Overview and Historical Perspective

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

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

In general, the broad outlines for the construction of both sets of DRGs were the same for each version. Actual patient stays in a sample of hospitals were the units of analysis. Each patient’s principal diagnosis—i.e., the reason (after study) that the patient was admitted—was coded using a detailed coding system that allowed for many thousands of possibilities. The first step, therefore, was to collapse the detailed diagnosis codes into meaningful, but broad, subcategories called “Major Diagnostic Categories” (MDCs). MDCs were then further subdivided, using a combination of statistical analysis and medical judgment, according to other characteristics that accounted for differences in resource use within the MDC.

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

In addition, there were a number of differences in the specific features of the development process. The differences were so extensive that there is very little correspondence between the two sets of DRGs. In the remainder of this appendix, the procedures used to create each of the two sets will be summarized and the similarities and differences among them will be examined.

Development of the “Original” 383 DRGs**

The original 383 DRGs were developed from data for approximately 500,000 patients. Most of these were from New Jersey hospitals, though additional data were also available from a large hospital in Connecticut and for a sample of patients reviewed under the Federal Professional Standards Review Organization program. Before the data were analyzed, cases thought to be misleading or unrepresentative were eliminated from further consideration. These included deaths, miscodes, and patients with extremely long lengths of stay (LOS). The reason for this exclusion was that the overriding objective of the process was to describe a “typical” patient. Apparently, aberrant cases were disregarded.

As a first step, clinicians classified the patient records into 83 mutually exclusive and exhaustive MDCs. MDCs were then further subdivided, using a combination of statistical analysis and medical judgment, according to other characteristics that accounted for differences in resource use within the MDC.

The major differences, however, may appear to outweigh the similarities. Significantly modified pro-

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

● Much of the substance of this section was derived from Fetter and colleagues (19).
Analysis was then performed to determine if each it was much more similar within groups of elderly and MDC should be further subdivided in order to reduce nonelderly patients, then a decision would be made the variance in length of hospital stay. LOS, as noted to divide the MDC “Pneumonia” further, according earlier, was selected because it was viewed as a key to whether or not the patient was 65 or older. A next indicator of resource use and the best such indicator step might involve examining the improvement in LOS for which data were available. Justification for its use only in the close correspondence between LOS, case com-variance” in LOS) when elderly and nonelderly parameters, and cost that has previously been reported in tients were further subdivided according to the pres-the literature (19). Particularly because of the per diem or absence of a secondary diagnosis. The process based rate structure of many of the existing reimbursee-would continue until it was determined that further ment systems, LOS was considered to be a reasonably subdivision would not significantly reduce the vari-accurate and accepted measure of resource use. In ad-ance, would not be medically meaningful, would result dition, the consistency with which this measure was in too many groups, or would result in too few cases reported was considered a practical advantage. More contained in a group.

direct measures of resource use, such as charges, were The number of variables investigated was deliberate-and are not only more difficult to obtain, but they ly limited to a small number to reduce the complexity are more difficult to make comparable across areas, of the analysis. However, they consisted of the vari-able to wage and price differences, and across hospitals, ables that the developers considered to be the key dis-due to differing markups. Consequently, LOS became criminators among the characteristics for which data the focus of the statistical analysis and strongly in-could be obtained from hospital records. Others, suchfluenced the final form of the diagnostic categories, as sex, were tested but did not prove to be important

Many have argued that while the number of days in explaining variations in LOS, a patient is hospitalized is one indicator of resource consumption, LOS by itself may not be an accurate and accepted measure of resource use. In ad-ance, would not be medically meaningful, would result dition, the consistency with which this measure was in too many groups, or would result in too few cases reported was considered a practical advantage. More contained in a group.

The process of defining DRGs was therefore a co- nsumption, LOS by itself may not be an accurate and accepted measure of resource use. In ad-ance, would not be medically meaningful, would result dition, the consistency with which this measure was in too many groups, or would result in too few cases reported was considered a practical advantage. More contained in a group.

The use of LOS as the primary measure of resource One result of this was that some MDCs, such as acute consumption also contributes to the lack of homog- myocardial infarction, were not subdivided at all and neity within the original DRGs formed. For instance, that, others, such as fractures, were subdivided into as the old DRGs grouped together in a single category many as 13 DRGs. Also, this approach resulted in vari-lung cancer patients with a short diagnostic workup, ation across MDCs in the criterion (or criteria) for sub-a lengthy chemotherapy treatment, or a terminal ad-division. For some, such as appendicitis, secondary mission (4). Researchers have suggested that clinical diagnosis was the criterion for subdivision; while for data in addition to those already used in the original pneumonia, age, surgery, and secondary diagnosis DRG classification system are needed in order to con-were all used. This difference was accepted insofar as struct groups that are more homogeneous from both it was seen to reflect variation in relevant patient char-a clinical and resource consumption standpoint (4,13).acteristics.

For instance, age, socioeconomic status, and type of admission have been suggested as important elements ly satisfied the criteria established for distinctiveness, in classifying patients into homogeneous groups (4). medical meaningfulness, and size. It is worth noting

Subdivision of MDCs resulted from an iterative once again that the 383 DRGs were based on input data process, during which statistical output was reviewed derived from a sample of patients that was mainly lim- by clinicians in order to determine which grouping al-ited to the northeast region. The result thus reflected ternatives best satisfied both medical and statistical cri-the composition of cases in that sample, as well as the teria. The statistical analysis involved the assessment medical practice patterns employed in the hospitals of whether within-group variance in LOS was signif- from which these patients were discharged. Also, judg-icantly reduced when the patients were subdivided ac-ments about the alternative grouping configurations cording to secondary diagnosis, primary and second- were made by a small group of clinicians whose views ary surgical procedures, and age. For example, if LOS may not have been representative of physicians na-among all pneumonia patients was highly variable, buttonially. The possibility that the initial set of DRGs
was not generalizable was raised by many critics (38, 106). This issue was addressed in the subsequent development process.

Most of the evaluation of DRGs to date has focused on the original 383, as the modified set was disseminated only recently. In part in response to criticism from the medical community, researchers, administrators, and others, modifications to the DRG development process were begun in the late 1970's. The modified procedures will be described in the following section.

**Development of the Modified 467 DRGs**

The revised set of DRGs has been described as a “major departure” from the original 383 (103). The opportunity for reevaluation and revision of the original set arose with the promulgation, in 1979, of a revised diagnostic coding scheme, known as the International Classification of Diseases, 9th Revision, Clinical Modification, or ICD-9-CM for short. The introduction of the ICD-9-CM coding convention was designed in part to increase precision in diagnosis and procedure coding. Its introduction meant that the two previous coding systems (known as ICDA-8 and HICD-2), in which the 383 DRGs had been defined, would be superseded in hospitals' medical records departments. As the ICD-9-CM system could be related back to the earlier coding schemes (though with some loss of information), the implementation of the new system did not in itself cause the 383 DRGs to become obsolete. However, the original DRGs were defined from a less refined coding system and thus could not benefit from the increased precision of ICD-9-CM unless they were redefined.

Revision of the DRG classification scheme was undertaken not only to take advantage of the improved diagnosis coding, but also to remedy perceived deficiencies in the earlier approach. Criticisms that had been leveled at the earlier system mostly concerned the limitations inherent in the input data, the small number of variables considered, and the lack of clinical homogeneity within some of the individual groups (32,103, 106). Some who used the original DRGs argued that they were not medically meaningful. For example, Williams and his colleagues (99) used a Delphi technique to identify the cause and significance of problems with the original 383 DRGs. Specifically, physicians, hospital administrators, and university researchers who were involved in the New Jersey DRG experiment were surveyed several times in order to reach a consensus on clinically inappropriate patient assignments within the original DRGs. The 38 experts who participated in this study identified and ranked 16 categories of problems that caused inappropriate assignments. Most important to them was that “some DRGs combine clinically similar patients, who nevertheless require different treatments” (99). They also identified 37 of the original 383 DRGs as categories likely to contain patients whose clinical status was not appropriately recognized. Physicians and hospital administrators identified different reasons for the problems, but they specified the same problematic DRGs (99).

Several of the problems identified by Williams and colleagues concerned the medical meaningfulness of DRG-based case-mix measures. Some who have used the original DRGs argue that they are not medically meaningful because patients with very different medical problems are grouped together (4). For instance, old DRG 301 groups together all patients whose principal diagnosis is “replacement of hip with prosthetic device, biopsy of bone, and spinal fusion.” In addition, DRGs fail to subdivide some broad diagnostic groups. The original DRG 121, for instance, includes all patients with acute myocardial infarction.

As a result of such criticisms, significant changes were made in the organization and orientation of the development process, the manner in which decisions were made, and the nature of the input data. Specifically, a more structured organization was used to administer the classification process and to guide the decisionmaking. A large number of participants from the medical profession, as well as other areas within the health industry (e.g., medical records) were involved. For this later phase, data were made available by the Commission for Professional and Hospital Activities from the Professional Activity Study (PAS). This meant that a nationally representative sample of patients could be selected and analyzed, thus improving the generalizability of the results.

DRG development procedures were substantially altered during this second phase. The major change was in the basic orientation of the decisionmaking, in that strong emphasis was placed on the clinical, rather than statistical, validity of DRGs. The first manifestation of this change was in the redefinition of MDCs. Rather than using the 83 MDCs defined previously, the revised approach redefined a total of only 23 MDCs, most of which were confined to a single organ system. To return to the example used earlier, diseases of the respiratory system, which were represented by six MDCs in the earlier methodology, were represented by a single MDC in the modified approach. Only a few MDCs (e.g., burns) remained the same. The purpose of this change was to bring MDCs into conformance with the organization of medical practice, in which, for the most part, specialties are defined around

---

* Much of the substance of this section was derived from Yale University (103).
the various organ systems. This also appeared to facilitate the increased use of expert clinical judgment in developing MDC subdivisions.

Additional changes were made in the process after the redefinition of MDCs. An important change in the analysis was the retention of patients who died. Another was the development and use of more precise guidelines for subdividing MDCs than had been used previously. An outgrowth of this was more consistency in the application of criteria for subdivision. For example, the guidelines required that the initial partition (when possible) be based on the presence or absence of a surgical procedure performed in an operating room. The need for grouping patients with clinically related diseases was continually stressed in the guidelines.

Several important modifications were made in the variables used to subdivide MDCs, primarily to capture severity of illness more precisely. One such modification instructed the expert panels to rank order surgical procedures according to resource use intensity and to assign patients with multiple procedures to the procedure involving the greatest intensity. This meant that the type of surgical procedure became an important consideration in the new DRG development, whereas in the original grouping procedure, only the presence or absence of surgery was taken into account. In addition to considering the type of surgery performed, the nature of existing comorbidity (i.e., coexisting conditions) and complications was explicitly evaluated based on the specific ICD-9-CM codes contained in the patient's discharge abstract. Again, "substantial" comorbidity and complications were distinguished from those considered to be less significant. "Substantial" was defined to include conditions likely to increase LOS by at least 1 day for at least 75 percent of the cases. In many instances, a composite variable indicating whether or not the patient was aged 70 or more and/or had substantial comorbidity or complications proved to be an important determinant of resource use.

Finally, other variables in addition to diagnosis, procedures, and age were taken into account when the experts judged that the additional factors were relevant. For example, for the MDC "Pregnancy, Childbirth, and the Puerperium," the initial division is made according to whether or not the patient was "delivered this admission." With respect to substance abuse, the initial split is according to whether or not the patient "left [the hospital] against medical advice." Death was also included as a possible criterion for subdivision.

While, as noted earlier, the modified process was much more dependent on clinical judgment, statistical analysis again was used to aid decisionmaking. Reduction in variance for LOS was again examined for each partitioning variable considered, but the fact that variance was significantly reduced by a particular variable did not guarantee that that variable would be included in the modified DRG definition. Clinical coherence, above all, dictated which measures would be used.

The modified approach resulted in the definition of 467 DRGs, * which bear little resemblance to the original 383. To the extent that the original groups can be "mapped" into the revised ones, it is clear that while some of the original groups were further subdivided by the new process, others were collapsed into fewer categories. For example, the original DRG 121 was "acute myocardial infarction" (AMI), undifferentiated. In the new configuration, AMI patients are classified into three DRGs:

- circulatory disorders with AMI and cardiovascular complications, discharged alive;
- circulatory disorders with AMI, without cardiovascular complications, discharged alive; and
- circulator disorders with AMI, expired.

In an example of the opposite effect, bronchitis and asthma were divided into three bronchitis- and three asthma-related DRGs under the original system. The modified set includes only a total of three DRGs for bronchitis and asthma combined.

Comparison of Alternative Sets of DRGs

Table B-1 presents a summary of the fundamental similarities and differences between the original 383 DRGs and the modified set of 467. Most of the specific areas shown in the table have been discussed in the previous section. This concluding section, therefore, will focus on the implications of the changes made.

Clearly, the major thrust of all of the methodological changes was to improve the medical meaningfulness of DRGs. To the extent that this was accomplished, it should result in the increased acceptability of the DRG scheme to physicians, who manage the medical care provided to hospital inpatients. As a consequence, the necessary interactions between clinical staff and hospital administrators should be improved. The development of a medical, meaningful grouping scheme has always been a clear objective of those who originated the concept of DRGs.

The enhancement of the "clinical coherence" of DRGs was attempted in several important ways:

- by increasing the number of clinician participants and the role of clinicians in DRG development;

* A 468th DRG has also been defined to account for patients who have received an operating room procedure unrelated to their MDC. This "outlier" category is not normally included in descriptions or evaluations of the system.
Table B.1.—Summary of Similarities and Differences Between the Original and Modified DRGs

<table>
<thead>
<tr>
<th>Original DRGs</th>
<th>Modified DRGs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on data for patients from New Jersey and Connecticut, and a sample of</td>
<td>Based on nationally representative sample of patients derived from PAS data</td>
</tr>
<tr>
<td>Medicare/Medicaid patients (deaths excluded)</td>
<td>(deaths retained)</td>
</tr>
<tr>
<td>Based on ICDA-8 (or HICD-2) diagnostic coding scheme</td>
<td>Based on CD-9-CM diagnostic coding scheme</td>
</tr>
<tr>
<td>Result from subdivision of 83 broad subcategories of diagnoses (MDCs) based</td>
<td>Result from subdivision of 23 broad subcategories of diagnoses (MDCs) based</td>
</tr>
<tr>
<td>on organ system and etiology</td>
<td>on organ system only</td>
</tr>
<tr>
<td>Subdivision of MDCs based on statistical analysis and clinical judgment</td>
<td>Subdivision of MDCs based on clinical judgment and statistical analysis</td>
</tr>
<tr>
<td>Variables used to subdivide MDCs include presence/absence of secondary</td>
<td>Variables used to subdivide MDCs include type of surgery and</td>
</tr>
<tr>
<td>diagnoses and surgery as well as age</td>
<td>comorbidity/complications, age, death, and other relevant criteria</td>
</tr>
<tr>
<td>Subdivisions not uniform across DRGs</td>
<td>When possible, first subdivision based upon presence/absence of operating</td>
</tr>
<tr>
<td></td>
<td>room procedure; generally, tighter guidelines for subdivision were applied</td>
</tr>
<tr>
<td>End result: 383 mutually exclusive and exhaustive DRGs</td>
<td>End result: 467 mutually exclusive and exhaustive DRGs</td>
</tr>
</tbody>
</table>

SOURCE: Office of Technology Assessment.

- by tying the initial subdivisions (MDCs) to major organ systems, in conformance with the delineation of medical specialties;
- by taking the specific nature of surgical procedures and comorbidity/complications into account in forming the groups; and
- by extending the number of characteristics used to partition MDCs when appropriate.

It would appear that significant strides have been made towards the objective of medical meaningfulness. These would also seem to be important strides if physician behavior is to be the target of management or cost control efforts.

It is unclear whether grouping according to clinical similarities was attained at the expense of statistical validity, and if so, whether it is important. The deemphasis of statistical analysis as a mechanism for decisionmaking in forming the groups implies that the new groups may be less internally homogeneous and distinct from each other than the original set. Admittedly, however, the basis for evaluation of within-group homogeneity and between-group heterogeneity—LOS—is an imperfect criterion for indicating resource consumption. Preliminary evidence does indicate, however, that the groups achieve similar reductions in variance for charges as they do for LOS (103). Data available for about 330,000 New Jersey discharges (including cost data) were analyzed, and the results indicated that the distribution of costs was similar to that of LOS for most DRGs. In the few instances where there were significant differences, modifications were made to the relevant DRGs.