

Variation in resource use within diagnosis-related groups: The severity issue

by Helen L. Smits, Robert B. Fetter, and Laurence F. McMahon, Jr.

Several authors have suggested that diagnosis-related groups (DRG's) make inadequate allowance for the severity of illness. Before modifications of DRG's are developed, the sources of within-group variation must be precisely defined; not all variation is attributable to the severity of illness. The limitations of the Uniform Hospital Discharge Data Set (UHDDS), of the International Classification of Diseases, Ninth Revision, Clinical Modification coding

system and of the original rules of DRG construction must be evaluated and, if necessary, corrected before new approaches to groupings are considered. The most promising potential modifications of existing groups and weights are those that make use of the UHDDS, or of the UHDDS plus additional diagnoses and procedures. The addition of entirely new data elements to the discharge abstract and the pricing process should be considered only as a last resort.

Introduction

Prospective payment based on diagnosis-related groups (DRG's) is the most radical change in hospital financing since the emergence of modern hospital financing mechanisms in the United States in the 1930's and early 1940's. Any change so profound is guaranteed to arouse controversy; one unintended side effect of prospective payment has been an epidemic of conferences, special seminars, speeches, and written comments in which the pros and cons of the new payment system are actively debated.

The aim of this article is to discuss one of the central elements in that debate: whether or not DRG's adequately reflect the severity of illness. To do so, we will review the concept of severity, describe the strategic decisions on which DRG grouping is based, and explore the major known causes for excessive within-group variation in DRG's. We will then describe the kinds of decisions which are basic to the development of a severity modification of DRG's and will outline those approaches to DRG modification which we believe are potentially the most productive.

Why severity?

The DRG's are designed to identify groups of patients with homogeneous patterns of resource use. Homogeneous does not mean identical; fifty patients within a given DRG cared for at a single hospital by a single physician would not each cost exactly the same amount. Assuming both efficient practice and a well-chosen DRG price, these patients' costs should vary moderately around the price without a strong skew toward either high or low cost. Some variation in cost within a DRG is not a cause for concern; variation only becomes a problem when it is both systematic and a function of identifiable patient attributes. If certain identifiable types of patients tend to cost either more or less than the DRG specific price, then there is a risk that hospitals will begin to select for the low-cost group and against the high-cost group.

Reprint requests: Helen L. Smits, Yale School of Organization and Management, Health Systems Management Group, 12 Prospect Place, New Haven, CT. 06520.

Actual within-group selection by hospitals is not as easily manageable an activity as some commentators seem to assume. Hospitals don't make admitting decisions; doctors do. A hospital can't reject the expensive portion of a physician's practice without risking the possibility that the physician will take the remainder of his practice elsewhere—hardly a desirable outcome in an era when admission volume is of increasing importance. In addition, most patients will not always be in the same group: today's high-risk (and costly) myocardial infarction may well return in a few months as a low-risk (and profitable) coronary artery bypass. Rejection of the patient the first time poses a real risk that he won't come back again. Finally, a hospital's need to maintain a positive and attractive image within its community will tend to reduce any large-scale efforts to reject certain patients.

Despite multiple forces which serve to modify within-group selection, there are compelling reasons to consider whether or not price adjustments for severity are needed. The first is simple equity. If more severely ill patients cost more, and if referral patterns tend to cluster these patients in certain hospitals, then the hospitals that accept these patients will receive DRG payments which are insufficient to cover the costs of care. On the other hand, hospitals that do not care for them will receive an undeserved bonus. The second is the problem of access. If certain patients with chronic diseases are always costly when compared with the established price, then discrimination against them may emerge. If such discrimination involves few individuals, and is done discreetly, the hospital may be able to avoid both adverse community and physician reaction.

Severity defined

Severity of illness is a definition based on a particular point of view. A physician's view of the severity of illness refers to the impact of the particular disease process on the patient's physiologic integrity. Is this patient more likely to experience long-run sequelae than others? Is he more likely to die? For a nurse, severity of illness encompasses the physician's parameters while also including psychologic and dependency

needs of the patient. The young, healthy woman hospitalized with pyelonephritis in the post-partum period may be seen by her internist as very low on a severity scale. Nurses are more likely to see her as at high risk for an episode of depression or anxiety and will plan their care of her accordingly. The final perspective on severity is that of the manager or payer: how many hospital resources will be used in care and, as a result, how much will care cost. The current literature on severity of illness makes two assumptions: that severity is a generic concept, failing explicitly to define whose severity of illness is being studied, and that "sicker" patients necessarily cost more.

Traditional physician concepts of severity do not always correlate directly with the cost of care. In the case of the sickest patients, the correlation will be negative since death soon after admission—an event which defines the most severely ill subgroup—can be a relatively economical outcome. Even when death is not the outcome, severe illness may in some patients be associated with parsimonious resource use in comparison with others who have the same diagnosis, because the risks of diagnostic testing may be higher, the diagnosis may be easier to establish, or the potential benefits of certain treatments may be reduced.

The nursing definition of severity, by contrast, is likely to have a more direct relationship to hospital cost. This arises because nursing time, the resource used in treating patients perceived by nurses as severely ill, represents a direct cost to the hospital.

The need to establish a severity measure, then, is not an absolute one. Severity modification of DRG's is needed only if, and when, distinct patterns of resource use within existing DRG's are associated with specific and definable variations in the severity of illness.

DRG construction

DRG groupings were developed using only two of the three definitions of severity described above: the doctor's and the manager's. The nursing perspective was conspicuously absent from DRG construction, both because the available data contained no dependency or psychology-of-illness measures and because hospital accounting methods do not reflect patient-specific variations in the use of nursing time. The expression of physician's perspective in the final system was limited by the decision to use the Uniform Hospital Discharge Data Set (UHDDS), and no other information, in the construction of groups.

When DRG's were first designed, they were intended as a management system for hospitals, as a tool for utilization review, and as a research tool. The use of available data permitted the development of a system which was economical and feasible for all users. Groupings based on the UHDDS provided the maximum opportunity to make comparisons across hospitals and doctors. In the later phases of DRG development, when their use for payment had been proposed, the decision to use only the UHDDS was reaffirmed both because of the availability of large data bases on which to test the system and because

the elements of UHDDS were standardized, well understood by the relevant experts, and easily subject to audit. These characteristics are of primary importance in a national payment scheme.

Details of DRG construction have been described elsewhere (Fetter *et al.*, 1980; Health Systems Management Group, 1982). The AUTOGRP program allowed physician input into group design to meet the requirement for groupings that were both clinically logical and relatively homogeneous in their use of resources, thus merging the physician's concept of severity with the managers (Mills *et al.*, 1976). One additional constraint on DRG design was the aim for a limited number of classes to keep the scheme both comprehensible and manageable.

Sources of unstable DRG's

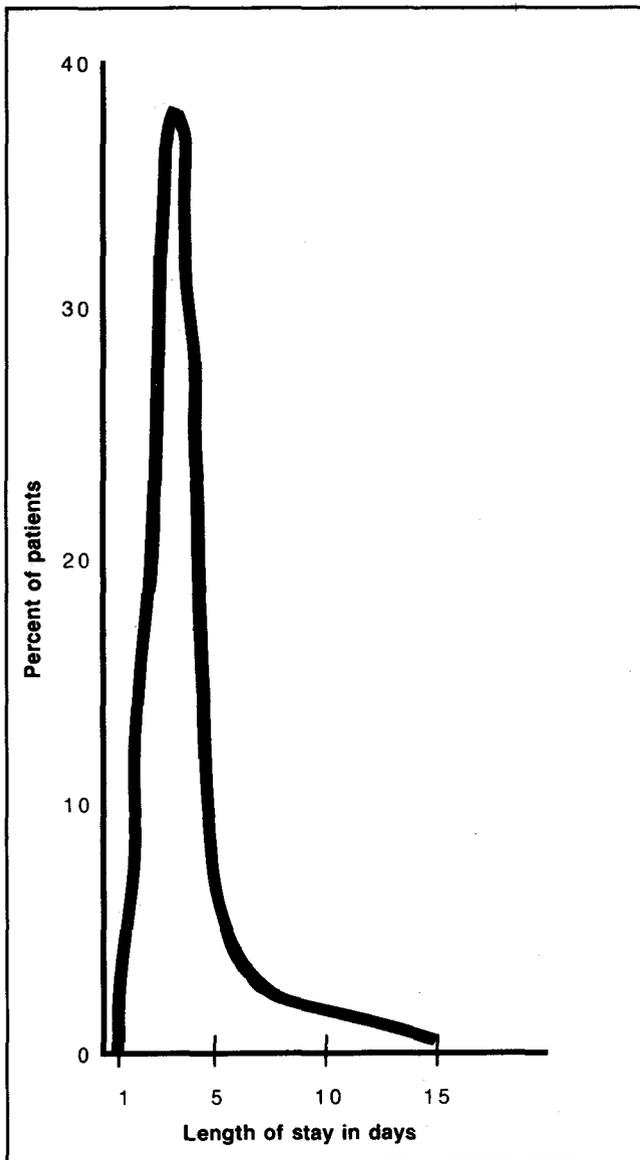
To place the issue of severity in perspective, the other known causes of DRG instability must be considered. If all DRG's from a large data base are plotted and inspected, some groupings appear more cohesive than others. Figure 1 illustrates a typical good, or stable, DRG with a tightly clustered pattern of resource use. Figure 2 shows a relatively unstable DRG which contains more variation. The stable DRG is 39, lens procedures, undertaken in patients judged capable of undergoing elective cataract surgery. The variations in operative approach in use at the time these data were collected had little effect on eventual patterns of resource use. DRG 14, specific cerebrovascular disorder except transit ischemic attack (TIA), by contrast, is much less cohesive. It appears that the patients within this DRG could be subdivided further if appropriate data were available.

A wide range of explanations have been postulated for variation in the degree of DRG cohesiveness in large cross-hospital data sets and within individual hospitals. (Thompson *et al.*, 1978; Smits *et al.*, 1984). The causes of this DRG instability vary, and as a result, the appropriate remedies will vary. Table 1 describes the possible causes for a pattern such as that observed in DRG 14 and describes in brief the treatment needed to correct each type of problem. Possible etiologies for DRG instability include:

Error

The first reason for apparent aberrancy is an error in either the abstract or the bill. High rates of coding error were demonstrated in the preprospective payment era (Institute of Medicine, 1977). Errors in recording resource use must also be considered, especially when bills are used as the source of information. Any critical analysis of DRG aberrancy must begin with an examination of the accuracy of DRG assignment. Erroneous records must be corrected or removed before proceeding with any analysis of the remainder of the group.

Figure 1
Percent distribution of length of stay of patients from Maryland in DRG 39, lens procedures



NOTE: The mean length of stay for this group of 1,729 patients is 3.7 days with a standard deviation of 1.9 days. Area under the curve equals 100 percent of the distribution.

Outliers

A second reason for unusually high- or low-cost cases is that something went wrong during the hospitalization. A patient may have experienced a complication of a diagnostic or therapeutic maneuver or may have developed a second or third illness while hospitalized. The causes of outlier status can be expected to vary widely. They may represent physician error, a rarely-occurring event that could not have been prevented, hospital-acquired disease, or the inefficient scheduling of hospital tests. It was to identify just such cases that DRG's were developed, all of them are important to review in detail. The results of review should be fed back to the hospital staff to

enhance the quality of care. Such cases, however, cannot be cited as indications that the hospital treats sicker patients than other hospitals or that the DRG price should be modified. Any good grouping system should produce outliers and should deal with them separately, because we know that there are many unclassifiable events within hospital practice and that attention to these events is an essential element in hospital management and quality control (McMahon, 1984).

Table 1
Types of DRG instability, their cause and treatment

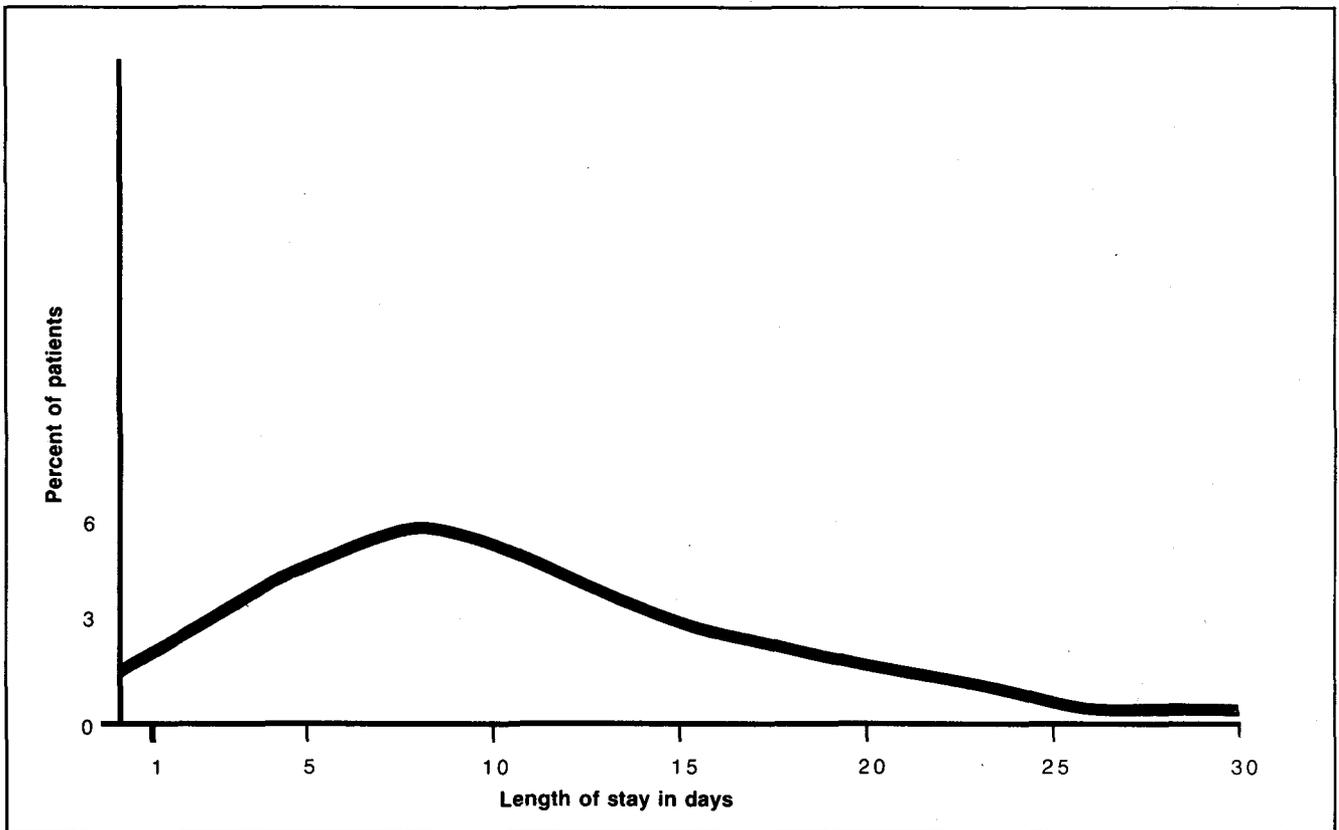
Type	Cause	Treatment
Error	Inaccurate discharge or cost information	Reclassify
True outlier	Aberrant case	None
Physician pattern	Variations in practice unrelated to patient condition	None at the national level
Small number	Group rejected by rules of DRG construction	Define if these patients cluster in specialized hospitals
Uniform hospital discharge data set (UHDDS) ICD-9-CM ¹	Data available, not now required	Expand UHDDS
Nursing severity	Dissimilar patients coded in same fashion	Revise ICD-9-CM or use special subcodes
Medical severity	No information in original data sets	Reevaluate groups and weights using nursing
	Disease description too vague	Add modifiers or use comorbidities

¹ International Classification of Diseases, Ninth Revision, Clinical Modification.

Physician practice patterns

The third reason is that the practice pattern of an individual physician or group of physicians may vary from the norm. This phenomenon, which has frequently been observed in utilization review efforts, can have a confounding effect on DRG patterns. Although DRG 39, lens procedures, is stable from a statistical standpoint, there is considerable variability in treatment patterns within the group. Our analysis of 1,729 records of cataract procedures performed in Maryland in 1981 demonstrated the extent of this variability even in this cohesive DRG. Overall mean length of stay at the time these data were collected was 3.7 days, with a standard deviation of 1.9 days with the outliers excluded. Yet, within the group, there are individual physicians whose patients' mean length of stay was 2 days, some for whom it was 3 days, and some for whom it was 4 and even 5 days. Physician-specific practice patterns which are unrelated to patient characteristics are naturally highlighted by DRG prospective payment; the resulting

Figure 2
Percent distribution of length of stay of patients from Maryland in DRG 14, cerebrovascular disorders except TIA



NOTE: The mean length of stay for this group of 5,550 patients is 11.57 days with a standard deviation of 7.4 days. Area under the curve equals 100 percent of the distribution.

variations in resource use are, self-evidently, not cause for DRG modification. The effects of practice patterns must, however, be considered in evaluating any grouping system proposed as a modification of, or substitution for, DRG's. Grouping methods which segregate different physician practice patterns will certainly appear to improve on DRG's; their use for pricing would be inappropriate.

Small subgroups

A fourth possible reason for variation within a single DRG at a given hospital is that there may exist classes of patients within some DRG's that were excluded from the original definitions because they occurred too rarely to warrant definition as a separate group. The criterion adopted in construction was that, if the expected number of patients seen in a 300-bed hospital in the United States was less than three per year, a class would not be defined. Exceptions were made for conditions known to involve the use of highly specialized resources and treated only in specialty hospitals. DRG 302, kidney transplant, is an example of a class that does not satisfy the frequency requirement, but was defined separately.

As a result of this relatively stringent frequency criterion, distinct patient subclasses probably exist within some DRG's which can be defined from the Uniform Hospital Discharge Data Set (UHDDS) using existing methods. If these subclasses tend to cluster in particular hospitals, then they should be defined and priced separately. Specialized hospitals need to examine in detail their outlier records to discern and identify such classes; pooled data can then be used to determine the prevalence of the subgroup, to examine its distribution across hospitals, and if appropriate, to develop a price.

Language failures

The remaining reasons why groupings may be inhomogeneous are best described as language failures. Either the clinical information supplied by the UHDDS is insufficient, clinically important distinctions are blurred in the process of coding, or the medical nomenclature itself is too imprecise to allow an appropriate distinction among patient groups.

Uniform hospital discharge data set

The clinical information in the UHDDS provides only principal diagnosis, up to four other diagnoses, and up to three surgical procedures. *Principal diagnosis* is defined as that diagnosis which is determined on discharge to have been the principal reason for admission. The use of principal diagnosis as an element in DRG construction means that those patients who enter the hospital with one condition but who remain because of another condition may become outliers. The patient who enters for cataract surgery, for example, and has a heart attack in the recovery room, is grouped as a cataract removal with complications. Grouping by *primary diagnosis*, which is defined as the diagnosis that accounted for the bulk of the hospital stay, would undoubtedly improve the classification of some patients. How much improvement would result is unclear, because we have been unable to identify a North American data set with reliable coding of primary diagnoses. The evidence from our French experiment suggests that the overall effect would not be great, but such clarification might well provide a marginal improvement for some individual hospitals.

Another possible change in the UHDDS would be an expansion of the coding of secondary diagnoses and surgical procedures. Although five diagnoses and three surgical events fully describe the majority of all hospital discharges, they may be insufficient for those patients who are most severely ill. Tests of an expanded UHDDS are needed to determine if grouping can be improved by using additional diagnoses and procedures.

ICD-9-CM

The coding system used for Medicare hospital discharges is that of the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM). The parent code, ICD-9, is widely used and subscribed to by multinational treaty. The ICD-9-CM is a system of diagnostic and procedural coding developed specifically for use in the United States. Major problems in the use of this system arise within the procedural codes.

The ICD-9-CM is a less specific procedural code than that used by Medicare for physician billing: the Common Procedural Terminology, Fourth Edition (CPT-4). The ICD-9-CM does not distinguish among different surgical approaches to the same problem. Because it is insensitive to the use of physician time, it has been deemed inappropriate for physician billing purposes. This same limitation may affect hospital billing under DRG's. The pattern within ICD-9-CM is to code on the basis of what is accomplished regardless of the technique or method used. This means that surgical procedures of different lengths and magnitude, and with dissimilar effects on patients and ultimately on resource use, may be coded in an identical fashion. The inclusion of dissimilar procedures under a single code is particularly likely to occur with the "grabbag" codes known as not elsewhere classified (NEC). As we have described in detail

elsewhere, the use of NEC codes may present a particularly serious problem for newly developed surgical procedures (Smits *et al.*, 1984).

One possible solution to the problems inherent in the *Clinical Modification* would be to revise DRG's using CPT-4 rather than ICD-9-CM as the source of procedural codes. Another somewhat simpler approach would be to modify the ICD-9-CM so that codes reflect the clinical impact of a given surgical procedure on patients. In a sense, this would mean the development of explicit surgical severity measures within the coding scheme. The development of improved NEC codes could improve both grouping and pricing for the affected patients.

Medical nomenclature and severity

Another challenging and intriguing problem arises in those instances when the language of medicine itself fails. As we have noted, the classic example of an unstable DRG is 14, specific cerebrovascular disorders except TIA. Behind 14 lies the word "stroke" or, more elegantly but no more precisely, "cerebrovascular accident," terms which contain little information on either the severity of the illness or its possible outcome. Out modern diagnostic ability to make better distinctions between thrombotic and hemorrhagic episodes has neither improved the capacity to treat the condition nor clarified prognostication.

To describe cerebrovascular accident more adequately both in terms of eventual outcome and expected resource use, more data are needed. The additional information needed for this diagnosis might include both the extent of the initial symptoms and the rapidity with which these symptoms change. One could conceive of a simple classification of cerebrovascular accident by both magnitude and change rate in which, for example, a patient with hemiparesis and marked improvement in 24 hours might be described as "CVA, large, rapid." One with loss of function in one arm with no change noted in the first days might be "CVA, small, slow."

Similar problems arise in many other conditions when a single etiologic event leads to a variety of outcomes. Myocardial infarctions, for example, range markedly in their effects on patients. Various severity scales based on physiological information are now in use that predict outcomes with considerable success (Killip *et al.*, 1967). Although there is, as yet, no firm evidence that these same scales could also be used to predict resource use, analysis of the relationships between severity as measured by such scales and cost is a logical extension of existing research.

Nursing time

The element most conspicuously absent from DRG construction is nurse time. This occurred because the use of nursing time by individual patients is not reflected by traditional hospital accounting methods. The only exceptions are in association with specialized care units where use by a patient indicates enhanced use of nursing time; costs associated with intensive

care units or recovery rooms are reflected in the appropriate DRG's. Because nursing time represents approximately 30 percent of all hospital costs and the use of nursing time varies widely among different patients on the same floor, considerable true cost variation attributable to nursing is to be expected.

Potential modification of diagnosis-related groups

The potential causes of DRG instability are listed in Table 1. Errors, outliers, and physician's practice patterns are not problems for DRG payment; they are included in this discussion only for clarity. The next three causes might best be described as DRG maintenance. Small groups, the limits of the UHDDS, and those of ICD-9-CM can be corrected without basic change in the grouping concepts now in use. The maintenance issues arising from the UHDDS and from coding are, of course, not peculiar to DRG's; they would arise in any grouping system which makes use of traditional hospital discharge information based on the ICD-9-CM.

Only the two final potential causes of DRG instability require consideration of a fundamental modification in DRG construction, either through the use of new data elements or through changes in the ways in which existing data are used and are directly related to severity of illness. Although many authors have discussed the medical and nursing aspects of severity as though they are the same, we prefer to maintain the distinction between them both for clarity and because of our expectation that appropriate solutions to the problems they pose may be quite different. In considering potential solutions, we need to keep in mind the constraints used in DRG design; of these, the two of most concern at this point are:

- How many groups are appropriate?
- What data elements are to be used? Are they reliable in the management sense of creating few areas of dispute and being easily subject to audit?

Nursing severity

Nursing severity measures cannot be developed without an accurate, reproducible, patient-specific measure of nursing time or nursing cost, an element which is absent from most traditional hospital accounts. Once such a data base is available, patterns of nursing practice within existing DRG's can be determined and modifications of existing weights can be developed. The use of nursing time as a well-defined part of the dependent variable may also lead to the development of distinct subgroups of existing DRG's, subgroups which could not be defined on the basis of routine and ancillary costs alone. In our view, research to introduce nursing weights into existing DRG's is the next logical step in DRG refinement.

A more radical approach to nursing severity would be to consider the use of new patient characteristics which commonly correlate with nursing time. At present, five States use some form of case-mix measure in paying for institutional long-term care services

(Stassen *et al.*, 1983). These schemes all use both measures of dependency and some actual nursing procedures in determining allowable costs. If dependent patients prove to have the same cost-enhancing qualities within hospitals that they have within nursing homes, then a fairly simple "dependency modifier" might eventually be designed for application within all DRG's. Entirely new data elements would be needed for such a scale; elements such as those in Katz's Activities of Daily Living measure (Katz *et al.*, 1963). Research is needed to determine whether or not dependency measures would improve hospital payments sufficiently to warrant their inclusion on the discharge abstract.

Medical severity

In those DRG's where the limits of medical nomenclature have proved to be the source of DRG instability, a more medical solution is needed. One possibility would be the adoption of a scale or severity score based on a set of clear-cut, identifiable clinical parameters such as blood pressure, heart rate, and laboratory tests on admission. A recent report has demonstrated the ability of such a scale to predict the need for use of the special care unit (Wagner *et al.*, 1983). Further study is needed to determine whether such measures can also predict resource use and whether they would be applicable to patients not cared for in special units. The problem with such an approach is that new data elements will be needed on the discharge abstract. Determination of the accuracy of the underlying measurements could become a serious problem were the method to be used in a nationwide payment scheme. In the long run, however, the use of such elements to define severity is a logical direction for the modification of existing prospective payment methods.

More promising short-term approaches may be found in methods which make use only of UHDDS data. A recent report of the Disease Staging concept developed by Gonnella and colleagues (Gonnella *et al.*, 1984) modifies their methods so that the stage of illness can be determined using existing discharge abstracts. Previous reports had required direct use of the medical record. The authors have reported an association between the stage of diabetes mellitus and the average length of stay of 5,842 patients. Their stage one patients had the lowest length of stay and stage four the highest. Stages two and three were intermediate and closely resembled one another in their length of stay. Because the patients studied would have fallen into a variety of different DRG's, this study does not demonstrate whether staging can explain within-group variation in patients segregated by DRG category.

When Pettengill and Vertrees reviewed potential case-mix measures for Medicare use in 1982, they rejected Gonnella's staging concept because it produced an unworkable number of groups, was not exhaustive, and did not necessarily produce groups with homogeneous patterns of resource use (Pettengill *et al.*, 1982). As a potential modifier of unstable

DRG's, however, disease staging may be more promising, particularly now that it can be performed using only discharge data. Further work is needed to determine whether clinical staging can explain enough within-group variation to make modification of certain DRG's via staging useful. If applied only to those DRG's that have been demonstrated to be unstable, the resultant system could be kept both comprehensible and manageable.

Another conceptually similar approach would be to use the comorbidities present on the UHDDS, or on an expanded UHDDS, in a more sophisticated manner than is now the case. In DRG construction, comorbidities were dealt with as a yes/no phenomenon. Either secondary diagnoses affected the principal diagnosis or they did not. As a result, comorbidities with a small clinical effect increase the weight of a given DRG as much as do those with a major effect. The construction of a hierarchical scale of comorbidities would be more clinically realistic. Another approach would be the formation of clusters of comorbidities treated as though they were a single, highly weighted variable. The aim in each instance would be the same: to produce subclasses of currently unstable DRG's using only the UHDDS, or possibly an expanded UHDDS, as the source of the severity scale.

The staging method which is currently generating the most interest is that of Horn and Sharkey (Horn *et al.*, 1983). This approach is described in detail elsewhere in this issue. We doubt that their system, as it now stands, could be used for payment purposes because of the subjectivity involved in the judgments made about patients, and the relative vagueness with which these judgments are converted to a scale. Their index requires decisions by raters, based on implicit criteria, in areas such as the stage of the principal diagnosis and the rate of response to therapy. At present, both scoring and the translation to a four-point scale are done by raters who work directly with the hospital chart following discharge. The inclusion of such a system in a national payment scheme, which must be based on hospital-reported data, seems to us improbable; the same judgment was made in the Department of Health and Human Services report to Congress on case-mix reimbursement (Schweiker, 1982).

Summary and conclusions

Diagnosis-related groups have been in use for only 4 years as a single-State payment system in New Jersey and for less than a year as an element in all Medicare hospital payments. Their use as a national payment scheme provides a rigorous test of the grouping methods. We fully expect that such use will reveal unanticipated weaknesses in groupings and highlight some of the known problems in the system. Further, we anticipate that continued use of DRG's for payment will lead to an active process of modification and improvement of groupings. DRG's have evolved rapidly since they were first developed in the early 1970's; this evolution should not only continue but accelerate.

Effective debate about DRG modification, however, requires both precision in discussing the grouping methods and clarity about the parameters of the discussion. Debates about the Federal payment weights must be carefully segregated from debates about the underlying grouping methods. Low levels of payment for outliers, for example, may lead to a demand for a system in which few outliers exist. This demand relates to payment, not to the nature of DRG's. Similarly, a definition of outliers which forces the payment of certain true outliers as though they were group members makes some DRG's appear more unstable than they are. This is not a problem inherent in or peculiar to DRG's; it is a payment problem, deserving of debate as such.

As we have indicated, discussions of severity issues are often confused by the inclusion of erroneous cases, by an assumption that the elimination of outliers is a desirable goal, or by a firm belief that increased sickness will of necessity correlate with increased cost. High levels of sickness and high levels of cost are not necessarily the same thing; an effective severity measure may well identify certain less costly subgroups of the severely ill.

With all these caveats in mind, the development of severity measures, or of some severity proxy, to reduce within-group variation in certain DRG's is an important step in DRG modification. By far, the most promising near-term approaches are those that make use of UHDDS information or of an expanded list of standard diagnostic and procedural information. The addition of even a single new element to the data set would require both extensive field testing and a long lead time. If some hospitals are seriously affected by within-group variation, then relief within the next few years is dependent on the development of modifications in grouping that are based on the same data elements as DRG's themselves.

References

- Fetter, R. B., Shin, Y., Freeman, J. L., Averill, R. F., and Thompson, J. D.: Case mix definition by diagnosis-related groups. *Med Care* (Supplement) 18(2):1-53, 1980.
- Gonnella, J. S., Hornbrook, M. C., and Louis, D. Z.: Staging of disease: A case-mix measurement. *Journal of the American Medical Association* 251:637-644, 1984.
- Health Systems Management Group, Yale School of Organization and Management. The new ICD-9-CM diagnosis-related group classification scheme, Final Report, May 1982.
- Horn, S. D. and Sharkey, P. D.: Measuring severity of illness to predict patient resource use within DRGs. *Inquiry* 20:314-321, 1983.
- Institute of Medicine. Reliability of Medicare hospital discharge records. National Academy of Sciences, NTIS No. PB281680, Nov. 1977.
- Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A. and Jaffee, M. W.: Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association* 185:94-100, 1963.

Killip, T. and Kimball, J. T.: Treatment of myocardial infarction in a coronary care unit. A two year experience with 250 patients. *American Journal of Cardiology* 20:457, 1967.

McMahon, L. F.: Diagnosis-related group prospective payment: Effects on medical quality assurance. *Evaluation and the Health Professions* 7(1):25-41, 1984.

Mills, R. E., Fetter, R. B., Riedel, D. C. and Averill, R. F.: AUTOGRP: An interactive computer system for the analysis of health care data. *Med Care* 14(7):603-615, 1976.

Pettengill, J. and Vertrees, J.: Reliability and validity in hospital case-mix measurement. *Health Care Financing Review*. HCFA Pub. No. 03149 Office of Research and Demonstrations, Health Care Financing Administration, Washington. U.S. Government Printing Office, Dec. 1982.

Schweiker, R. S.: Report to Congress, Hospital Prospective Payment for Medicare. Dec. 1982.

Smits, H. L. and Watson, R. E.: Do DRG payments regulate the practice of surgery? *New England Journal of Medicine*. To be published.

Stassen, M. and Bishop, C. E. Incorporating case mix in prospective reimbursement for skilled nursing facility care under Medicare: Critical review of relevant research. Center for Health Policy Analysis and Research, Brandeis University, 1983.

Thompson, J. D., Fetter, R. B., and Shin, Y.: One strategy for controlling costs in university teaching hospitals. *Journal of Medical Education* 58:167-175, 1978.

Wagner, D. P., Knaus, W. A. and Draper, E. A.: Statistical validation of a severity of illness measure. *American Journal of Public Health* 73(8):878-884, 1983.