

A research paradigm for severity of illness: Issues for the diagnosis-related group system

by Paul M. Gertman and Steven Lowenstein

The new Medicare Prospective Payment System has been challenged with regard to its fairness in reimbursing hospitals adequately, given the true resource needs in caring for patients. Most of these criticisms are now labelled as issues about adjustments for severity of illness. Critics point to the large amount of unexplained variation in charges and length of stay within the existing DRG's as indirect support for their contentions about inadequate adjustments.

A paradigm is presented which argues that the key questions on the types of severity of illness measures to be utilized in future refinements of DRG's revolve around the extent and type of data which can feasibly be included in any workable reimbursement approach. A paradigm is presented on how these questions about information define a series of research options in the severity of illness arena.

Introduction

The new DRG-based prospective payment system (PPS) for Medicare is a revolutionary concept in health care finance. It introduces a new approach to reimbursement, which is based upon a defined "clinical product." In contrast, under retrospective cost-based reimbursement, each component of the process of care (each visit, x-ray, procedure, and day of care) was paid for separately. That open-ended payment system has proved too inflationary; Federal outlays for Medicare have increased from \$1.1 billion in 1966 to \$51 billion in 1982, a rate of inflation which has outpaced the gross national product, the Consumer Price Index, the median family income, and other benchmarks of the economy's growth.

Under the prospective payment system there will eventually be a National market price for each clinical product. The shifts in the Medicare system should stimulate major efficiencies, as providers of care begin to consider which components should be introduced into the clinical production process, and which should be rejected. It is expected that there will be a major break in the upward-spiraling overutilization of services which produce no improvement in clinical care. It is hoped that the DRG system will create incentives for physicians and hospitals to shorten hospital lengths of stay and incentives to decrease resource consumption per episode of illness. However, considerable debate has arisen over whether the current DRG formulations in the prospective payment system represent the best approach for achieving these objectives.

With tens of billions of dollars at stake each year, the Health Care Financing Administration (HCFA) will face innumerable challenges to the design and administration of the prospective payment system. These challenges are likely to encompass a broad spectrum: on one hand, there may arise overt challenges to the equity of the DRG's, particularly with respect to their ability to adjust fully for differences in severity of illness. On the other hand, there is the

possibility of covert challenges, in particular, those which involve clinical opportunism. For example, under the PPS, providers may seek to: 1) admit larger numbers of patients, especially patients with easily treated illnesses and short anticipated lengths of stay; 2) split illnesses into two parts, in order to spread a patient's care over two hospital admissions; 3) unbundle diagnostic procedures, shifting some to the ambulatory setting (outside the PPS); 4) upgrade primary and secondary diagnostic codes, in order to obtain a higher-paying DRG assignment; 5) perform more complex surgical procedures to inflate the DRG (procedure inflation); and 6) prolong the hospital stay of patients with lingering illnesses, as the outlier trim point is approached. When a patient has more than one clinical problem, there arises the possibility of gaming the data about the correct principal diagnosis; there also arises the possibility of outright fraudulent representations.

HCFA can respond to these challenges by improvements in design of the basic DRG formulation and administrative rules and procedures. It is likely that changes in DRG structure and design can best deal with the equity and overt challenges to the system; administrative controls will likely have their greatest strength in dealing with the covert challenges.

A cornerstone of the new prospective payment system for Medicare is the use of a diagnosis-related group case classification system. The DRG system employs existing discharge abstract data on the patient's principal diagnosis, secondary diagnoses, surgical procedures, age, sex, and discharge status, in order to classify patients into different groups that are "clinically coherent and homogenous with respect to resource use" (*Federal Register*, Sept. 2, 1983).

The current DRG's have been criticized for failing to account for severity of illness. Controversies about severity are based upon the assumption that, within each DRG, there are sicker patients who will cost more. Hospital administrators and physicians have expressed concern that they will be unfairly penalized if their patients are sicker and consume greater resources. Medicare administrators have concerns of their own: To the extent that providers can measure

Reprint requests: Paul Gertman, The Health Data Institute, 7 Wells Ave., Newton, Mass. 02159.

or anticipate patient severity, they may reject patients with severe illnesses in order to maximize their profit. Among policy analysts and economists, severity has become the buzzword to describe the sources of otherwise unexplained variations in resource use within the DRG's.

It is the authors' opinion that the objectives of a system for defining severity of illness under DRG's (with respect to Medicare reimbursement) should be threefold: 1) to place patients into categories which are homogeneous with respect to *resource need*; 2) to categorize patients in such a way that manipulations of data are minimized; and 3) to avoid producing distortions of provider behavior that could adversely affect the outcome of patient care. (For example, with respect to the third objective, HCFA has refrained from using death as a classification factor in most DRG's. Unless this is a common outcome—e.g., in heart attacks—HCFA does not want to reward hospitals for letting patients die.)

The distinction between resource use and resource need is critical. One of the major assumptions by many critics of the DRG system, particularly those who point to large variations in resource use within DRG's, is that the major source of this variation is the absence of equitable severity of illness measures. However, there are other clinically-related sources of variation. These may be responsible for a substantial portion (or even the majority) of the variation in costs and length-of-stays observed to date in DRG's and may be even more important as covert challenges to restraining Medicare costs. For example: 1) situations exist where there are no clear rules for determining an individual's correct principal diagnosis or the validity of the presence of a complication; 2) existing practice patterns frequently involve substantial proportions of medically inappropriate use, and 3) clinical practice patterns exist where definition of an appropriate episode of medical care can be split to produce two or more reimbursable episodes.

Stated another way, variation in resource use within DRG's may be due not only to inadequacies of measurements and adjustments, but also to the above types of problems. In this article, these three basic sources of confounding will be reviewed, and the measurement of severity of illness will be discussed, focusing on an applied research paradigm for testing new methods of incorporating such measurement approaches into the Medicare prospective payment system.

Clinical data definitions

The major factor which determines the DRG assignment, and thus determines the payment of the provider, is the physician's assessment of the patient's diagnosis. The accuracy and completeness of the diagnostic and procedure data are fundamental to the DRG system. Regulations require that physicians record all diagnostic and procedures data on the face

sheet in an honest and responsible manner. Peer review organizations are charged with monitoring the validity of diagnostic information provided by the hospital for the purposes of payment. However, HCFA has accumulated evidence suggesting serious errors in the description and coding of the principal diagnoses and principal procedures. According to studies by the Institute of Medicine and the Department of Management and Research of the University Hospitals of Cleveland, errors and discrepancies in the listing of the principal diagnoses range from 20-40 percent (Demlo *et al.*, 1978, Doremus and Mickenzi, 1983). Much of this error was probably due to sloppiness and to use of unqualified record abstractors. These errors represent a threat to the validity of all analytic studies employing historical diagnostic data, and almost certainly contribute to the variance found in some previous DRG analyses. The extent of the sloppiness problem should, however, rapidly diminish under pressure from concerned administrators and from HCFA oversight through Professional Review Organizations (PRO's) and fiscal intermediaries.

A more fundamental problem is illness definition. Creation of each DRG category starts with consideration of the patient's primary diagnosis and the performance or nonperformance of an operating room surgical procedure. A substantial number of the DRG categories are then further divided, depending upon the presence or absence of significant complications or comorbid conditions.

Comorbidity is defined as an underlying condition that existed before the patient was hospitalized. A complication is a condition other than the principal diagnosis which occurs after the onset of the principal diagnosis (and usually after the patient is admitted to the hospital). Many (if not most) Medicare patients admitted to the hospital have one or more comorbid or complicating conditions in addition to their primary diagnosis. In the planning of the DRG system, it was reasoned that the length of stay and the consumption of resources (and, of course, reasonable reimbursement) are likely to be higher if a substantial complication or comorbidity is present. Therefore, specific lists of such conditions have been developed, based upon the judgment of the physician teams who assisted the DRG design team at Yale. A "substantial comorbidity or complication" is defined as any secondary diagnosis likely to prolong the hospital stay of at least 75 percent of patients by one day or more (Yale University, 1982). One example of a substantial comorbid-complicating condition is *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) code 202.98 (Hodgkin's disease). Another example is ICD-9-CM code 250.50 (diabetes with ophthalmologic manifestations). If one or more of these is present and is listed on the discharge data abstract sheet as a secondary diagnosis, the patient's care needs are considered more resource-intensive. Therefore, he or she is usually placed in a

DRG category for which the hospital is awarded a higher reimbursement for the episode of illness.¹

The concept of a complicating condition or a comorbid condition requiring additional resources has validity and should contribute to a reduction in variance among DRG categories. At the same time, operationalization of the concept in a prospective payment system raises some serious medical and definitional problems. One obvious problem is that the source of the information used to classify each patient into a DRG is the face sheet of the medical record. The accuracy and completeness of the face sheet are of utmost importance. For example, the presence or absence of a substantial complication or comorbid condition can only be determined from the list of all secondary diagnoses on the face sheet. However, as noted earlier, there is considerable error in hospital discharge data.

Another serious problem is whether rewarding hospitals for more complications is a fundamentally sound approach. If a comorbid condition is present upon admission, its inclusion as a factor discriminating between types of patients may be valid. For complications which arise after admission, however, their inclusion may reward less careful practitioners, while penalizing those who make investments to prevent such untoward events. Might a hospital which reduces its infection control procedures benefit if its infection rates subsequently rose (thus sending additional patients into higher reimbursement DRG categories)? The answer, unfortunately, may be yes. The current DRG classification system, however, does not distinguish a comorbid problem present at the time of admission from a complicating condition which occurs subsequently. This dichotomy, although conceptually attractive, would be extremely difficult to operationalize; identification of many comorbid conditions at admission might require a series of investigational procedures, which are not normally done and which might involve additional cost and risk for patients.

An even more difficult problem is how to define and validate the actual presence of a substantial comorbidity or complication. For example, diabetes mellitus with renal manifestations is one of the DRG substantial comorbid/complicating conditions. However, it is not clear when diabetes mellitus should be considered as substantial. Does the presence of an abnormal glucose tolerance test and the finding of trace proteinuria constitute ICD-9-CM 250.4 for the purposes of reimbursement? Similarly, ICD-9-CM 780.0, the diagnosis of coma (including stupor, drowsiness and somnolence) is a substantial comorbid/complicating condition. How is drowsiness validated by the PRO? How is respiratory failure (ICD-9-CM 799.1) defined? Is it by discovering a low arterial blood oxygen concentration? If so, how low?

¹A large number of the DRG's include the variable "age >70 and/or c/c." The translation is "age greater than seventy and/or the presence of any substantial comorbidity or complication." If this variable is present, the patient is in one DRG; if the variable is absent, a lower DRG is assigned. The variable is dichotomous.

How is hepatitis (ICD-9-573.3) proven? Is it enough to have a single abnormal liver function test?

Another example involves the question of sufficient evidence to indicate a postoperative wound infection. Would the presence of a positive culture from the wound site for any organism be sufficient? Would it depend upon the colony count of infecting organisms or the types of organisms found? Would it require that the infection be significant to the point where fever occurs or antibiotic therapy is instituted? If the lowest level of this hierarchy of definitions is chosen, every patient will have their wound site cultured, and many with clinically insignificant bacterial organisms (many of which are normally present on the skin) will suddenly start to become included in the complicated condition category. If sufficient evidence is antibiotic treatment, we are likely to see many elderly patients (who are already at higher risk of adverse drug reactions) having such agents administered more frequently, because of the bias introduced by the DRG reimbursement system.

As long as the definitions of comorbid and complicating conditions remain vague, clinicians and data abstracters have an incentive to include larger numbers of such conditions on the data abstract sheet, whether or not such conditions are clinically relevant. Indeed, most of the current DRG scheme makes no distinction between complicating/comorbid conditions related and not related to principal diagnosis. Related comorbidities or complications may contribute more (or less) to resource consumption than unrelated conditions (Louis and Heineccius, 1983).

It should be noted that these definitional issues plague not only the assessment of complicating or comorbid conditions, but also affect the determination of the principal diagnosis. For example, when does a patient with peptic ulcer disease fit into DRG 175 (gastrointestinal hemorrhage, age less than seventy without c/c), rather than into one of the other peptic ulcer categories (DRG 176 or DRG 178)? Is hemorrhage evidenced by a positive stool guaiac test for blood? Is it, instead, any aspiration of hematest-positive material through a nasogastric tube (and if so, does it count any fresh bleeding which may have been induced by insertion of the nasogastric tube)? Or is "hemorrhage" only frank vomiting of blood? Adjacent DRG categories are frequently separated only by such terms as "hemorrhage" or "complication." Different parties are likely to define such terms differently, depending upon economic and other pressures. Obviously, many of the adjacent DRG's require more precise separation and definition, for they rank among the high cost diagnoses for Medicare patients. The boundaries separating many such DRG's are indistinct and, potentially, malleable. In Maryland, for example, the mean charge for DRG 140 (angina) was \$2,409 in 1981 versus a mean charge for DRG 143 (chest pain) of \$1,925. There are no well-accepted rules delineating the correct assignment of a patient to one or the other of these categories. Thus, the incentives to label a case as angina, with a 25 percent greater reimbursement, will be substantial. There will

still remain, however, a more serious basic issue—how to relate a patient's clinical reasons for hospitalization to the ICD-9-CM coding system used in DRG groupings.

The *International Classification of Diseases*, now in its ninth revision, is still based on an anatomic approach to disease developed in the late nineteenth century by the great French forensic pathologist, Dr. Jacques Bertillon. This system, despite its decennial updating and an American clinically modified version, was never designed to deal with clinical severity of illness. It is still not well-suited to this task. Nor was the ICD-9-CM coding structure even intended to accommodate a reimbursement system or withstand the pressures for precision that this requires. ICD-9-CM's central flaw is that there are no clear rules for determining when a patient has a particular problem which fits an ICD-9-CM code. Everyone in the past has used the "common wisdom" approach that avoids precise operational definition.

While a PRO policing effort may prevent massive definitional abuse, the pressing need is for a research and development program to develop at least consensus guidelines and criteria for use of ICD-9-CM terms. Where there are objectively verifiable criteria, these should be disseminated by HCFA throughout the hospital and PRO community Nationwide. Where there are no such criteria, definitional guidelines can be generated through consensus conferences (such as the National Institutes of Health utilizes), opinion surveys of specialists and generalists, and other such means. Certainly, the medical community will continue to be plagued with serious definitionally-induced variance unless such steps are initiated.

Appropriateness of utilization

In the current health system patients are often admitted to the hospital for problems such as low back pain, which can be treated at home, and cancer chemotherapy, which may be safely administered on an outpatient basis. If these current patterns of unnecessary admissions—those not *severe* enough to warrant hospital admission—are allowed to continue, Medicare will continue to spend billions of dollars unnecessarily.

Many of the most common DRG's, particularly those not associated with surgery, involve conditions with a wide spectrum of clinical manifestations and needs for acute care. The decision whether or not patients in these categories require hospitalization instead of outpatient treatment involves a set of physiologic and clinical assessments which are often not captured by the diagnostic, procedure, and demographic factors used to designate DRG's. For example, the category heart failure (DRG 127) might include a patient with such excess fluid load that he or she was having severe difficulty breathing and was in need of aggressive therapy in a hospital environment. However, this DRG category would also apply to a patient with long-standing heart disease who had gained some extra fluid weight as a result of dietary

indiscretion and was noticing a mild increase in shortness of breath when walking a short distance. The latter case would not require hospitalization in most circumstances, but with the DRG system there might be an incentive to admit such a patient for several days of therapy for the patient's convenience and for the financial benefit of the provider. This situation involves no manipulation of the diagnosis as suggested by the term "DRG creep," but rather involves playing on a central weak point of the DRG system—a lack of severity of illness standards which distinguish inappropriate hospital admissions. Other high frequency categories such as DRG 182 (gastroenteritis and miscellaneous digestive diseases) and DRG 132 (arteriosclerosis in patients over 70) involve the same potential problem of classification of disease by diagnosis, by procedure, and by age which could run the gamut from life-threatening acute situations to chronic, stable manifestations which can be effectively treated on an ambulatory basis. While it may be fraudulent to deliberately change the diagnosis of a patient so as to gain higher DRG reimbursements, there are no accepted legal or procedural barriers to physicians lowering their severity standards for admitting patients within these medical DRG categories.

A study of over 8,000 Medicare and non-Medicare patients hospitalized in 41 Massachusetts hospitals in 1973 and 1978 found that almost two out of every five days for medical patients with a length of stay (LOS) under ten days were inappropriate—a rate higher than that of medical patients staying longer than ten days (Gertman *et al.*, 1982). Studies done by the Boston University School of Medicine indicated that in 1980, among selected Professional Standards Review Organizations, levels of inappropriate hospital admissions among Medicare beneficiaries ranged from a low of 12 percent to over 31 percent (Restuccia *et al.*, 1984). Both of these studies employed the Appropriateness Evaluation Protocol methodology developed for HCFA (Gertman and Restuccia, 1981). Part of the mysterious variation in days of care per 1,000 Medicare beneficiaries across the country may represent variation in inappropriate admissions. Assessments by SysteMetrics Incorporated, using their standardized MedReview instrument (SMI) found that a smaller but still substantial amount of Medicare admissions were inappropriate and that levels varied geographically (SysteMetrics, Inc., January 1983). Given that neither of these instruments challenged the necessity of elective surgery (i.e., was the hysterectomy indicated), these must be conservative estimates. The true levels of unnecessary admissions may be even higher.

Within episodes of illness, substantial resource use variation may be due to inappropriate ancillary service ordering. The joint Massachusetts Blue Cross-Massachusetts Hospital Association's Ancillary Services Review Program has documented average levels of inappropriate laboratory tests, electrocardiograms and respiratory therapy in the 20 percent to 30 percent range (Hughes *et al.*, 1984). After adjustments for

certain severity of illness factors and for pricing differences, the dollar amount of inappropriate use of an ancillary service for a specific diagnosis varied over tenfold from the highest to the lowest hospital.

These data suggest that a major source of variation among institutions for any given diagnostic grouping may not be due to inadequacies in the grouping method, but rather, largely due to inappropriate admissions, inappropriate days of care and inappropriate ancillary utilization. This possibility can be addressed by a major program of research on the correlation of unexplained DRG variation with measures of inappropriate use. If this, along with the other basic factors, is the principal source of variation, then the current DRG system may, in the near term, need only modest refinement rather than a dramatic overhaul.

The episode of illness

The DRG case-mix classification system defines the relevant unit of service for reimbursement as the illness episode, which lasts from the time of admission to the hospital until discharge. The greater the number of admissions a hospital has for a fixed number of bed days and available staff, the greater its revenue will be. This could create incentives for shortening hospital lengths of stay and decreasing consumption of resource *per episode*.

For some conditions a medical episode of illness has a clear beginning and end point which correspond fairly closely with initial admission to the hospital and subsequent discharge. For many conditions, however, the potential exists for deliberate abuse of the admission payment concept, by artificially dividing an episode of illness into two hospitalizations. In other words, patients may be admitted for several days or weeks, then discharged, and later readmitted for a procedure or for continued medical treatment. The Medicare regulations recognize the potential for such deliberate abuses, and the admission pattern monitoring program is specifically charged with looking for such attempts. Unfortunately, for some conditions, the proper end points of hospital care are less clearly defined. This represents a third major source of confounding in research on severity of illness.

The single admission-to-discharge concept remains most problematic in the following two areas. The first area involves medically-acceptable splits. There are a substantial number of common conditions where medical opinion may differ and individual physician judgment is used to determine whether a patient's course of treatment should be completed during one or more hospitalizations. A simple example is acute cholecystitis due to gallstones. Some physicians believe that patients should first be admitted for medical treatment, which includes fluids, antibiotics and nasogastric suction. Patients are then discharged and readmitted several weeks later for elective cholecystectomy. Other physicians and surgeons believe it is safe and proper to operate on the patient who is improving

during a single hospitalization. At the present time, no medical review panel could fairly penalize competent physicians for choosing one course over the other. Similarly, a patient hospitalized with symptoms of coronary insufficiency may undergo cardiac catheterization during his hospitalization and be discharged for a period of recovery and risk-factor modification (e.g., stopping smoking) prior to coronary bypass surgery; alternatively, these procedures might occur in the same admission. Another example is the procedure-linked diagnosis of benign prostatic hypertrophy; patients may be admitted with symptoms of urinary obstruction and then scheduled for prostatectomy either during the same hospitalization or at a subsequent admission in the near future.

To illustrate this issue, the authors reviewed the top 100 DRG's and identified 32 examples of medically acceptable situations where, in their opinion, multiple procedure-linked admissions might occur; these are listed in Table 1. In each case, for some physicians, it might seem preferable to complete the medical treatment, the diagnostic evaluation and the operative therapy in a single hospitalization. However, depending upon the habits and style of practice of the physicians in charge and depending upon the economic incentives which are allowed to come to the fore, a single process of diagnosis and treatment might be split into two care episodes. There are usually no well established guidelines to determine whether split admissions are appropriate for a given clinical problem. The older the patient and the more complicated his medical care, the more often physicians may decide to partition care into two or more hospitalizations.

The second problematic area related to illness episode involves chronic diseases requiring episodic hospital care. In managing many illnesses, no curative therapy is administered. Rather, symptomatic or palliative therapy is given. Such episodes of illness are very difficult to monitor or regulate under the DRG system. Particular areas of difficulty are psychiatric illnesses such as alcoholism. If, after seven days, an alcoholic patient signs out of the hospital against medical advice, how hard should the hospital or physician fight to convince him to stay? Similarly, if a psychiatric patient is almost well, a potentially inappropriate incentive exists to allow him to return home too early. In such cases, there is a high probability the patient will return in the days or weeks to come. Indeed, there are no clearly defined lengths of stay for inpatient treatment of alcoholism and psychiatric illnesses.

Other examples are patients with heart failure and patients with chronic respiratory disease. Such patients are usually treated to the point where they are able to function outside of the hospital environment. The more vigorously the acute episode of heart failure or chronic lung disease is treated, the less likely that the patient will be readmitted in the near future. A study conducted at University Hospital and Boston City Hospital demonstrated that readmissions for

Table 1
Procedure-linked split admissions: Possible problem diagnosis-related groups (DRG's)¹

Category	DRG	Name	Second Admission DRG	Name
Nervous	009	Spinal disorder and injury	004	Spinal procedure
Ear, nose and throat	072	Nasal trauma/deformity	056	Rhinoplasty
	070	Otitis media/and upper respiratory infection	062	Myringotomy
Respiratory	082	Respiratory neoplasm	077	Operating room procedure on respiratory system
	092	Interstitial lung disease	077	Operating room procedure on respiratory system
Circulatory	130	Peripheral vascular disorder	110	Major reconstructive vascular procedure
	135	Cardiac congenital valvular disease	105	Cardiac valve procedure with pump
	140	Angina pectoris	106	Coronary bypass with cardiac catheterization
	125	Circulatory disorder with cardiac catheterization	107	Coronary bypass with cardiac catheterization
	141	Syncope and collapse	116	Permanent cardiac pacemaker implant
	138	Cardiac arrhythmia/conduction disorder	116	Permanent cardiac pacemaker implant
	172	Digestive malignancy	149	Major small and large bowel procedure
Digestive	182	Esophagitis/gastroenteritis/miscellaneous	149	Major small and large bowel procedure
	180	Gastrointestinal obstruction	150	Adnesiolysis
	174	Gastrointestinal bleeding	157	Anal procedures
Hepatic/biliary	203	Malignancy of helato biliaryruben or pancreas	191	Major pancreas, liver or shunt procedure
	204	Disorder of pancreas except malignancy	191	Major pancreas procedure
	207	Disorder of biliary tract	195	Total cholecystectomy
	204	Disorder of pancreas except malignancy	200	Hepatobiliary diagnostic procedure
Musculo/skeletal	243	Medical back problem	215	Back and neck procedure
	271	Skin ulcers	263 110	Skin graft or vascular major reconstructive procedure
Endocrine/metabolic	272	Major skin disorder	267	Perianal/pilonidal procedure
	294	Diabetes age greater than 36	287	Wound debridement/skin graft for metabolic disorder
Kidney and urinary tract	300	Endocrine disorder	290	Thyroid procedure
	318	Kidney and urinary tract neoplasm	303	Major bladder procedure for malignancy
	325	Kidney/urinary tract signs and symptoms	306	Prostatectomy
Male/reproductive	346	Reproductive malignancy	334	Major pelvic procedure
Female/reproductive	348	Benign prostatic hypertrophy	336	Prostatectomy
	366	Malignancy, female reproductive system	357	Uterus/adnexa procedure for malignancy
Myeloproliferative/blood	397	Coagulation disorder	392	Splenectomy
	403	Lymphoma/leukemia	400	Lymphoma or leukemia with major operating room procedure
Infection/parasitic	419	Fever of unknown origin age greater than 70	415	Operating room procedure for infectious disease

¹These "split-admission" problem DRG's among the top 100 DRG's are provided to illustrate an issue. They are not a definitive categorization based on empirical research.

heart failure were inversely related to the vigor with which excess fluid was removed during the hospital stay (Gertman and Stanton, 1975). Patients who were discharged early with persisting but modest signs of fluid accumulation had a higher rate of readmission in the ensuing year. Yet, examining only an individual case after the fact, a medical panel could not have successfully proven undertreatment of any single patient.

Relatively little research has been done, even at a descriptive level, on patterns of care for older individuals over time; this deficiency is particularly great for those with multiple active chronic health problems. Prior use, as discussed later, may in part explain subsequent hospital admission resource use. More important are concerns about whether breaking care into separate admissions may enhance outcomes.

How patterns of care affect resource use and outcomes is a topic which requires extensive research. At the least, such research is necessary to minimize confounding efforts to properly adjust for severity of illness; beyond this, any study of DRG impacts on quality of care must develop such information.

Severity of illness analysis methods

There is considerable concern among providers about whether the current DRG system optimally achieves the stated objectives of the Medicare prospective payment system. Specifically, many fear that the PPS may not equitably deal with systematic differences in severity of patient illnesses across all hospitals. As much as one might wish that 467 DRG's controlled for interhospital differences in patient severity, the large amount of unexplained variance in some DRG's is disturbing. Of course, much of it may be simply *random* variation, with zero net effect across all admissions. On the other hand, there may be systematic variations by one or more hospital characteristics (e.g., teaching status, ownership, location, etc.), resulting in windfall gains or unfair losses. Inevitably, however, an inaccurately defined output leads to inequities and, what is worse, gaming of the system.

While all parties to this debate over the equity of DRG's generally understand the term "severity" and would agree that in any group of patients there are those with "more severe" and "less severe" illnesses, there is no agreement on how to:

- Quantify differences in severity on a continuous scale of measurement;
- Translate such severity measures into uniform resource need measures; and
- Fairly monetize the resource measures in Medicare reimbursement procedures.

In fact, the authors would argue that there is no fully valid way to accomplish any one of these three measurement tasks. "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. The best one can do is make approximations of the concept which reduce

the extent of disagreement. Recognizing that, even if one could remove all confounding factors, no perfect severity measure exists, HCFA and health care providers must move forward to make the best approximations possible subject to a host of practical constraints.

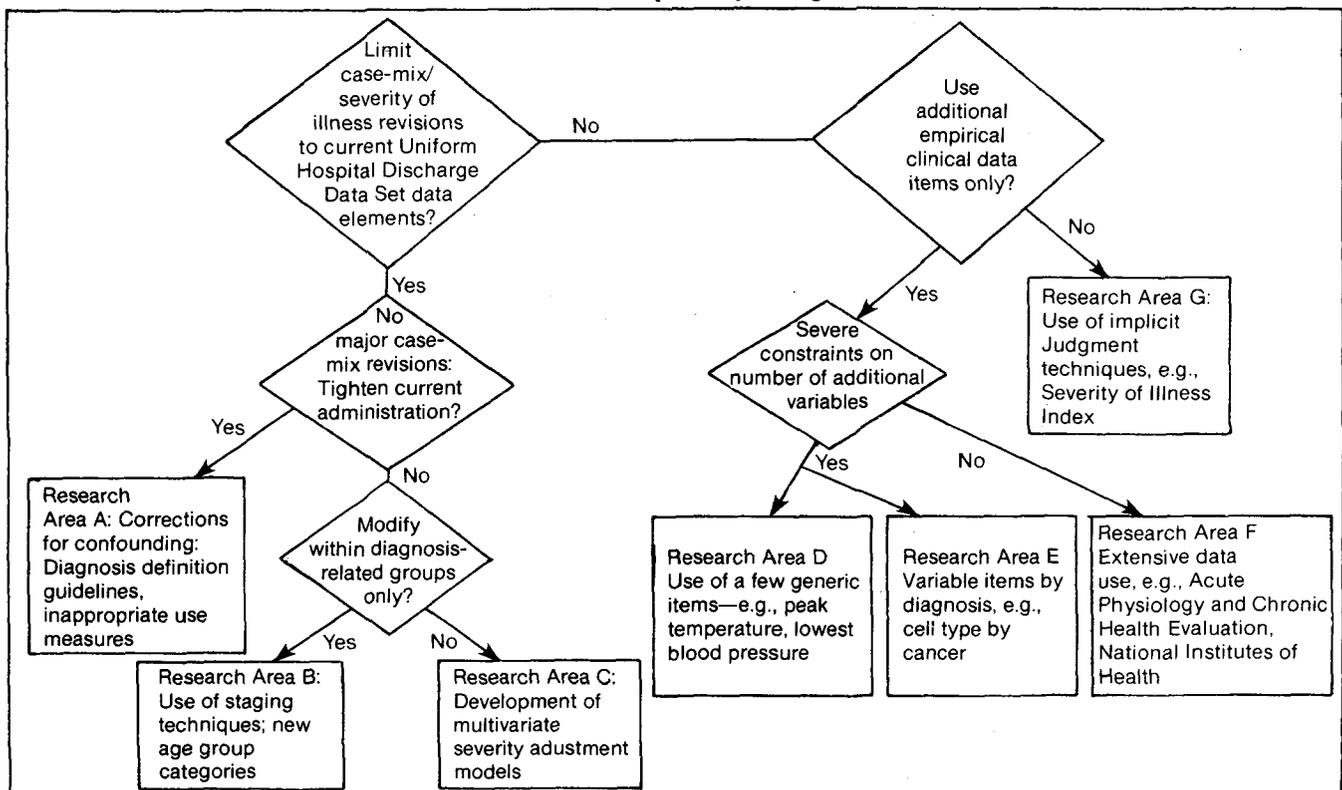
The old DRG system devised by the Yale University group was subjected to extensive methodological criticism for not adequately dealing with major differences in severity among patients. Several research organizations, including Blue Cross of Western Pennsylvania, Systemetrics, Inc., Susan Horn's team at the Johns Hopkins School of Public Health and others have demonstrated that they can achieve superior reductions in length of stay variance by alternative methods compared to the DRG's used in the old system (Young, 1979; Garg *et al.*, 1978; Horn, 1983). For example, Dr. Horn has compared four existing classifications systems (ICD-9-CM DRG's, the old New Jersey DRG's, disease staging, and generalized patient management paths) with her severity of illness index for their respective variance reduction characteristics on a disease-specific basis across different hospitals. Generally, Horn's chart abstract method achieves far greater subgroup homogeneity (total charges and length of stay) than do any of the other computerized, discharge abstract-based classification systems. Horn and her co-workers have also tested whether their approach is superior to the new Medicare DRG system which has attempted to address some of the prior criticisms on severity adjustment. In a study of ten new DRG's at four hospitals, they have reportedly been able to achieve better than 40 percent additional reduction in variance by their severity of illness index (Horn *et al.*, 1983).

The issue for HCFA is not whether to simply stand pat with the DRG system as currently defined versus attempting to identify possible improvements. Rather, the issue is: What is the most cost effective way to proceed in enhancing the DRG's. While basic research must proceed on the sources of confounding, such as definitional imprecision, inappropriate use and variance in practice patterns, efforts to explore alternative severity of illness adjustments in the Medicare PPS must move forward as well.

To assist in illustrating how a DRG/severity of illness policy research program could be organized, Figure 1 shows a decision tree model for questions which might be explored in the PPS severity of illness revisions. The authors believe that *the central issue structuring a practical severity of illness research paradigm is the amount and type of information available.*

The most important practical question to address is whether to limit enhancements of the case-mix method to those measures available only in the Uniform Hospital Discharge Data Set (UHDDS). While some might argue that this is only a procedural constraint and should not have such an important place in the research agenda paradigm, the availability of information and the logistics of acquiring that information is in fact the most critical decision that Medicare must address now if it wants to have an operationally revised system in the near future.

Figure 1
Severity of illness revisions:
A research options paradigm



Within the current information set available to Medicare, the next question which must be addressed is whether the focus of revisions should be to fine-tune the existing DRG system or to alter it? The former course leads to research option area A; here the principal research work would be to evaluate better trim points, test use of dollar (cost) dependent variables more extensively than was possible in the initial DRG design, structure definitional requirements more tightly to prevent abuse, develop a system of potential correction factors for inappropriate admissions, etc.

Alternatively, under research option area B, HCFA could decide to modify the DRG's by creating additional DRG subcategories. An example of this is the current pilot study by Systemetrics, Inc. applying the primary staged conditions and staging levels to selectively create subcategories where there is a clear need for reduction in currently large variances in costs per admission (Systemetrics, Inc., September 1983). Revisions might also be designed that anticipate issues of major quantitative importance in the near future. An example of the latter would be DRG 209, major joint replacement. A pioneering institution in the field, Brigham and Women's Hospital in Boston, has now started to extensively perform multiple major joint replacements in a single hospital admission episode rather than having two or more admissions. In their own internal, DRG-oriented management information system, they have now subdivided DRG 209

into two groups based on whether there is a single major joint replacement or multiple major joint replacements. This is an important, common surgical procedure where cost-effective and quality-of-care advances in surgical practice, involving multiple joint replacements per admission, are likely to disseminate fairly rapidly over the next several years. Thus, this DRG might be a priority candidate for potential revision.

Research option area C, using the current UHDDS dataset, would set aside the current Yale DRG framework and evaluate completely different ways of defining severity of illness differentials between institutions for purposes of reimbursement. These research options might involve use of multivariate quantitative models as opposed to a step-wise classification methodology to determine severity-adjusted reimbursement formulas. Carol Fernow at Health Care Systems International, for example, has taken Commission on Professional and Hospital Activity (CPHA) data and developed log-linear models which apply specific quantitative coefficients to factors such as specific secondary diagnoses (e.g., diabetes as a secondary diagnosis in cholelithiasis), types of secondary operation procedures, etc., to compute case-mix intensity and expected lengths of stay (Fernow, 1983).

The potential flexibility of a multivariate model based on Medicare UHDDS data sets is shown in Tables 2, 3, and 4 which were part of a teaching example developed at the Boston University School of

Table 2
Heart attack patients discharged alive: Hypothetical hospital claims file data

Obs	Hospital	DRG	Age	Sex	Non-OR procedure	Pacemaker	History	Total charges
1	A	115	64	M	No	Yes	No	\$ 9,976
2	A	115	58	M	No	Yes	No	9,721
3	A	121	52	F	Yes	No	Yes	9,830
4	A	121	60	M	Yes	No	No	8,965
5	A	121	47	F	No	No	Yes	7,780
6	A	122	40	M	No	No	No	6,493
7	A	122	45	F	Yes	No	Yes	9,555
8	A	122	52	M	No	No	No	6,982
9	A	122	58	F	No	No	No	7,320
10	B	115	60	F	No	Yes	Yes	10,441
11	B	115	51	M	No	Yes	No	8,988
12	B	121	40	M	Yes	No	Yes	7,765
13	B	121	57	M	No	No	No	6,690
14	B	122	52	F	Yes	No	Yes	9,388
15	B	122	63	F	No	No	No	6,997
16	B	122	62	F	No	No	No	7,012
17	B	122	61	M	No	No	No	6,845
18	B	122	58	F	Yes	No	No	8,588
19	C	115	62	F	Yes	Yes	Yes	12,404
20	C	115	60	F	No	Yes	Yes	10,213
21	C	121	55	M	Yes	No	Yes	9,154
22	C	121	57	M	Yes	No	No	8,130
23	C	122	48	F	Yes	No	Yes	8,870
24	C	122	58	M	No	No	Yes	7,501
25	C	122	62	F	No	No	Yes	7,682
26	C	122	64	M	No	No	Yes	7,664
27	C	122	45	F	No	No	No	6,075

NOTES:

Non-OR = Any non-operating room surgical procedure billed (e.g., Swan-Ganz catheterization, endoscopy, etc.)
 History = Prior admission in the past twelve months for a heart attack.

Obs = Patient observation number.

DRG = Diagnosis-related groups

Table 3
Model of heart attack charges per admission
(N = 27)

Factors	Model A diagnosis-related groups (DRG)		Model B (Multivariate)	
	p	R ²	p	R ²
Overall	0.009	0.63	0.0001	0.98
Hospital	0.27	-	0.0001	-
DRG category	0.0004	-	-	-
Hospital x DRG	0.38	-	-	-
Age	-	-	0.0001	-
Non-operating room procedure	-	-	0.0001	-
Pacemaker implant	-	-	0.0001	-
History of prior heart attack	-	-	0.0001	-

Table 4
Heart attack patients: Charges per admission

Hospital	All cases: Unadjusted	DRG 122 ¹ comparisons	All cases: Multivariate model
Hospital A	\$8,514	\$7,588	\$8,844
Hospital B	8,074	7,756	8,242
Hospital C	8,633	7,558	8,134
Differences between hospitals (p)		N.S.	p<0.0001

¹There were too few cases in other diagnosis-related group (DRG) categories to show any significant differences.

Medicine in the spring of 1983. In this hypothetical case situation, 27 patients who were discharged alive were admitted for heart attacks to three different hospitals in the same community. The current DRG algorithm categorizes such live discharges sequentially by whether or not they have had a permanent pacemaker implant, and whether or not they have had cardiovascular complications. Under the existing DRG framework for classifying heart attacks, the only surgical data used is whether or not there is a permanent pacemaker implant and the complications factors are related solely to the cardiovascular system. As shown in Table 2, after adjusting for the highly significant class variable (p = .0004) of the current DRG category, no differences could be demonstrated among the hospitals (p = .27). In contrast, a simple multivariate model incorporated an age factor, a factor for whether or not the patient had any nonoperating room procedures (such as insertion of a Swan-Ganz catheter, an endoscopic procedure, etc.) and whether or not the patient had a prior history of a heart attack (which could be determined fairly readily in a longitudinal data file). In this model, while the pacemaker variable is still the most important single factor explaining variation in charges, all of the other factors contributed and produce an R² value 50 percent greater than a DRG hospital comparison. Additionally, the conclusion of the analysis using the multivariate model is that there are significant differences in severity among the three

different hospitals; in fact, it reverses the unadjusted results. What appears to be the most expensive hospital in terms of the unadjusted charges is actually the least expensive hospital after adjustment through the multivariate model. While this is obviously a simplified example, it does illustrate the potential to use more clinical diagnostic and procedure data than is employed in the current DRG framework and to potentially obtain superior reductions in variance through these techniques. Revisions of severity adjusters—potentially unfettered by the serious clinical constraints of current DRG branched groupings—might provide substantial additional explanatory power and could, if useful, be quickly implemented because the data is readily at hand.

Another example of a factor which could be incorporated is whether there is a history of any socio-medical problem such as alcohol abuse, even though that may not be an active clinical issue during the admission. Data collected by The Health Data Institute, Inc. on auto workers indicated that for major nonsubstance abuse admissions, individuals with a history of alcoholism have more expense per admission than persons admitted without such a prior history, controlling for all other relevant case-severity factors (McGuire and Fairbank, 1984).

If the research on revision and enhancement of DRG's is not constrained to working with just the current UHDDS data items, a series of other research options become available. Again, data issues are key. The next practical question is whether any expansion of the items needs to be limited to a small number of additional fields (e.g., five to ten characters) or whether the process can be more open-ended in the types of data elements that are used. A further subdivision is whether common elements are to be used as adjusters for all diagnoses or procedures; if so, then the focus would be on generic, nondiagnostic clinical measures such as vital signs (research option area D). Factors like the admission temperature, the highest peak temperature, the lowest systolic blood pressure during the hospitalization, etc., would frequently indicate the severity of a condition, the presence of comorbid conditions or the development of complications.

Table 5 shows how one of these clinical factors, peak temperature, can considerably enhance the explanation of variance within a single DRG, uncomplicated heart attacks. In a project for Alcoa, CPHA data sets were obtained for six southern community hospitals which were the sole source of health care in the community, as part of an effort to understand why the Alcoa Corporation's length of stay experience at its main aluminum plant in Marysville, Tennessee was so high (Lind *et al.*, 1979). As part of that effort, an analysis was done for a set of common conditions to pinpoint differences in resource utilization which employed standard diagnostic and procedure information. The analysis also attempted to adjust for severity of illness by use of the medical audit component of the CPHA Hospital Discharge Abstract

Table 5
Use of clinical variables as length-of-stay¹ adjusters:

Variable	P
Peak temperature	.0001
Sex	.0002
Total number of drugs	.003
Use of cardiac drugs	.001
Hospital	.0001
Model R ²	.23

¹For diagnosis-related group 122

SOURCE: Lind, K., Gertman, P. M., Anderson, J. J., Egdahl, R. H. Alcoa Project: Study of six sole source southern community hospitals over 300 beds. Boston University Center for Industry and Health. Unpublished data, 1979.

form.² Peak temperature was by far the most significant factor in explaining variations in length of stay among hospitals for this DRG and also turned out to be the most significant factor in explaining variations in charges among physicians at Alcoa's principal community hospital. Temperature also was a frequently significant adjuster for other standard high-risk conditions. Table 4 shows that it is possible to obtain additional reductions in variance by use of other treatment and clinical factors within a defined DRG category.

Clinical variables not currently in the UHDDS data set may be conceptually more attractive as severity adjusters because: 1) they are more directly tied to a patient's actual clinical status than is a diagnosis, and 2) they can be selected to more likely be exogenous rather than endogenous variables. Thus, an admission blood pressure level cannot readily be directed by the provider, while performing surgery can be done with considerable discretion.

If variable severity elements are allowed for different conditions (research option area E), type and extent-of-disease spread variables could be incorporated for malignancies, medically derived cardiovascular classes could be identified (e.g., American Heart Association functional rating for general cardiac patients, Killip classifications for heart attacks, Plum-Pozner codes for strokes), Trauma Severity Index for injuries and other limited data elements shown in clinical trials to be important prognostic severity of illness measures. If variable special elements could be incorporated, the availability of four, six or ten additional characters of data might dramatically improve the severity of illness adjusters and thus the clinical homogeneity of the groups created for reimbursement purposes. In fact, HCFA need not be limited to adding data fields to the UHDDS data set, but rather might replace fields that

²Despite years of data collection, remarkably little has been done to use the data from this MAP portion of the standard CPHA Hospital Discharge Abstract as severity of illness adjusters. This is a potentially rich source of data for severity grouping research.

are currently reserved for other purposes. For example, by dropping the five digits reserved for the fifth diagnosis—an item with potentially extremely small marginal utility—one could encode in two character fields the peak temperature (e.g., 99.6, 103.5, etc.) and the lowest systolic blood pressure (which would occupy only three character fields).

If extensive additional data elements could feasibly be added to the UHDDS dataset for Medicare, then much more clinically-oriented physiologic parameters could be used in the severity adjustment methodology (research option area F). One example of this is the acute physiology and chronic health evaluation (APACHE) system used at George Washington University. In APACHE, more than thirty additional clinical measurements, including blood gas determinations, are incorporated into the classification weighting scheme for intensive care patients (Knaus *et al.*, 1981). Finally, if one is not limited to the UHDDS dataset, severity of illness classification based on judgmental criteria such as the Severity of Illness Index of Susan Horn and her associates and the relative intensity measures (RIMS) nursing care weighted system, etc., could be tested (research option area G).

A critical policy decision on the approach to incorporating severity measures into the DRG revision process is whether an attempt should be made to develop a *comprehensive* set of revisions which would be introduced *en masse*, or whether revisions should be promulgated through a piecemeal or phased-in approach. (By the latter, we do not mean a phase-in of a comprehensive set of revisions because, if evidence was present of a superior method of classification, it would probably be politically unacceptable to the losers under the old approach for HCFA to not introduce all of the revisions at once.) The arguments for and against either of these approaches are basically political rather than technical. Yet, the decision on which is the preferable course to the Government and the health care industry is important to provide guidance for an efficient research strategy and would prioritize among the areas.

Some approaches, such as the Severity of Illness Index technique, are essentially comprehensive approaches and do not lend themselves to partial introduction for the DRG list of categories. On the other hand, approaches which would further subdivide or minimally reorganize terminal DRG's within a major diagnostic category, such as the SystemeMetrics approach, using adjacent diagnosis-related groups or additional age splits, are more reasonable on a selective introduction basis.

Conclusion

There is a final issue which will plague the DRG reimbursement system even if all data confounding and severity problems could be resolved perfectly—the issue of effectiveness of treatment. There has been an assumption in many quarters that if any group of patients had exactly the same clinical needs, their physicians or any outside physicians would agree on

the “product” to be delivered. The actual situation today in American medicine is that there are sometimes vast, well-reasoned and sincere differences among outstanding physicians about the best course of diagnosis and treatment. If these differences are associated with substantial and systematic differences in resource use within a DRG or other adjustment schemes, then American medicine will continue to resist the system, labelling it as inequitable and as potentially containing adverse quality of care consequences.

References

- Demlo, L., Campbell, P., and Brown, S.: Reliability of information abstracted from patients' medical records. *Med. Care* 1978; 16:995.
- Doremus, Harvey and Mickenzi, Elana.: Data quality: an illustration of its potential impact upon a diagnosis-related group's case mix index and reimbursement. *Med. Care* 1983; 21:1001.
- Fernow, C. Health Care Systems International. Personal communication, 1983.
- Garg, M. L., Louis, D. Z., Glibe, W. A., *et al.*: Evaluating inpatient costs: the staging mechanism. *Med. Care* 1978; 16:191.
- Gertman, P. and Stanton, B.: Preventable Hospital Admissions. Presented at the American Public Health Association Annual Meeting, Minneapolis, Minn., 1975.
- Gertman, P. M., and Restuccia, J. D.: The Appropriateness Evaluation Protocol: A Technique for Assessing Unnecessary Days of Hospital Care. *Med. Care* 1981; 19, 8:855-871.
- Gertman, P. M., Restuccia, J. D., Dayno, S. J., Resnick, A. L., Lenhart, G. M., Anderson, J. J.: *A Comparative Assessment of Inappropriate Hospital Use*. Contract No. SSA-600-75-0209. Prepared for U.S. Department of Health and Human Services. Washington, D.C., 1982.
- Horn, S. D.: Measuring severity of illness: comparisons across institutions. *American Journal of Public Health* 1983; 73:25.
- Horn, S. D., Bulkley, G., Schramm, C., *et al.*: Inter-hospital differences in patient severity: problems for prospective reimbursement based on diagnostic groups. Baltimore, Maryland: Johns Hopkins University, 1983.
- Hughes, E., Gertman, P. M., Anderson, J. J., Friedman, R. H., Rosen, M., Ward, A., Kreger, B. E.: The Ancillary Services Review Program in Massachusetts: Experience of the 1982 Pilot Project. *Journal of American Medical Association*, 1984; 252:1727.
- Knaus, W., Zimmerman, J., Wagner, D., Draper, E., *et al.*: APACHE—acute physiology and chronic health evaluations: A physiologically based classification system. *Critical Care Medicine* 1981; 9:591.
- Lind, K., Gertman, P. M., Anderson, J. J., Egdahl, R. H.: Alcoa project: study of six sole source southern community hospitals over 300 beds. Boston University Center for Industry and Health. Unpublished data, 1979.
- Louis, D. and Heineccius, L.: Refinements to diagnosis related groups based on severity of illness and age. Santa Barbara, California: SystemeMetrics, Inc., July 1983.

McGuire, T. and Fairbanks, A.: The Cost of alcoholism: A study for Auto Workers in Michigan. Newton, Massachusetts: The Health Data Institute, Inc., 1984.

Restuccia, J. D., Gertman, P. M., *et al.*: A Comparative Analysis of Appropriateness of Hospital Use. *Health Affairs*, June 1984.

SysteMetrics, Inc.: The measurement of inappropriate hospital utilization in a nationally representative sample. Santa Barbara, California: SysteMetrics, Inc., January 1983.

SysteMetrics, Inc.: Refinements to diagnosis related groups based on severity of illness and age. Santa Barbara, California: SysteMetrics, Inc., September 1983.

Yale University: The new ICD-9-CM diagnosis related groups classification scheme: final report. New Haven, Connecticut: May 1982.

Young, W. W.: Measuring the cost of care using generalized patient management paths. Year 1 Final Report. Pittsburgh, Pennsylvania: Blue Cross of Western Pennsylvania, November 1979.