-term severity adjustment of DRGs is attempted, the most fruitful efforts will need to focus on improving the ICD-9-CM coding system, and to test the ability of an objective system such as APACHE to modify DRGs. An APACHE-like approach (which includes elements of MEDISGRPS) is preferred because it is objective, has been clinically validated, and makes extensive use of existing data (laboratory data, diagnostic data, age).

Ultimately, to capture the various aspects of severity of illness, *multiple* dimensions of a patient's illness will need to be measured. For example, it will be necessary to describe the patient's disease using a coding system such as ICD-9-CM, in addition to identifying selected physiologic variables to differentiate important clinical differences. Furthermore, evaluation of some diseases (e.g., strokes) will require measurements of the patient's motor, self care, and cognitive functions utilizing an approach such as the KCFs in MEDISGRPS. Finally, it may be necessary to evaluate the effect on patients' disease states of their overall health status or socioeconomic strata. It is likely that this ideal severity stratification system is many years away and that incremental gains will be made focusing on one or two dimensions of severity and developing valid systems for general hospital patients along these selected dimensions.

To capture the interaction of these unique aspects of severity of illness will require the collections of new variables. It is doubtful that focusing on only one dimension of this multidimensional problem will prove fruitful in the long run across a wide variety of illnesses. One quickly approaches diminishing returns as one adds more of the same type of variables to a stratification system. This fact was evident in the developments of both the DRG definitions and the APACHE system: in each instance, multiple variables along one dimension were initially utilized, and after evaluation, a subset of variables was found to capture most of the important interactions.

In the short term, there will be increased pressure to adopt a severity of illness system, even if there are only limited data to substantiate its utility among the three dimensions of severity measurement. The challenge of rapid development of a valid measure of severity of illness is compounded by the proprietary nature of such prominent severity of illness measurement systems as MEDISGRPS and the Computerized Severity Index. If the field of severity of illness measurement becomes dominated by proprietary systems that limit independent evaluation, the field will suffer. While the area of severity of illness measurement is of utmost importance, the funds available for health services research are limited. This precludes a sustained effort of the type necessary to build, test, and modify a new severity of illness measure.[56] Thus, open evaluation of the existing systems is essential. The importance of an open discussion of significant health care system modifiers was demonstrated by the lively and valuable discussion of the strengths and weaknesses of the DRG patient classification system. If severity of illness systems emerge as proprietary "black boxes," such frank and open discussion will be impaired.

In the long term, physicians will need to become more actively involved in the development of clinically meaningful stratification systems to segregate patients more precisely. Such clinical classification systems will be necessary not only to adjust payment systems, but also to improve the validity of quality assurance programs and of clinical trials that attempt to evaluate efficacy and cost effectiveness of new diagnostic and therapeutic modalities. Such severity measures will also be essential to monitor the quality of care, so that poor outcomes that result from a patient's increased severity of illness can be differentiated from those caused by poor medical practice.[47] Without the development of such clinical stratification systems, physicians will find themselves locked out of the health policy debate for want of credible data.

The authors thank Drs. Jeoffrey K. Stross and Joel Howell and anonymous reviewers for their helpful comments on an earlier version of the manuscript, and Laura Bersuder for her assistance in preparing the manuscript.

REFERENCES

1. Mendenhall S. DRG's must be changed to take patient's illness severity into account. Modern Health Care 1984; Nov 15:86-8.
2. Muller C. Paying hospitals: how does a severity measure help (editorial). Am J Public Health 1983;73:14-5.
3. Eisenberg BS. Diagnosis-related groups, severity of illness and equitable reimbursement under Medicare (editorial). JAMA 1984;25:645-6.
4. Schweiker RS. Report to Congress. Hospital prospective payment for Medicare. Department of Health and Human Services. 1982; Dec: 1-107.
5. Horn SD, Bulkley G, Sharkey PD, Chambers AF, Horn RA, Schramm CJ. Interhospital differences in severity of illness. Problems for prospective payment based on diagnosis-related groups (DRGs). N Engl J Med 1985;313:20-4.
6. Smits HL, Fetter RB, McMahon LF. Variation in resource use within diagnosis-related groups: the severity issue. Health Care Financing Review. 1984; (suppl) Nov: 71-8.
7. Leikin JB, Polin KS, Schiff RL, et al. Transfers to a public hospital: a prospective study of 467 patients. N Engl J Med 1986;315:552-7.
8. McMahon LF. Diagnosis-related groups prospective payment: effects on medical quality assurance. Evaluation and the Health Professions 1984;7:24-41.
9. Griner PF, Liptzen B. Use of the laboratory in a teaching hospital: the implication for patient care education, and hospital costs. Ann Intern Med 1971;75:157-63.
10. Linn LS, Yager J, Leake BD, Gastaldo G, Palkowski O. Differences in the numbers and costs of tests ordered by internists, family physicians, and psychiatrists. Inquiry 1984;22:266-75.
11. Eisenberg JM. Physician utilization. The state of research about physician's practice patterns. Med Care 1985;23:461-83.
12. Cornell FA, Blide LA, Hanken MA. Clinical correlates of small area variation in population-based admission rates for diabetes. Med Care 1984;22:939-49.
13. Wennberg JE, Gittleson A. Variations in medical care among small areas. Sci Am 1982;120-34.
14. Wennberg JE, McPhearson K, Caper P. Will payment based on diagnosis-related groups control hospital costs? N Engl J Med 1984;311:295.
15. Stockwell H, Vayda E. Variations in surgery in Ontario. Med Care 1979;17:390-6.
16. Roos LL, Roos NP. Assessing existing technologies: the Manitoba study of common surgical procedures. Med Care 1983;21:454-62.
17. Horn SD, Horn RA, Sharkey PD. The Severity of Illness Index as a severity adjustment to diagnosis related groups. Health Care Financing Rev 1984;(suppl):33-45.
18. McMahon LF, Newbold R. Variation in resource use within diagnosis related groups: the effect of severity of illness and physician practice. Med Care 1986;24:388-97.
19. Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. Med Care 1980; (suppl)18:1-53.
20. Health Systems Management Groups. The new ICD-9-CM diagnosis related group classification scheme. Final report. United States Department of Health and Human Services. Health Care Financing Administration Publication #03167, September 1984.
21. Uniform hospital discharge data; minimum data set. Report to the National Committee on Vital Health Statistics, DHEW Pub. No. PHS 80-1157.
22. Commission on Professional Hospital Activities. ICD-9-CM international classification of diseases, 9th revision, clinical modification. Ann Arbor, MI: Commission on Professional Hospital Activities, 1978.
23. McMahon LF, Smits HL. Can the Medicare prospective payment system survive the ICD-9-CM disease classification system? Ann Intern Med 1986;104:562-6.
24. Mullin RL. Diagnosis-related groups and severity: ICD-9-CM, the real problem. JAMA 1985;254:1208-10.
25. Gertman PM, Lowenstein S. A research paradigm for severity of illness: issues for the diagnosis-related group system. Health Care Financing Rev 1984; (suppl), 85.
26. Young WW. Incorporating severity of illness and comorbidity in case-mix measurement. Health Care Financing Rev 1984; (suppl):23-31.
27. Young WW, Swinkola RB, Zurn DM. The measurement of hospital case mix. Med Care 1982;20:501-12.
28. Roveti GC, Horn SD, Kreitzer SL. "AS-SCORE": a multi-attribute clinical index of illness severity. Quality Rev Bull 1980; July; 25-31.
29. Horn SD, Horn RA. Reliability and validity of the severity of illness index. Med Care 1986;24:159-78.
30. Horn SD. Measuring severity of illness: comparisons across institutions. Am J Public Health 1983;73:25-31.
31. Horn SD, Sharkey PD. Measuring severity of illness to predict resource use within DRGs. Inquiry 1983;20:314-21.
32. Horn SD, Sharkey PD, Bertram D. Measuring severity of illness: homogeneous case-mix groups. Med Care 1983;21:14-31.
33. Horn SD, Horn RA, Sharkey PD, Chambers AF. Severity of illness within DRG's homogeneity study. Med Care 1986;24:225-35.
34. Schumacher DN, Parker B, Kofie V, Munns JM. Severity of Illness Index and the Adverse Patient Occurrence Index: a reliability study and policy implications. Med Care 1987;25:695-704.
35. Jenks SF, Dobson A, Willis P, Feinstein PH. Evaluating and improving the measurement of hospital case mix. Health Care Financing Rev 1984;(suppl):1-11.
36. Horn SD. Measuring severity: how sick is sick? How well is well. HealthCare Financial Management October 1986, 21-32.
37. Backofen J, Ashworth MA, Horn SD. The computerized severity of index. A new sophisticated tool to measure hospital quality of care. HealthCare Form, March/April 1987, 35-37.
38. Gonella JS, Louis DZ, McCord JJ. The staging concept — an approach to the assessment of outcome of ambulatory care. Med Care 1976;14:13-21.
39. Gonella JS, Hornbrook MC, Louis DZ. Staging of disease: a case-mix measurement. JAMA 1984;637-44.
40. Systemetrics Inc. Disease staging: a clinically based approach to measurement of disease severity, Vol 1. Executive Summary National Technical Information Service: PB 83-254656. 1983;(Aug):1-25.
41. Kelly JV, Hellinger FJ. Physician and hospital factors associated with mortality of surgical patients. Med Care 1986;24:785-800.
42. Conklin JE, Lieberman JV, Barnes CA, Louis DZ. Disease staging: implications for hospital reimbursement and management. Health Care Financing Rev 1984;(suppl):13-22.
43. Coffey RM, Goldfarb MG. DRGs and disease staging for reimbursing Medicare patients. Med Care 1986;24:814-29.
44. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE — Acute Physiology and Chronic Health Evaluation: a physiologically based classification system. Crit Care Med 1981;9:591-7.
45. Wagner DP, Draper EA. Acute Physiology and Chronic Health Evaluation (APACHE II) and Medicare reimbursement. Health Care Financing Rev 1984;(suppl):91-105.
46. Wagner DP, Knaus WA, Draper EA. Statistical validation of a severity of illness measure. Am J Public Health 1983;73:878-84.

47. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. Ann Intern Med 1986;104:410-8.
48. Brewster AC, Jacobs CM, Bradbury RC. Classifying severity of illness by using clinical findings. Health Care Financing Rev (suppl) 1984; 107-8.
49. Brewster AC, Karlin BG, Hyde LA, Jacobs CM, Bradbury RC, Chac YM. MEDISGRPS®: a clinically based approach to classifying hospital patients at admission. Inquiry 1985;22:377-87.
50. Iezzoni LI, Ash AS, Moskowitz MA. MEDISGRPS®: a clinical analytic assessment. Waltham, MA: Health Policy Research Consortium, 1987.
51. Calore KA, Iezzoni L. Disease staging and PMCs: can they improve DRGs? Med Care 1987;25:724-35.
52. Thomas JW, Ashcraft MLF, Zimmerman J. An evaluation of alternative severity of illness measures for use by university hospitals. Ann Arbor, MI: Department of Health Services Management and Policy, School of Public Health, University of Michigan, 1986.
53. Jenks SF, Dobson A. Refining case-mix adjustment: the research evidence. N Engl J Med 1987;317:679-86.
54. Prospective Payment Assessment Commission. Medicare prospective payment and the American health care system report to Congress. February 1986; 19.
55. Levit KR, Lazenby H, Waldo DR, Davidoff LM. National health expenditures. 1984. Health Care Financing Rev 1985;7:1-35.
56. Shapiro MF, Larson EB. Funding for medical care research. J Gen Intern Med 1987;2:113-8.

REFLECTIONS

High-tech Dying

I KNOW AN OLD MAN IN IRELAND whose health is rapidly failing. If he lived here, he would probably be in a hospital. Instead, he lives by a lake in the cottage his father built. His family lives nearby. Altogether, he is a content man. He has never been in a hospital, and vows he never will be. I thought of this man a while ago when another elderly friend of mine lay dying of lung cancer in a West Coast hospital. He wanted to stay at home, but our health care system discourages that. And a nursing home would have wiped out his savings in a few weeks. So a bizarre medical game was played out. Machines bubbled and burped by his bed. Each morning he was bundled onto a hospital cart and taken downstairs for radiation. This was the necessary fiction. As long as those rays kept zapping, my friend remained eligible for insurance benefits. As soon as the treatment stopped, he would be out on his ear, with no money for nursing home care, and no family close by to care for him. So we pretended. One morning, the doctor in charge of zapping held up for me an x-ray showing a shriveled tumor. He marveled proudly, "Isn't this incredible? The tumor is half the size it was two weeks ago." So is my friend, I thought. "He is still dying, isn't he?" I asked. The doctor's voice turned as crisp as his white coat. "That's not my department," he said, "I just deal with tumors." He knew the code and he was playing the game. The problem with this modern minuet is that death sooner or later stops the dance. It comes barging right through the door, scatters comforting x-rays, and takes what it wants. When death comes for my friend in Ireland, however, it won't seem so bizarre or out of place. No one will be caught playing insurance games, because the dying man has not had to choose between losing his life savings and unnecessary care. Nobody should turn away from medical care that can really help. But high-tech dying makes me wonder if maybe death isn't better met in a cottage on the bank of an Irish lake.

The comments of Pat O'Brien, a writer living in Washington, DC. This news commentary was originally broadcast on National Public Radio's news and information magazine "Morning Edition" [January 4, 1988] and is printed with the permission of National Public Radio. Any unauthorized duplication is prohibited.