Structure and Performance of Different DRG Classification Systems for Neonatal Medicine

John H. Muldoon, MHA

ABSTRACT. There are a number of Diagnosis-Related Group (DRG) classification systems that have evolved over the past 2 decades, each with their own strengths and weaknesses. DRG systems are used for case-mix trending, utilization management and quality improvement, comparative reporting, prospective payment, and price negotiations. For any of these applications it is essential to know the accuracy with which the DRG system classifies patients, specifically for predicting resource use and also mortality.

The objective of this study was to assess the adequacy of the three most commonly used DRG systems for neonatal patients—Medicare DRGs, All Patient Diagnosis-Related Groups (AP-DRGs), and All Patient Refined Diagnosis-Related Groups (APR-DRGs). A 2-part methodology is used to assess adequacy. The first part is a descriptive analysis that examines the structural characteristics of each system. This provides a framework for understanding the inherent strengths and weaknesses of each system and for interpreting their statistical performance. The second part examines the statistical performance of each system on a large nationally representative hospital database.

The analysis identifies major differences in the structure and statistical performance of the three DRG systems for neonates. The Medicare DRGs are structurally the least developed and yield the poorest overall statistical performance (cost $R^2 = 0.292$; mortality $R^2 = 0.083$). The AP-DRGs are structurally the most developed and yield the best statistical performance (cost $R^2 = 0.627$; mortality $R^2 = 0.416$). The AP-DRGs are intermediate to Medicare DRGs and AP-DRGs, although closer to APR-DRGs (cost $R^2 = 0.507$; mortality $R^2 = 0.304$).

An analysis of payment impacts and systematic effects identifies there are major systematic biases with the Medicare DRGs. At the patient level, there is substantial underpayment for surgical neonates, transferred-in neonates, neonates discharged to home health services, and neonates who die. In contrast, there is substantial overpayment for normal newborns. At the facility level, there is substantial underpayment for freestanding acute children’s hospitals and major teaching general hospitals. There is overpayment for other urban general hospitals but this pattern varies by hospital size. There is very substantial overpayment for other rural hospitals. The AP-DRGs remove the majority of the systematic effects but significant biases remain. The APR-DRGs remove most of the systematic effects but some biases remain.

ABBREVIATIONS. DRG, Diagnosis-Related Group; UB, uniform bill; LOS, length of stay; PPS, prospective payment system; HCFA, Health Care Financing Administration; NACHRI, National Association of Children’s Hospital and Related Institutions; PM-DRG, Pediatric Modified Diagnosis-Related Group; MDC, major diagnostic category; AP-DRG, All Patient Diagnosis-Related Group; CC, comorbid-complicating condition; RDRG, Refined Diagnosis-Related Group; SR-DRG, Severity Refined Diagnosis-Related Group; APR-DRG, All Patient Refined Diagnosis-Related Group; DR, operating room; UHDDS, Uniform Hospital Discharge Data Set; RCC, ratio of costs-to-charges; GME, graduate medical education.
nesses of each system and for interpreting the statistical performance of each system.

The second part of the methodology examines the statistical performance of each system. This is done in two ways. First, the overall predictive power of each DRG system for resource use and mortality is measured through explanation of variance statistics. Second, payment impacts and systemic effects are measured at the patient level as defined by various patient attributes and at the hospital level as defined by overall results for various groupings of hospitals. To assist with interpretation, the results by hospital type are linked back to the results by patient type.

This article is organized into seven sections. The first, “Evolution of DRG Classification Systems,” explains how DRG systems have evolved since the late 1970s with an emphasis on the different approaches to severity adjustment within a DRG category. The second, “Data Elements for DRG Classification Systems,” describes the data elements, their source, and some of their strengths and limitations. The third, “Comparison of Structural Characteristics of DRG Classification Systems,” provides a summary description and a detailed technical description of each system. The fourth, “Database for Comparative Analysis of DRG Classification Systems,” explains the representativeness of hospitals in the study sample frame, the source of the data, and strengths and weaknesses of the available measures for resource consumption. The fifth, “Explanation of Variance (R^2) for Resource Use and Mortality,” presents the overall explanatory power of each DRG system for predicting cost, length of stay (LOS), and mortality. The sixth, “Payment Impacts and Systematic Effects by Patient Type and Hospital Type,” presents results for a detailed series of payment simulations for each DRG system. The seventh, “Conclusion,” presents the overall conclusions from the study.

**EVOLUTION OF DRG CLASSIFICATION SYSTEMS**

The design and development of DRGs began in the late 1960s at Yale University. The initial motivation was to create an effective framework for monitoring the utilization of services in a hospital setting. The first large-scale application of DRGs was conducted in the late 1970s by the State of New Jersey in its hospital prospective payment system (PPS). In 1984, a DRG-based PPS was implemented for the Medicare program. Subsequently, a number of states and large payors implemented DRG-based PPS for non-Medicare patients. In addition, DRGs have been used as the basis for global budget allocation and payment in several countries in Western and Eastern Europe as well as Australia.1

The initial DRG system developed by Yale was intended to describe all types of patients seen in an acute care hospital. There was an inherent problem, however, in that the database used for its development attempted to be representative of a cross section of community hospitals. This ensured there would not be sufficient case volume of complex low-volume pediatric and neonatal conditions to detect certain problems or to develop solutions. Of note, freestanding acute children’s hospitals were not included in the Yale study database.2

The updating of the DRG system used for Medicare PPS in the 1980s and 1990s has been done by the US Health Care Financing Administration (HCFA). The database used to evaluate potential updates has been a Medicare patient only database and the focus has been on problems related to the elderly and other Medicare recipients. So, the shortcomings of the original DRGs for pediatric and neonatal patients have never been addressed by the Medicare DRGs. HCFA acknowledged in its annual rulemaking process in August 1991 that its DRG research was designed to develop improvements for the Medicare population and it should not be assumed that the Medicare DRGs are appropriate for other patient populations. Nevertheless, many state Medicaid programs, Blue Cross plans, and other payors have used the Medicare DRGs to classify newborns, children and non-elderly adults for hospital payment purposes.3

A first attempt to systematically evaluate the Medicare DRG system and develop refinements for children took place from 1984 to 1986 in a study entitled “Children’s Hospital Case-Mix Classification Study Project,” funded by the National Association of Children’s Hospital and Related Institutions (NACHRI), the HCFA, and The Pew Memorial Trust. This led to follow-on work by NACHRI and the release of the Pediatric Modified Diagnosis-Related Groups (PM-DRGs) in 1988. PM-DRGs contained an extensive set of modifications to the Medicare DRGs for pediatric patients and a whole new structure for major diagnostic category (MDC) 15, the primary diagnostic category for neonatal patients.4 (Note: an MDC is a body system or major disease process. There are 25 MDCs in most DRG systems.)

Part of the pediatric and neonatal modifications of the PM-DRGs were incorporated into the All Patient Diagnosis-Related Groups (AP-DRGs) adopted by the New York State Health Department in 1988. The AP-DRG system also introduced other new categories applicable to patients of all ages, including a new MDC for patients with multiple significant traumas and a new MDC for patients with human immunodeficiency virus. The AP-DRGs have been used for prospective payment by several state agencies, Blue Cross plans and other payors. The US Department of Defense has also incorporated the neonatal definitions of the initial AP-DRGs into its DRG-based PPS for the Civilian Health and Medical Program of the Uniformed Services.5

The initial generation of DRG systems provided only modest differentiation for severity within a DRG category. Certain of the DRG categories were split into two categories based on the presence or absence of a secondary diagnosis from a list of co-morbid-complicating (CC) conditions that included approximately 3000 of the 12 000+ International Classification of Diseases, 9th Revision Clinical Modification (ICD-9-CM) diagnosis codes. The CC split was a simple yes/no split. No differentiation was made as to whether certain of the CC diagnoses represented more extreme conditions or whether the patient had multiple CC diagnoses.6
The effort to develop a more advanced severity adjustment methodology began in the mid-to-late 1980s when HCFA funded a 2-year study project by Yale University. This project produced the first “Refined” Diagnosis-Related Group (RDRG) system, which is a DRG system with multiple CC (or severity) levels within each DRG category. Nearly all the DRG categories were given either three or four severity subclasses (mild, moderate, major, extreme) based on the presence of certain secondary diagnoses.

The AP-DRGs implemented by New York State in 1988 designated a subset of secondary diagnoses as major CCs. These diagnoses were similar to those classified as catastrophic by the initial RDRGs. To avoid significantly increasing the number of AP-DRG categories, AP-DRG major CC categories were formed for groups of surgical AP-DRGs and medical AP-DRGs in a body system.8

In 1993 HCFA funded another 2-year study project to develop a severity adjustment methodology for the Medicare DRGs. The study, which was conducted by 3M Health Information Systems, produced a system at the end of 1994 entitled “Severity Refined” DRGs (SR-DRGs). This system was developed using Medicare patient data only and excluded DRGs associated with pregnancy, newborns, and pediatric patients. Major CC subclasses were developed for many of the DRG categories. As of June 1998, HCFA had not announced plans to implement the SR-DRGs, nor have the SR-DRGs been updated since 1994.9

The severity-adjusted version of DRGs that has come into widest acceptance and use in the 1990s is the All Patient Refined DRGs (APR-DRGs), first introduced in 1991. As of June 1998, there were >1400 hospitals and other organizations using the APR-DRGs. This included 17 state health departments and data commissions using the APR-DRGs for comparative profiling of hospitals. The 17 states are identified in Table 1. One of the first and most extensive hospital profiling reports produced by a state agency was the 1996 Guide to Hospitals in Florida.10

The APR-DRGs are developed and updated through the combined research activities of 3M Health Information Systems and the NACHRI. The APR-DRGs are different from other DRG systems in a number of respects including: 1) definitions for DRG categories; 2) revisions to surgical hierarchies; 3) updates to diagnoses on the CC list; 4) assignment of all diagnoses to one of four CC (or severity) levels; 5) severity subclass algorithm that takes into account the interactive effect of multiple secondary diagnoses; 6) a severity subclass methodology specifically developed for neonatal patients; and 7) a separate subclass methodology for risk of mortality.11

### DATA ELEMENTS FOR DRG CLASSIFICATION SYSTEMS

There are six data elements that are the building blocks for all these DRG classification systems. They are the same for each system.

- principal diagnosis
- secondary diagnoses
- operating room (OR) and non-OR procedures
- age (derived from date of birth and admission date)
- sex
- discharge disposition

These data elements come from or are derived from the Uniform Hospital Discharge Data Set (UHDDS), which defines the core data set for all hospital inpatient medical record abstracting systems. The UHDDS data elements are part of the UB-92 (Uniform Bill 1992), the required electronic billing format for a hospital to submit a bill to a payor. The UHDDS and the UB-92 bill are both considered administrative data sets as they are used for the billing of health care services. They are also called secondary data sets because they are generated as a by-product of a nonresearch activity. They contain a great deal of demographic, diagnostic, and treatment information and so can be used for a number of clinical as well as financial applications.12

The other data elements that are part of the UHDDS are: hospital identifier, patient identifier, date of birth, race, residence (ZIP code), admission date, discharge date, attending physician identifier, surgeon identifier, procedures with dates, and expected source of payment. These additional variables, while not used to group patients into DRG categories, can be used as part of DRG-based analyses. For example, data variables such as race and expected payment source (eg, Medicaid, self-pay) can be used to further describe patient attributes.

Of note, birth weight is not a UHDDS data element and there is no field on UB-92 for birth weight. However, birth weight range information can be obtained from the fifth digit of the ICD-9-CM prematurity diagnosis codes. The birth weight ranges are: <500 g, 500 to 749 g, 750 to 999 g, 1000 to 1249 g, 1250 to 1499 g, 1500 to 1749 g, 1750 to 1999 g, 2000 to 2499 g, >2499 g, and unspecified. In addition, if a hospital medical record abstract system is designed with a separate field for birth weight the software for both AP-DRGs and APR-DRGs is programmed to read birth weight as a separate variable.

The UHDDS diagnosis and procedure codes are both from the ICD-9-CM.13 As of October 1997, there

---

**TABLE 1. State Agencies With APR-DRG Software Licenses**

<table>
<thead>
<tr>
<th>1.</th>
<th>Arizona Department of Health Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>California Office of State Health Planning and Development</td>
</tr>
<tr>
<td>3.</td>
<td>Delaware Bureau of Health Planning</td>
</tr>
<tr>
<td>4.</td>
<td>Florida Agency for Health Care Administration</td>
</tr>
<tr>
<td>5.</td>
<td>Hawaii Health Information Corporation</td>
</tr>
<tr>
<td>6.</td>
<td>Indiana Department of Health</td>
</tr>
<tr>
<td>7.</td>
<td>Iowa Health Data Commission</td>
</tr>
<tr>
<td>8.</td>
<td>Maryland Health Services Cost Review Commission</td>
</tr>
<tr>
<td>9.</td>
<td>New Mexico Health Policy Commission</td>
</tr>
<tr>
<td>10.</td>
<td>Ohio Office of Medicaid Policy and Planning</td>
</tr>
<tr>
<td>11.</td>
<td>Oklahoma Health Care Authority</td>
</tr>
<tr>
<td>12a.</td>
<td>South Carolina State Budget and Control Board</td>
</tr>
<tr>
<td>12b.</td>
<td>South Carolina Workman’s Compensation</td>
</tr>
<tr>
<td>13.</td>
<td>Tennessee Department of Health</td>
</tr>
<tr>
<td>14.</td>
<td>Texas Health Care Information Council</td>
</tr>
<tr>
<td>15.</td>
<td>Utah Department of Health</td>
</tr>
<tr>
<td>16.</td>
<td>Virginia Health Information</td>
</tr>
<tr>
<td>17.</td>
<td>Wisconsin Office of Health Care Information</td>
</tr>
</tbody>
</table>

were 12,766 valid diagnosis codes and 3560 valid procedure codes. At some point, possibly in the year 2001, the United States will probably convert from ICD-9-CM to ICD-10-CM. Some European countries have already implemented ICD-10. The DRG systems will at that time be converted to read the ICD-10-CM codes. The draft version of ICD-10-CM has many more codes with more specificity than ICD-9-CM and thus will present an opportunity to introduce further improvements to the existing DRG systems.

The predictive power and clinical utility of a DRG classification system is limited or empowered to a large extent by the precision of the available diagnosis and procedure codes. The ICD-9-CM coding system is often criticized for lacking sufficient specificity and although this is true to a certain extent, there is also a wealth of information among the many thousands of ICD-9-CM codes. There is also a set of Official ICD-9-CM Guidelines for Coding and Reporting developed by the American Hospital Association, American Health Information Management Association, HCFA, and the National Center for Health Statistics and an established process through the ICD-9-CM Coordination and Maintenance Committee to update the codes annually. Thus, there is the opportunity that many physicians and other pediatric health professionals might not be aware of to develop proposed reformulations for individual diagnosis and procedure codes. To illustrate, this is how the fifth digit code for birth weight range was added to the prematurity diagnosis codes.\(^\text{14}\)

In addition to the information content of individual ICD-9-CM codes, a DRG classification system is limited or empowered by the way it uses the combined information from all diagnosis and procedure codes reported for a patient. In instances where an individual code is overly broad with respect to a specific diagnostic condition, the DRG grouper can examine for the presence of certain additional diagnoses to differentiate a more severe illness. The DRG grouper can also examine for the presence of certain procedure codes in instances where the procedure is consistently associated with a more severe illness and there is minimal practice pattern variation.

DRG classification systems have purposely restricted their data elements to those defined by UHDDS and used as part of UB-92. This is done to make it possible to assign DRGs to all hospital discharge data sets and to take advantage of existing processes for establishing coding guidelines and annually updating the diagnosis and procedure codes. At the same time, it is important to recognize that additional clinical data elements collected as part of primary data sets (eg, clinical trials, medical record) might add to the predictive power and/or clinical utility of DRG classification systems. For neonatology this might include gestational age, Apgar score (at specified time intervals), blood gas values (at specified time intervals), and the presence of or possibly certain types of prenatal care shown to affect outcomes. For this to be a realistic option for broad scale implementation there would have to be agreement among the provider and payor communities on the data elements and definitions and a way to expand the current UB-92 billing form. Whether this is likely to happen is difficult to say but it would probably require a demonstration that the additional data elements provide significant predictive power and clinical utility above that which can be obtained from existing data elements. There would also be a lead time for incorporation into the ongoing billing forms and computer claims processing systems.

In sum, the data elements used by the existing DRG systems are the same. The difference between the DRG systems is in how extensively and effectively they use these data elements to form clinically and statistically coherent classifications.

**COMPARISON OF STRUCTURAL CHARACTERISTICS OF DRG CLASSIFICATION SYSTEMS**

The structural characteristics of each DRG classification system identify at a broad level the strengths and weaknesses of each system. They also provide a conceptual framework for understanding and interpreting the statistical performance of each system.

Table 2 provides a comparison of the structural characteristics of Medicare DRGs, AP-DRGs, and APR-DRGs for neonatal patients. Appendices 1, 2, and 3 provide a full listing of all the neonatal categories in each DRG system. This section will begin with a summary description of the structural characteristics of each DRG system followed by a more detailed description.

The Medicare DRGs for neonates are vastly different from both the AP-DRGs and the APR-DRGs. The Medicare DRGs do not use age to define neonates (MDC 15); do not use birth weight as an initial grouping variable; do not distinguish surgical from medical patients; and do not provide breakouts for multiple problems or severity subclasses. There are only seven DRG categories for neonates including one defined for neonates who are either transferred or die. The Medicare DRGs for neonates have not been substantively updated since their initial implementation in 1984. Given these structural characteristics, there is little reason to expect the Medicare DRGs to yield a high statistical performance or be very meaningful clinically for neonates.

The AP-DRGs are structurally similar to the APR-DRGs in a number of respects, somewhat different in other respects, and entirely different in still other respects. They both use age to define neonates (MDC 15) and both use birth weight as an initial grouping variable. They both use surgery as a grouping variable but do so differently. They both have major problem/other significant problem/other problem diagnoses lists but these lists are quite different. They are also used differently. The APR-DRGs use them to form severity (costliness) subclasses for each DRG category. The AP-DRGs do not form severity subclasses but do provide separate DRG categories for neonates with multiple major problems who are >1500 g. Both systems use mechanical ventilation but do so differently. Both systems use LOS combined with transfer-out status to create DRG categories for neonates who are early triage and transfer.
<table>
<thead>
<tr>
<th>Structural Elements</th>
<th>Medicare DRGs Version 12.0</th>
<th>AP-DRGs Version 12.0</th>
<th>APR-DRGs Version 12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition of Neonate (ie, definition of MDC 15)</td>
<td>Newborn V30 code, perinatal diagnosis code, and a subset of approx. ten congenital anomaly diagnoses.</td>
<td>Age &lt;29 days at admission.</td>
<td>Age &lt;29 days at admission.</td>
</tr>
<tr>
<td>Use of birth weight</td>
<td>No</td>
<td>Yes, six birth weight ranges.</td>
<td>Yes, six birth weight ranges.</td>
</tr>
<tr>
<td>Use of surgery Types of surgical lists</td>
<td>No, N/A</td>
<td>Yes, for neonates &gt;1000 g. Significant OR procedures; Minor abdominal OR procedures.</td>
<td>Yes, for all neonates. Major OR procedures; Other OR procedures; Minor OR procedures.</td>
</tr>
<tr>
<td>Major/extreme problem diagnosis lists Multiple major problem classifications</td>
<td>Yes</td>
<td>Yes, for neonates &gt;1500 g, with combined major/extreme problem list.</td>
<td>Yes, for all neonates, with separate lists for major diagnosis problem and extreme diagnosis problem.</td>
</tr>
<tr>
<td>Other significant problem diagnosis list</td>
<td>No, is default category for all neonates not classified as major problem or normal newborn.</td>
<td>Yes, is “other problem” and “minor problem” categories.</td>
<td>Yes, is “other significant problem” category.</td>
</tr>
<tr>
<td>Other problem diagnosis list</td>
<td>No</td>
<td>No</td>
<td>Yes, through subclasses to differentiate from normal newborns.</td>
</tr>
<tr>
<td>Reevaluation of diagnosis problem lists</td>
<td>No</td>
<td>Limited</td>
<td>Comprehensive</td>
</tr>
<tr>
<td>Normal newborns</td>
<td>Newborns whose only diagnosis is a V30 newborn code or whose only other diagnoses are from a list of several dozen very minor diagnoses.</td>
<td>Default category for newborns who do not have diagnoses assignable to other categories.</td>
<td>Default category for newborns who do not have diagnoses assignable to other categories.</td>
</tr>
<tr>
<td>Severity subclasses</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of severity subclasses</td>
<td>N/A</td>
<td>N/A</td>
<td>4</td>
</tr>
<tr>
<td>Interactive effect of multiple diagnoses in assigning severity subclass</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td>Use of Non-OR procedure, mechanical ventilation</td>
<td>No</td>
<td>Yes, as one of the criteria for major problem categories which exist for neonates &gt;1500 g.</td>
<td>Yes, as severity subclass modifier for all neonates, with differential weighting based on neonate’s birth weight and whether duration &gt;96 hours.</td>
</tr>
<tr>
<td>Use of LOS</td>
<td>No</td>
<td>Yes, for transfer-out patients and early death patients.</td>
<td>Yes, for transfer-out patients only.</td>
</tr>
<tr>
<td>Use of death</td>
<td>Yes</td>
<td>Yes, for patients who die within first day of life and also those who are &lt;1500 g.</td>
<td>Not used in definition of categories so that mortality rates within each category can be examined.</td>
</tr>
<tr>
<td>Risk of mortality subclass</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of mortality subclasses</td>
<td>N/A</td>
<td>N/A</td>
<td>4</td>
</tr>
<tr>
<td>Interactive effect of multiple diagnoses in assigning mortality subclass</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of base DRG categories</td>
<td>7</td>
<td>28</td>
<td>43 (Version 12.0) 35 (Version 15.0)</td>
</tr>
<tr>
<td>Total number of DRG categories (with multiple major problem or severity subclass breakouts)</td>
<td>7</td>
<td>34</td>
<td>172 (Version 12.0) 140 (Version 15.0)</td>
</tr>
</tbody>
</table>

The version 12.0 DRG classification systems are those which were applicable for the ICD-9-CM diagnosis codes in effect for the time period October 1994–September 1995 and were applied to this study’s databases. Each October 1, each DRG classification system needs to be updated to read the annual updates to the ICD-9-CM diagnoses codes. In addition, sometimes the DRG system is substantively updated. Of the above three DRG classification systems, the only substantive update to neonatal category definitions occurred with the version 15.0 APR-DRGs which were applicable for October 1997–September 1998.
patients. AP-DRGs use death as a grouping variable. APR-DRGs do not use death as a grouping variable so that death can be used as an outcome variable.

The number of base DRG categories is 28 in AP-DRGs and the total number of categories is 34 with multiple major problem breakouts for certain of the base AP-DRGs. The number of base DRG categories is 35 for APR-DRGs and the total number of categories is 140 with 4-way severity subclass breakouts for all the base APR-DRGs. The APR-DRG system maintains a large number of categories for neonatal patients for three reasons: 1) to generally increase the clinical meaningfulness of the DRG categories; 2) to maintain a consistent 4-tiered severity subclass hierarchy even although there are some low-volume cells (eg, neonate with major surgery, subclass one); and (3) to support risk of mortality as well as resource use applications (eg, separate categories for neonate with major anomaly/hereditary condition and neonate with respiratory distress syndrome).

The APR-DRGs for neonates are comprehensively reviewed and updated with each substantive update to the APR-DRG system, which occurs every second or third year. The AP-DRGs for neonates are not substantially updated on a regularly scheduled basis but there have been updates.

Following is a summary of how one of the DRG systems, the APR-DRG, actually groups a neonatal patient, and then a detailed comparison of the structural characteristics of each DRG system. Refer to Appendix 3 for a full listing of the APR-DRG categories.

**Step 1: Prebirth Weight APR-DRG Assignment**

If a neonate is transferred-out within the first 4 days of life, receives an organ transplant or extracorporeal membrane oxygenation, the neonate is assigned to one of the pre-birth weight APR-DRGs.

**Step 2: Birth Weight Range Classification**

If not assigned to a prebirth weight APR-DRG, the neonate is classified to one of six birth weight ranges.

**Step 3: Surgical Versus Medical Classification**

All neonates assigned to a birth weight range are next classified as either surgical or medical. For neonates <2500 g (5.5 pounds) this includes only those with a major OR procedure. For neonates >2499 g, this includes neonates with any OR procedure.

**Step 4: Surgical APR-DRG Assignment**

All neonates classified as surgical are assigned to a specific APR-DRG in their birth weight range. For neonates >2499 g there are three surgical APR-DRGs numbered in hierarchical order. If a patient has more than one OR procedure, the patient is assigned to the highest category in the APR-DRG surgical hierarchy.

**Step 5: Medical APR-DRG Assignment**

All neonates classified as medical are assigned to a specific APR-DRG in their birth weight range: because most neonates do not have a principal diagnosis in the usual sense of the term, a hierarchy among medical APR-DRGs must be specified for patients with multiple significant diagnostic conditions. Principal diagnosis is defined by the UHDDS to be “that condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital for care.” For newborns, the principal diagnosis code is a V30-V39 code indicating newborn birth as this is what led to the hospital admission. This provides no information about medical problems and because there is no significance to the sequencing of secondary diagnoses it is necessary to create a hierarchy among medical APR-DRGs. If a diagnosis is not used to assign the neonate to a medical APR-DRG, it is available to be used in the severity subclass algorithm.

**Step 6: Assignment to APR-DRG Severity Subclass**

Once a patient is assigned to an APR-DRG, then the severity subclass assignment is made based on all the diagnoses, interactions between multiple diagnoses, and select non-OR procedures.

**Step 7: Assignment to APR-DRG Risk of Mortality Subclass**

Same process is followed as for severity subclass assignment, except that the subclass values assigned to individual diagnoses are often different.

The Medicare DRGs define neonates (MDC 15) by the presence of certain diagnoses codes. There must be either an ICD-9-CM code indicating that the patient is a newborn (codes V30.XX–V39.XX), or a perinatal diagnosis code (codes 760.XX–779.XX), or one of approximately 10 congenital anomaly diagnosis codes. This definition causes certain patients to be excluded from MDC 15 that are considered neonates by the conventional definition of age 0 to 28 days and causes other patients to be included in MDC 15 who are older than 28 days. The neonatal patients most frequently excluded from Medicare MDC 15 are full term neonates who are either transfer-in or readmission patients. Many of the infants who require surgery in the first month of life are full-term babies who are transferred to a facility that can perform neonatal surgery and wind up in a non-neonatal MDC such as the digestive or circulatory MDC. The AP-DRGs and APR-DRGs both define neonates based on age 0 to 28 days at time of admission. The diagnosis codes have no bearing on the definition of a patient as a neonate.

The Medicare DRGs do not use birth weight as an initial grouping variable. Instead, the very broad diagnosis codes for extreme prematurity and other prematurity are used. The AP-DRGs and APR-DRGs both use birth weight (six ranges) as the initial grouping variable.

The Medicare DRGs do not use surgery as a grouping variable for MDC 15 patients. The AP-DRGs use surgery as a grouping variable but only for neonates in birth weight ranges >1000 g (2.2 pounds). The surgical category includes all OR procedures except minor abdominal procedures, with a 2-way breakout for multiple major problems for neonates >1500 g (3.3 pounds). For neonates >2499 g (5.5 pounds) there is also a category for minor abdominal procedures. This is different from the APR-DRGs, which...
define a list of major OR procedures and create a major surgery category for neonates in all birth
weight ranges, with a 4-way severity subclass break-
out. The APR-DRGs also create another surgery
category for neonates >2499 g (5.5 pounds).

The Medicare DRGs do not have severity (costli-
ness) subclasses for neonates. The AP-DRGs also do
not have severity subclasses but do provide separate
DRG categories for neonates with multiple major
problems for those in birth weight ranges >1500 g
(3.3 pounds). The APR-DRGs have a 4-way severity
subclass breakout that is applied to all DRG catego-
ries in all birth weight ranges. The severity subclass
algorithm considers all the patient’s diagnoses in-
cluding interactive effects of multiple diagnoses.

The Medicare DRGs have a major problem diagno-
sis list for neonates that it applies to create two DRG
categories, preterm neonate with major problem and
full-term neonate with major problem. The list has not
been updated in any significant way since its
initial implementation in fiscal year 1984. The AP-
DRGs have updated the major problem diagnosis list
and have applied the list to create DRG categories for
neonates with major problems and multiple major
problems, but only for neonates in birth weight
ranges >1500 g (3.3 pounds). The APR-DRGs have
also updated and use a major problem diagnosis list
for neonates, which it further differentiates into ma-
jor problems and extreme problems. For example,
neutropenia and thrombocytopenia are considered
major problems and disseminated intravascular co-
agulation is considered an extreme problem. All the
diagnosis lists and the severity subclass algorithms
are comprehensively reviewed and updated with ev-
ery substantive update of APR-DRGs, which occurs
every second or third year.

The Medicare DRGs do not have an other significant
problem list per se, but instead define “significant
problem” neonates as the default category for full-
term neonates who are not assigned to either the
major problem or normal newborn category. It con-
tains many neonates who on close inspection appear
in birth as normal newborns. The AP-DRGs provide DRG
categories for “minor problems” and “other prob-
lems” for neonates >1500 g (3.3 pounds). The diag-
noses included in these categories roughly corre-
spond to the “other significant problem” categories
in the APR-DRG system, except that they are less
inclusive.

The Medicare DRG approach to defining normal
newborns is entirely different from the AP-DRGs and
APR-DRGs. The Medicare DRGs define normal new-
borns as those newborns whose only diagnosis is a
V30 to V39 newborn code or whose only other diag-
noses are from among a list of several dozen very
minor diagnoses (eg, 7746 unspecified fetal/neonatal
jaundice). If a newborn has any other of the 12 766
ICD-9-CM diagnosis codes and is not classified as
major problem or premature, it is assigned to the
default category of other significant problem. For
example, diagnoses such as transient tachypnea and
hypoglycemia will get a neonate assigned to the
Medicare DRG of other significant problem. The re-
result of this approach is that the neonates classified as
normal newborns are almost all truly normal new-
borns, but many other normal newborns wind up
classified as other significant problem. In contrast,
the AP-DRGs and APR-DRGs define normal new-
borns to be a default category for newborns who do
not have a diagnosis assignable to one of the other
more specific problem categories.

Medicare DRGs do not use LOS as a grouping
variable. The AP-DRGs and APR-DRGs both use
LOS to create categories for neonates transferred to
another acute hospital in the first 4 days of life so as
to identify early triage and referral neonates. There
are other neonates who are transferred to another
acute hospital at a later point in time but neither
classification system addresses this. To illustrate,
there are low birth weight and other complex neo-

nates who receive several weeks or months of care at
a tertiary facility and are then sometimes transferred
back to a community hospital closer to home for
growth and development care. The acute hospital-
ization episode is thereby split up between several
hospital stays. This often reflects local geographic
delivery system factors, which can not be ad-
dressed through a diagnostic classification system.

The transfer and back-transfer issues are very impor-
tant but have to be addressed through policies and
methodologies tailored to the local environment for
specific applications such as prospective payment or
outcomes analysis.

The Medicare DRGs use death as a grouping vari-

able. It combines into one DRG all neonatal deaths, a
very heterogeneous group, with all neonates trans-
ferred to another acute hospital, another very heter-
ogeneous group. Among neonates who die there are
several prominent subgroups. There are newborns
who are judged to be nonviable and receive only
comfort measures and often die within the first or
second day of being born. These are mostly ex-
tremely premature infants and newborns with cer-
tain very severe anomalies. There is another group
that receives medical or surgical treatment but who
nonetheless die, often within 1 week or several
weeks of birth. Then there is another group that dies
after many weeks or months of life. The Medicare
DRGs place all these neonates together with those
who are transferred to another acute hospital.

The AP-DRGs create five categories for neonates
who die. There are two categories for neonates who
die within the first day of life and three categories for
other neonates who die and are in the lower birth
weight ranges. The APR-DRGs do not use death as a
grouping variable. To be able to examine death as an
outcome within the individual DRG categories, it is
necessary to not use death to define the categories.
For payment system or other purposes it may be
indicated to consider separately neonates who die,
but this can still be done and tailored to the specific
application. This can be done having the benefit of all
the information from the classification system to un-
derstand the characteristics of neonates who die.

Related to the differences with the use of death as a
grouping variable, only the APR-DRGs have risk of
mortality subclasses. These subclasses consider all a
patient’s diagnoses and the interactive effect of mul-

308 SUPPLEMENT
multiple diagnoses to predict the likelihood of a patient dying. The assignment of diagnoses to risk of mor-
tality subclasses is sometimes the same but is often
different from that for severity (costliness) subclasses.
The reason for this is that some diagnoses may indicate a patient is likely to need a lot of
treatment, but is not likely to die. Other diagnoses may indicate a high likelihood of dying and resource
use might be lower for that reason.

Finally, the three DRG systems differ with respect
to the use of the mechanical ventilation. The Medicare
DRGs do not classify neonates based on the use of
mechanical ventilation although they do for older
patients. The AP-DRGs use mechanical ventilation as
an additional means to identify neonates with a ma-
jor problem when it differentiates based on major
problems (birth weight ranges >1500 g). The APR-
DRGs use mechanical ventilation as a severity sub-
class modifier with differential weighting based on a
neonate’s birth weight and whether the duration of
mechanical ventilation is ≥96 hours, a distinction
available in the ICD-9-CM procedure codes. For ne-
onates <1000 g (2.2 pounds), the code for mechanical
ventilation <96 hours does not add much informa-
tion to what is already known about the infant’s
health condition and therefore is not used to modify
the severity subclass. For larger infants, particularly
those >2499 g, mechanical ventilation <96 hours is
very distinguishing of the infant’s health condition
and therefore is considered as a possible modifier to
the severity subclass. The code for mechanical ven-
tilation ≥96 hours adds information about the health
condition of all neonates and is considered as a pos-
sible modifier to the severity subclass. It is especially
distinguishing among larger infants and is given
particular weight for them as a severity subclass
modifier.

In sum, the structure, definitions and logic for the
classification of neonates are very different for Medi-
care DRGs, AP-DRGs, and APR-DRGs. The process
for review and refinement of each DRG system is
also very different. It is therefore reasonable to ex-
pect that the statistical performance and clinical util-
ity of the three DRG systems will also be consider-
ably different.

DATABASE FOR COMPARATIVE ANALYSIS OF
DRG CLASSIFICATION SYSTEMS

The database used for this study was a calendar
year 1993 hospital medical record abstract discharge
database that included 675 acute general hospitals and
40 freestanding acute children’s hospitals. To be
nationally representative, the study sample frame
included all patients from the 675 general hospitals
and a 20% random sample from the 40 children’s
hospitals.

The 675 general hospitals were generally represen-
tative of acute general hospitals in the United States
with respect to bed size, teaching status, and urban/
rural location although there was a slight underrep-
resentation of rural hospitals. The 675 general hospi-
tals included a generally representative number of
hospitals from the four Census Bureau regions of the
country with the exception of the Northeast, which
was underrepresented. The 40 children’s hospitals
were very representative of freestanding acute chil-
dren’s hospitals in the United States. The edited data
set for this study contained 4 203 646 inpatient dis-
charges of which 492 558 were newborns and other
neonates (age <29 days at admission).

The study database was built for NACHRI by
HCIA. It included all children’s hospitals participat-
ing in NACHRI’s Case-Mix Comparative Reporting
Program and all general hospitals for whom HCIA
had medical record abstract data at the time the
database was created. HCIA collects the general hos-
pital data from a variety of sources including state
agencies, state hospital associations, and individual
hospitals. All data must pass a series of clinical,
demographic and financial edits developed by the
NACHRI Case-Mix Comparative Reporting Pro-
gram. The data from each hospital include the
UHDDS data elements plus total charges, admission
source, and birth weight if reported as a separate
data element by the hospital. The database also in-
cluded many calculated variables such as LOS, area
wage-adjusted charges, area wage-adjusted costs,
HCFA DRG, AP-DRG, and APR-DRG.

Four of the variables in the study database are
measures of hospital resource use: LOS, total
charges, area wage-adjusted charges, and area wage-
adjusted costs. The measure selected for most of the
study’s analyses was area wage-adjusted costs. This
was selected because, of the measures available, it
best represented total resource use. Two sets of cal-
culations are performed to create this variable. First,
total charges are converted to area wage-adjusted
charges by applying the HCFA area wage index. This
adjusts for differences in a hospital’s charge structure
that relate to wage levels in its locale. HCFA
publishes this information each year as part of its
rule-making process to update Medicare hospital
progressive payment system rates. Second, area
wage-adjusted charges are converted to area wage-
adjusted costs by applying a hospital-wide ratio of
costs-to-charges (RCC). This adjusts for overall hos-
pital-wide differences from one hospital to another
in the markup of charges over costs. The RCCs are
obtained from HCFA’s hospital Medicare cost report
files that are updated annually and made publicly
available. The specific set of hospital-wide RCCs se-
lected for this study were the operating plus capital
costs RCCs. These RCCs include all Medicare allow-
able hospital costs except graduate medical educa-
tion (GME). GME costs were excluded to avoid the
skewing effect of these costs when comparing teach-
ing hospitals and nonteaching hospitals. It should be
noted that these cost figures contain hospital costs
only; no physician costs are included.

As a measure of total resource use, area wage-
adjusted operating plus capital costs are far more
accurate for comparisons across hospitals than total
charges. It adjusts for three major sources of noncom-
parability: area wage differences, overall hospital-
wide charge-to-cost markup differences, and GME
costs. Still, it is only a rough approximation of total
resource use and will tend to underestimate the true
costs of neonates treated in intensive care units and

Downloaded from by guest on August 30, 2017
readmission neonates treated in infant-toddler or other medical-surgical patient care units.

An extensive study on the effect of different cost accounting methods has identified that cost accounting and pricing methods commonly used by acute general hospitals tend to underestimate the true cost of caring for children, especially young children, those treated in a pediatric or neonatal intensive care unit, and those with serious congenital and chronic health conditions. The reasons for this are many and very complicated. In brief, the actual unit costs of acute care hospital services for children tend to be higher than for adult patients but is often not broken out discretely in the hospital’s cost accounting and pricing methods. This is often the case for patient care services such as nursing and respiratory care, and to a lesser extent, for ancillary services and for indirect service costs allocated using statistics from all age patients combined (eg, social services, plant operations, general administrative). The tendency toward average costing (versus more discrete costing) is strongly reinforced by the Medicare hospital cost reporting system that, for example, averages the nursing care costs of all intensive care unit patients and averages the nursing care costs of all medical-surgical unit patients.

To simplify the data analysis for this study, all patients with the discharge destination of transferred to acute hospital or left against medical advice were removed from the edited database unless they were assigned to an APR-DRG defined on that basis. These are patients with an incomplete hospitalization so it is not reasonable to expect a DRG system to predict how much care a patient receives before being discharged. For any payment system application it is of course necessary to develop payment policies for these patients or exclude them from prospective per discharge payment. This is an especially important issue for neonates who are often transferred to tertiary facilities for diagnosis and treatment and are then sometimes transferred back to the community hospital for growth and development care. It is also important to include transfer patients in any delivery system or outcomes analysis that is examining the total cohort of neonates.

**EXPLANATION OF VARIANCE (R²) STATISTICS FOR RESOURCE USE AND MORTALITY**

One of the most commonly used statistics to measure the performance of DRG classification systems is reduction of variance (R²), often referred to as explanation of variance. It is also a commonly misunderstood statistic. R² provides a summary measure of the extent to which a DRG system is able to predict the value of a dependent variable such as resource use or mortality, based on the characteristics of individual patients. A technical explanation of how the R² statistic is calculated including the actual formula is provided in Appendix 4.

In simplified terms, the denominator in the R² equation is the total variation in the dependent variable for all patients in the database. The numerator is the amount of total variation that can be explained by classifying each patient into a DRG category. If, for example, 40% of total variation can be reduced or explained by assigning a patient to a DRG category, then the R² equals 0.400. If, for example, 60% of total variation can be reduced or explained, then the R² equals 0.600. If 100% of total variation could be explained, then the R² would equal 1.00.

The denominator in the R² formula, total variation, is calculated by summing the square of the difference between each individual patient value and the average patient value for all patients in the database. The numerator is calculated by summing the square of differences between each individual patient value in a DRG category and the average value for all patients in the same DRG category. It is important to realize that because the R² formula calculates variation by summing the square of differences, it is very sensitive to extreme values. In other words, if there are subgroups of patients with predictably very high costs and these patients can be classified into their own DRG categories, a particularly high R² value can be generated. This would suggest that neonatal and circulatory MDCs might achieve high R² values, neonatal MDCs because of the predictably very high cost of extremely premature infants and surgical neonates and circulatory MDCs because of the predictably high costs of certain types of cardiovascular surgery.

There are limits to how high an R² a DRG classification system is likely to achieve. In the instance of costs there are at least five reasons why it is unlikely for a DRG system to generate an R² that is close to 1.00. First, the available diagnostic and procedure code information is not always as specific and precise as might be desired. Second, even if the codes always had the ideal specificity and coding practices by physicians and hospital personnel were perfect, it is not possible to predict the exact course for each patient. Third, there are differences in physician and hospital practice patterns that affect costs. Fourth, there are differences in hospital operating and capital costs. Fifth, there are limitations to the precision with which costs are measured for individual patients by existing hospital cost accounting and pricing methodologies. The first three of these reasons are also applicable constraints to the ability of a DRG classification system to achieve a high R² for explaining mortality. Actually, because death only occurs for a small subset of patients, one might expect the level and type of information from diagnosis codes alone to be more of a constraint. In other words, the R² achievable by DRG systems for mortality might be less than that which is achievable for costs.

Table 3 provides cost R² statistics for neonatal patients. It also provides cost R² statistics for all age patients to enable a perspective as to whether the different DRG classification systems seem to perform better or worse for neonatal patients than other patients. Finally, it provides LOS R² statistics for all age patients to show the relationship between LOS R² and cost R² statistics.

Table 3 shows that there are very large differences between the three DRG systems in their overall ability to explain variation in resource use. This is true for all age patients but is especially dramatic for
neonatal patients. For all age patients, the cost $R^2$ is 0.408 for Medicare DRGs, 0.468 for AP-DRGs, and 0.531 for APR-DRGs. For neonatal patients, the cost $R^2$ is extremely low at 0.292 for Medicare DRGs, increases to 0.507 for AP-DRGs, and to 0.627 for APR-DRGs. So, the cost $R^2$ for neonates is much lower than for other patients in Medicare DRGs, increases to a value a little higher than for other patients in AP-DRGs, and increases to a value that is much higher than for other patients in APR-DRGs. The increase in $R^2$ for AP-DRGs is thought to be attributable to the separate categorization for low birth weight neonates, multiple major problem neonates, and surgical neonates. The additional increase in $R^2$ for APR-DRGs is thought to be attributable to the more refined set of base categories and the 4-tiered severity subclasses. The neonatal MDC $R^2$ of 0.507 in AP-DRGs is the second highest among the 25 MDCs with circulatory the highest at 0.533. The neonatal MDC $R^2$ of 0.627 in APR-DRGs is the highest followed by circulatory at 0.571.

Table 3 shows as might be expected that the LOS $R^2$ is less than the cost $R^2$ in all DRG systems. For all age patients, the LOS $R^2$ is 0.313 for Medicare DRGs, 0.369 for AP-DRGs, and 0.421 for APR-DRGs. These $R^2$ values are all three-quarters to four-fifths of the cost $R^2$ values. The reason for this is that DRG systems are intended to explain variation in total resource use, ie, costs. A patient’s LOS represents a major component of a patient’s total resource use but there are many other components of costs as well. To illustrate, a surgical patient may not have an extremely long LOS but may still have fairly high costs given the intensity of service while an inpatient. Because the measure of resource use that DRG systems are intended to predict is cost, this article focuses on costs although certainly additional LOS analyses might also be insightful.

This study also produced cost $R^2$ statistics for a trimmed data set, that is, with outliers removed. This was done to show how much the $R^2$ statistics can be effected by removing extreme or unusual patients. For this analysis, it was the intent to remove only very extreme outliers. High-cost outliers were defined as the highest 1% of patients in each DRG category in each DRG system. Low-cost outliers were defined as the lowest 1/2% of patients in each DRG category in each DRG system. Approximately 7400 of the database’s 492,558 neonates were removed from the trimmed data set for each DRG system, although it is important to note it was a different group of 7400 neonates identified as extreme outliers for each DRG system.

The $R^2$ results from the trimmed data set show the same overall pattern as the untrimmed database. The $R^2$ values are 0.357, 0.573 and 0.655, respectively for Medicare DRGs, AP-DRGs, and APR-DRGs. There is also another important pattern. The lower the predictive performance of the DRG system with the untrimmed data set, the greater its improvement in the trimmed data set. The $R^2$ improves from 0.292 to 0.357 for Medicare DRGs, from 0.507 to 0.573 for AP-DRGs, and from 0.627 to 0.655 for APR-DRGs. The primary reason for this is that the outlier patients are even more extreme in the poorer performing DRG systems. So although the outlier definition used in this study removes 1% of high-cost patients from each DRG system, there is more unexplained variation removed in the instance of the poorer performing DRG systems. It is important to be aware of the effect of trimming outliers from a data set as many comparative analyses are done from trimmed data.

The purpose of the $R^2$ statistical analysis for mortality is to compare the overall ability of the three DRG systems to explain variation in inpatient hospital mortality. The DRG systems, in particular the APR-DRGs, are increasingly being used or considered for use for this purpose. Numerous states are now publishing hospital risk-adjusted mortality rates to permit consumers to compare hospital outcomes. For example, the State of Florida Agency for Health Care Administration included an APR-DRG-based mortality analysis as part of its 1996 Guide to Hospitals in Florida. The Medicare DRGs and AP-DRGs systems were developed for predicting resource use and were not really intended for mortality prediction. In these systems mortality is one of the variables used to define the DRGs (ie, patients are assigned to certain DRGs depending on whether they lived or died). This is not appropriate for a mortality prediction model because it would be circular logic to use mortality to predict mortality. Therefore, for the mortality analyses, the data were regrouped eliminating all mortality distinctions, (ie, patients are grouped into the DRG to which they would have been assigned if the patient had not died). Because APR-DRGs do not use mortality as a grouping variable, regrouping was not necessary for APR-DRGs.

The APR-DRGs have a separate set of severity subclasses that group patients based on risk of mortality. These APR-DRG risk of mortality subclasses were used for the mortality analyses instead of the APR-DRG severity subclasses that were used for the analyses of cost and LOS.

Mortality in the hospital is used in this analysis.
Mortality in the hospital is commonly available in hospital administrative records, but if the patient dies the day after discharge from the hospital, this would not be reflected in the hospital’s records. Ideally, mortality subsequent to discharge would have been merged with the data, but this information was not available.

The $R^2$ for mortality is computed by assigning each patient a value of 0 or 1 indicating whether they were discharged alive or dead, respectively. The predicted mortality for the patient is equal to the average value of the 0/1 variable in the DRG to which the patient is assigned. The average value of the 0/1 value is equivalent to the fraction of patients who die in the DRG. Based on the 0/1 variable, the $R^2$ for mortality is computed in the same manner as the $R^2$ for cost or LOS.

The $R^2$ for mortality for all age patients was 0.108 for Medicare DRGs, 0.1507 for AP-DRGs, and 0.2638 for APR-DRGs. The $R^2$ for mortality for neonates was 0.083 for Medicare DRGs, 0.304 for AP-DRGs, and 0.4155 for APR-DRGs. The improvement in mortality $R^2$ for AP-DRGs is thought to be attributable primarily to the low birth weight categories and secondarily to the categories for multiple major problems and surgical neonates. The additional improvement in mortality $R^2$ for APR-DRGs is thought to be attributable primarily to the risk of mortality subclasses and secondarily to the refined base categories, in particular, those for major anomalies and hereditary conditions.

As risk of mortality analysis is a more recent application of DRG systems than cost analyses, there is not as full of an understanding of how it works and its most appropriate interpretation. In addition, in the case of the APR-DRGs, there have been many revisions introduced to the risk of mortality subclasses in version 15.0 APR-DRGs released in the spring of 1998. Finally, it should be noted that a DRG-based mortality model is quite different from some of the other neonatal mortality models that may include additional variables such as gestational age (along with birth weight), Apgar score (at specified time intervals), blood gas values (at specified time intervals), presence of prenatal care, or certain specific types of prenatal care. Some of the other neonatal mortality models also differ in using only certain types of diagnoses or only diagnoses within a certain time period after birth. It would be useful to explore these differences further in future analyses.

In sum, the differences in overall predictive power are very large. The Medicare DRGs have very modest power for predicting the costs of neonatal inpatient care or inpatient mortality. The APR-DRGs have much greater predictive power, in fact, more than double for costs and five times greater for mortality. The overall predictive power of AP-DRGs is intermediate to the Medicare DRGs and APR-DRGs, although somewhat closer to APR-DRGs. The next section examines how well each DRG system predicts costs for specific groups of neonates and hospitals.

### PAYMENT IMPACTS AND SYSTEMATIC EFFECTS

**BY PATIENT TYPE AND HOSPITAL TYPE**

A key question in evaluating the performance of alternate DRG classification systems is whether they predict equally well the cost for all groups and subgroups of patients and hospitals. To the extent that patients of certain age ranges or service lines or other attributes have costs that are greater or less than predicted, there is a systematic bias. If certain hospitals predominate in the care of patients whose resource needs are systematically understated, then these hospitals will likewise have their case-mix intensity understated and will be underpaid in a payment application.

The statistical measure used in this study to evaluate the systematic effects or biases of each DRG classification system is a ratio of simulated payment to actual cost for various groupings and subgroupings of patients and hospitals. A ratio $>1.00$ indicates that the payment is greater than the actual cost. A ratio $<1.00$ indicates that the payment is less than the actual cost.

Essentially, the hypothesis being tested is that the AP-DRGs and especially the APR-DRGs will show a payment to actual cost ratio closer to 1.00 than the Medicare DRGs for various groups and subgroups of patients and hospitals. To the extent this is true, they provide a fairer and less biased classification of patients. For the comparative analysis of payment impacts for neonates, a neonate is defined as any patient whose age at time of admission is less then 29 days regardless of how classified by the DRG system.

One caveat is in order. This statistical testing is not intended to represent the testing of a full prospective payment model. For this, additional payment policies would have to be tested, most notably for outlier patients and for transfer patients if they are included in the prospective payment system. Facility level adjustments would also need to be considered. Rather, it is the purpose of this analysis to test the classification system component of a prospective payment model. Secondly, it will provide insight as to where payment system adjustments may be most needed and whether they seem less important when the AP-DRGs or APR-DRGs are used.

Table 4 presents the overall results for neonates as well as separate results for surgical neonates, medical neonates, and normal newborns.

The ratio of payment to costs for all neonates is 0.899 for Medicare DRGs and increases to 0.994 and 1.000 for AP-DRGs and APR-DRGs. This is because the Medicare DRGs define neonates by the presence of certain newborn and perinatal diagnosis codes and place certain very expensive neonates in non-neonatal DRGs. The AP-DRGs and APR-DRGs define neonates by age (0–28 days at admission) and so the sum of predicted costs for all neonates will be approximately equal actual costs.

The ratio of payment to costs varies dramatically by service line. The change is most dramatic for surgical neonates whose payment ratio increases from an extremely low value of 0.318 for Medicare DRGs to 0.869 for AP-DRGs and 1.004 for APR-DRGs. From
the standpoint of systematic risk, this is very important because only a very small number of hospitals offer neonatal surgical services. The opposite pattern presents for normal newborns whose payment ratio for Medicare DRGs is very high at 1.231 and decreases to .988 for AP-DRGs and 1.016 for APR-DRGs. In Medicare DRGs, many neonates who are really normal newborns are grouped to other significant problems, and as a result, the payment is much higher than actual cost. Although these newborns are relatively inexpensive and dollar differences per case are small, total case volume is very large so total dollar volume is significant.

The classification of normal newborns is presented in more detail in Table 5. In Medicare DRGs, 65.3% of neonates are classified as normal newborns. In APR-DRGs, 86.0% of neonates are classified as normal newborns. In AP-DRGs, 78.3% are classified as normal newborns, and another 7.5% are classified in a subclass entitled other problem that is intermediate to normal newborn and other significant problem.

Based on a review of the case counts, costs, and diagnosis codes for neonates classified as normal newborns and the contiguous DRG categories, it is clear that the Medicare DRGs are classifying a substantial number of neonates as having significant problems who really are normal newborns. An opposite pattern presents with the AP-DRGs where it appears that many neonates classified as normal newborns really have problems that properly should place them in a category intermediate to normal newborn and other significant problem. The most accurate classification of normal newborn is provided by the APR-DRGs.

Table 6 presents another very striking pattern. Neonates who are admitted from another acute hospital (transfers-in) have an extremely low payment ratio of 0.548 with Medicare DRGs. This improves to 0.816 with AP-DRGs and to 0.948 with APR-DRGs. This again is very important from the viewpoint of systematic risk. In an area such as neonatology where there is a relatively high regionalization of services,
the DRG system needs to be able to describe fully the characteristics of patients who are transferred to tertiary facilities for diagnosis and treatment.

Table 7 identifies two important patterns. Neonates who are discharged from the hospital with home health services have an extremely low payment ratio of .565 with Medicare DRGs. This improves to 0.845 with AP-DRGs and to 0.899 with APR-DRGs which is a significant improvement but still short of the desired level of 1.00.

The second important pattern by discharge destination is for neonates who die. Their payment ratio is extremely low with Medicare DRGs at 0.494, increases to 0.843 with AP-DRGs and then to 1.495 with APR-DRGs. The best results are for AP-DRGs that might be expected because the majority of neonates who die are classified into one of the two AP-DRGs for neonates died within first day of life or one of the three neonate died categories for low birth weight neonates. The high level of overpayment with the APR-DRGs represents that more often than not, neonates who die cost less than other neonates in the same APR-DRG category. From a payment system perspective, this suggests two options to consider: option one, remove neonates who die from prospective per discharge payment and pay on another basis such as percent of charges; or option two, develop a short stay/low-cost outlier policy. Option one is simplest and probably fairest. Option two is somewhat complex because some early death neonates receive comfort only measures and other early death neonates receive major surgery and other expensive interventions.

Table 8 identifies important patterns of systematic risk that are significantly but not entirely resolved as the AP-DRGs and then the APR-DRGs are applied.

The payment ratios are by far the lowest for free-standing acute children’s hospitals with a value of 0.440 for Medicare DRGs, 0.769 for AP-DRGs and 0.910 for APR-DRGs. This is consistent with the payment ratios in Tables 4 and 6 for neonates who are surgical and neonates who are admitted from another acute hospital. Because children’s hospitals generally do not offer an obstetric service, nearly all their neonates are transfers-in from another acute hospital or readmissions. Many are surgical patients. These are the

### TABLE 6. Comparison of Payment to Actual Cost for Three DRG Systems for Neonates by Admit Source

<table>
<thead>
<tr>
<th>Neonates by Admission Source</th>
<th>Case Totals</th>
<th>Ratio of Payment to Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Cases</td>
<td>Total Costs (000s)</td>
</tr>
<tr>
<td>Neonates (age 0–28 days):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admit from other acute hospital</td>
<td>7423</td>
<td>$148 593</td>
</tr>
<tr>
<td>Admit from a skilled nursing facility</td>
<td>976</td>
<td>$82 859</td>
</tr>
<tr>
<td>Admit from other source (most are newborn admissions, some are readmissions)</td>
<td>484 159</td>
<td>$909 438</td>
</tr>
<tr>
<td>Total</td>
<td>492 558</td>
<td>$1 066 290</td>
</tr>
</tbody>
</table>

This is simplified payment simulation with entire payment based on DRG weight and no outlier or facility level adjustments. Costs include hospital operating and capital costs, exclusive of graduate medical costs, adjusted by HCFA area wage index. Many of the neonates coded as admitted from a skilled nursing facility are probably miscoded. In all likelihood, the majority of these neonates were actually admitted from another acute hospital.

### TABLE 7. Comparison of Payment to Actual Costs for Three DRG Systems for Neonates by Discharge Destination

<table>
<thead>
<tr>
<th>Neonates by Discharge Destination</th>
<th>Case Totals</th>
<th>Ratio of Payment to Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Cases</td>
<td>Total Costs (000s)</td>
</tr>
<tr>
<td>Neonates (age 0–28 days):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharged to home'</td>
<td>482 608</td>
<td>$906 627</td>
</tr>
<tr>
<td>Discharged to home with home health services</td>
<td>5224</td>
<td>$71 571</td>
</tr>
<tr>
<td>Discharged died</td>
<td>2371</td>
<td>$67 929</td>
</tr>
<tr>
<td>Discharged to other destinations</td>
<td>2355</td>
<td>$20 163</td>
</tr>
<tr>
<td>Total</td>
<td>492 558</td>
<td>$1 066 290</td>
</tr>
</tbody>
</table>

This is simplified payment simulation with entire payment based on DRG weight and no outlier or facility level adjustments. Costs include hospital operating and capital costs, exclusive of graduate medical costs, adjusted by HCFA area wage index. There are no statistics above for neonates discharged to other acute hospital. This is because all patients discharged to other acute hospitals were removed from the study’s analytic data files to ensure that only patients with a complete acute care hospitalization experience were included. Discharged to other destinations includes skilled nursing facility, intermediate care facility, other institution, and left against medical advice. Four-fifths of these patients are discharged to “other institution” and on closer inspection it appears that many of these patients were probably discharged to other hospitals (but were miscoded), thus explaining why total payment exceed total costs. (ie, full payment for partial acute care hospitalization).
patients most inadequately classified by the Medicare DRGs and for whom the improvements in the AP-DRGs and APR-DRGs have the greatest impact.

The payment ratios for major teaching general hospitals ($N = 28$) are similar to freestanding acute children's hospitals but not as extreme. For this analysis, major teaching general hospitals were defined as those with a ratio of interns to residents to beds $\geq 0.25$ the definition commonly used by the US HCFA. The payment ratio is very low at 0.623 for Medicare DRGs, improves to 0.871 for AP-DRGs and to 0.892 for APR-DRGs. Major teaching general hospitals show a very large improvement for surgical neonates and transfer-in neonates, similar to that for freestanding acute children's hospitals. They also show a large improvement for medical neonates (excluding normal newborns), but not nearly as large as for freestanding acute children's hospitals. This is because as general hospitals they offer obstetric services and as a result see mildly and moderately ill medical neonates along with more extremely ill medical neonates.

The payment ratios in Table 8 for other urban general hospitals ($N = 413$) show a mixed pattern that varies by hospital bed size. The composite payment ratio for all other urban hospitals is very high at 1.119 for Medicare DRGs, decreases to 1.099 for AP-DRGs, and decreases to 1.068 for APR-DRGs. The composite pattern is one of overpayment by Medicare DRGs, moderated somewhat by AP-DRGs and APR-DRGs. In interpreting these and other hospital payment ratios it is important to remember that this is a simplified payment simulation with no payment adjustments for outlier patients or any facility level adjustments.

The composite pattern for other urban hospitals is actually not very meaningful because there are very distinctive patterns among subgroups of these hospitals. One way although certainly not the only way to categorize these hospitals is by total bed size. Table 9 shows the percent change in case-mix index for neonates by hospital type with breakouts by total bed size. Case-mix index measures the average cost weight for a subgroup of patients compared with that for a larger group of patients. In this instance, the average cost weight for neonatal patients at a

### Table 8: Comparison of Payment to Actual Cost for Three DRG Systems for Neonates by Hospital Type

<table>
<thead>
<tr>
<th>Hospital Type</th>
<th>Cases</th>
<th>Medicare DRGs</th>
<th>AP-DRGs</th>
<th>APR-DRGs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children's hospitals ($N = 40; 20%$ sample)</td>
<td>5751</td>
<td>1.910</td>
<td>3.349</td>
<td>3.961</td>
</tr>
<tr>
<td>Major teaching general hospitals ($N = 28$)</td>
<td>61.906</td>
<td>0.494</td>
<td>0.693</td>
<td>0.708</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $&lt; 100$ ($N = 84$)</td>
<td>15.869</td>
<td>0.230</td>
<td>0.165</td>
<td>0.162</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $100-199$ ($N = 115$)</td>
<td>64.508</td>
<td>0.208</td>
<td>0.181</td>
<td>0.170</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $200-299$ ($N = 96$)</td>
<td>87.630</td>
<td>0.237</td>
<td>0.216</td>
<td>0.201</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $300-449$ ($N = 75$)</td>
<td>100.478</td>
<td>0.309</td>
<td>0.299</td>
<td>0.289</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $\geq 450$ ($N = 43$)</td>
<td>99.842</td>
<td>0.409</td>
<td>0.450</td>
<td>0.447</td>
</tr>
<tr>
<td>All other rural general hospitals, beds size $&lt; 100$ ($N = 167$)</td>
<td>25.045</td>
<td>0.166</td>
<td>0.145</td>
<td>0.136</td>
</tr>
<tr>
<td>All other rural general hospitals, beds size $100-199$ ($N = 52$)</td>
<td>20.605</td>
<td>0.182</td>
<td>0.159</td>
<td>0.149</td>
</tr>
<tr>
<td>All other rural general hospitals, beds size $200-299$ ($N = 10$)</td>
<td>4921</td>
<td>0.173</td>
<td>0.151</td>
<td>0.142</td>
</tr>
<tr>
<td>All other rural general hospitals, beds size $300-449$ ($N = 5$)</td>
<td>6003</td>
<td>0.201</td>
<td>0.163</td>
<td>0.150</td>
</tr>
</tbody>
</table>

Case-mix index measures the average cost weight for a subgroup of patients compared with that for a larger group of patients. In this instance, the average cost weight for neonatal patients at each of the hospital types (numerator) is divided into the average cost weight for all patients from all hospitals (denominator). The case-mix indices for neonatal patients at each of the general hospital types is $<1.00$ because the majority of neonates are normal newborns.

Costs include hospital operating and capital costs, exclusive of graduate medical costs, adjusted by HCFA area wage index.

Children's hospitals include freestanding acute children's hospitals with separate Medicare provider numbers.

Major teaching hospitals include general hospitals with a ratio of interns and residents to beds $\geq 0.25$. 

### Table 9: Comparison of Percent Change in Case-Mix Index for Neonates by Hospital Type and Bed Size

<table>
<thead>
<tr>
<th>Hospital Type and Bed Size</th>
<th>Cases</th>
<th>Medicare DRGs</th>
<th>AP-DRGs</th>
<th>APR-DRGs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children's hospitals ($N = 40; 20%$ sample)</td>
<td>5751</td>
<td>1.910</td>
<td>3.349</td>
<td>3.961</td>
</tr>
<tr>
<td>Major teaching general hospitals ($N = 28$)</td>
<td>61.906</td>
<td>0.494</td>
<td>0.693</td>
<td>0.708</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $&lt; 100$ ($N = 84$)</td>
<td>15.869</td>
<td>0.230</td>
<td>0.165</td>
<td>0.162</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $100-199$ ($N = 115$)</td>
<td>64.508</td>
<td>0.208</td>
<td>0.181</td>
<td>0.170</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $200-299$ ($N = 96$)</td>
<td>87.630</td>
<td>0.237</td>
<td>0.216</td>
<td>0.201</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $300-449$ ($N = 75$)</td>
<td>100.478</td>
<td>0.309</td>
<td>0.299</td>
<td>0.289</td>
</tr>
<tr>
<td>All other urban general hospitals, beds size $\geq 450$ ($N = 43$)</td>
<td>99.842</td>
<td>0.409</td>
<td>0.450</td>
<td>0.447</td>
</tr>
<tr>
<td>All other rural general hospitals, beds size $&lt; 100$ ($N = 167$)</td>
<td>25.045</td>
<td>0.166</td>
<td>0.145</td>
<td>0.136</td>
</tr>
<tr>
<td>All other rural general hospitals, beds size $100-199$ ($N = 52$)</td>
<td>20.605</td>
<td>0.182</td>
<td>0.159</td>
<td>0.149</td>
</tr>
<tr>
<td>All other rural general hospitals, beds size $200-299$ ($N = 10$)</td>
<td>4921</td>
<td>0.173</td>
<td>0.151</td>
<td>0.142</td>
</tr>
<tr>
<td>All other rural general hospitals, beds size $300-449$ ($N = 5$)</td>
<td>6003</td>
<td>0.201</td>
<td>0.163</td>
<td>0.150</td>
</tr>
</tbody>
</table>

Case-mix index measures the average cost weight for a subgroup of patients compared with that for a larger group of patients. In this instance, the average cost weight for neonatal patients at each of the hospital types (numerator) is divided into the average cost weight for all patients from all hospitals (denominator). The case-mix indices for neonatal patients at each of the general hospital types is $<1.00$ because the majority of neonates are normal newborns.

Costs include hospital operating and capital costs, exclusive of graduate medical costs, adjusted by HCFA area wage index.

Children's hospitals include freestanding acute children's hospitals with separate Medicare provider numbers.

Major teaching hospitals include general hospitals with a ratio of interns and residents to beds $\geq 0.25$. 

---

**SUPPLEMENT 315**

Downloaded from by guest on August 30, 2017
given hospital type (numerator) is divided into the average cost weight for all patients from all hospitals (denominator). A percent change in case-mix index will correspond to the same percent change in payment ratio. According to Table 9, the small other urban general hospitals show a very substantial decrease in their case-mix index with the movement from Medicare DRGs to AP-DRGs and APR-DRGs. The mid-sized other urban general hospitals show the same pattern but the decreases in case-mix index are not as dramatic. The large other urban general hospitals (bed size \( \geq 450 \) beds) show an increase in case-mix index with the AP-DRGs and APR-DRGs but not as large as that for major teaching general hospitals.

The payment ratios in Table 8 for other rural hospitals \((N = 234)\) show a distinctive pattern that is consistent by hospital bed size. The composite payment ratio for all other rural hospitals is extremely high at 1.388 for Medicare DRGs, decreases to 1.195 for AP-DRGs, and decreases to 1.122 for APR-DRGs. The pattern is one of substantial overpayment by Medicare DRGs, moderated somewhat by AP-DRGs and APR-DRGs. It is important to point out again that this is a simplified payment simulation with no payment adjustments for outlier patients or any facility level adjustments.

The pattern for other rural hospitals is shown to be consistent across rural hospitals of different bed sizes in Table 9. It is important to note that there are few large rural hospitals in the study database and so this part of the results should be interpreted with caution.

**CONCLUSION**

There are major differences in the structure and statistical performance of Medicare DRGs, AP-DRGs, and APR-DRGs for neonatal patients. The Medicare DRGs are structurally the least well-developed and yield the poorest statistical performance. The APR-DRGs are structurally the most developed and yield the best statistical performance, both for cost and risk of mortality. The AP-DRGs are intermediate to Medicare DRGs and APR-DRGs although closer to the APR-DRGs.

The APR-DRGs remove most of the systematic biases in DRG classification for neonatal patients. Of the existing DRG systems, which group patients based on existing information in the hospital inpatient medical record abstract, the APR-DRGs clearly provide

**APPENDIX 1.** Listing of Neonatal Categories in Version 15.0 HCFA DRGs

<table>
<thead>
<tr>
<th>HCFA DRG No.</th>
<th>DRG Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>385</td>
<td>Neonate, died or transferred to another acute hospital</td>
</tr>
<tr>
<td>386</td>
<td>Extreme immaturity or respiratory distress syndrome</td>
</tr>
<tr>
<td>387</td>
<td>Prematurity with major problem</td>
</tr>
<tr>
<td>388</td>
<td>Prematurity without major problem</td>
</tr>
<tr>
<td>389</td>
<td>Full-term neonate with major problem</td>
</tr>
<tr>
<td>390</td>
<td>Neonate with other significant problem</td>
</tr>
<tr>
<td>391</td>
<td>Normal newborn</td>
</tr>
</tbody>
</table>

**APPENDIX 2.** Listing of Neonatal Categories in Version 15.0 AP-DRGs

<table>
<thead>
<tr>
<th>AP-DRG No.</th>
<th>AP-DRG Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>602</td>
<td>Neonate, birth weight &lt;750 g, discharged alive</td>
</tr>
<tr>
<td>603</td>
<td>Neonate, birth weight &gt;7500 g, died</td>
</tr>
<tr>
<td>604</td>
<td>Neonate, birth weight 750-999 g discharged alive</td>
</tr>
<tr>
<td>605</td>
<td>Neonate, birth weight 750-999 g, died</td>
</tr>
<tr>
<td>606</td>
<td>Neonate, birth weight 1000-1499 g, with significant OR procedure</td>
</tr>
<tr>
<td>607</td>
<td>Neonate, birth weight 1000-1499 g, without significant OR procedure</td>
</tr>
<tr>
<td>608</td>
<td>Neonate, birth weight 1000-1499 g, died</td>
</tr>
<tr>
<td>609</td>
<td>Neonate, birth weight 1500-1999 g, with significant OR procedure, with multiple major problems</td>
</tr>
<tr>
<td>610</td>
<td>Neonate, birth weight 1500-1999 g, with significant OR procedure, without multiple major problems</td>
</tr>
<tr>
<td>611</td>
<td>Neonate, birth weight 1500-1999 g, without significant OR procedure, with multiple major problems</td>
</tr>
<tr>
<td>612</td>
<td>Neonate, birth weight 1500-1999 g, without significant OR procedure, with major problem</td>
</tr>
<tr>
<td>613</td>
<td>Neonate, birth weight 1500-1999 g, without significant OR procedure, with minor problem</td>
</tr>
<tr>
<td>614</td>
<td>Neonate, birth weight 1500-1999 g, with significant OR procedure, with other problem</td>
</tr>
<tr>
<td>615</td>
<td>Neonate, birth weight 2000-2499 g, with significant OR procedure, with multiple major problems</td>
</tr>
<tr>
<td>616</td>
<td>Neonate, birth weight 2000-2499 g, with significant OR procedure, without multiple major problems</td>
</tr>
<tr>
<td>617</td>
<td>Neonate, birth weight 2000-2499 g, without significant OR procedure, with multiple major problems</td>
</tr>
<tr>
<td>618</td>
<td>Neonate, birth weight 2000-2499 g, without significant OR procedure, with major problem</td>
</tr>
<tr>
<td>619</td>
<td>Neonate, birth weight 2000-2499 g, without significant OR procedure, with minor problem</td>
</tr>
<tr>
<td>620</td>
<td>Neonate, birth weight 2000-2499 g, without significant OR procedure, normal newborn</td>
</tr>
<tr>
<td>621</td>
<td>Neonate, birth weight 2000-2499 g, without significant OR procedure, with other problem</td>
</tr>
<tr>
<td>622</td>
<td>Neonate, birth weight &gt;2499 g, with significant OR procedure, with major problems</td>
</tr>
<tr>
<td>623</td>
<td>Neonate, birth weight &gt;2499 g, with significant OR procedure, without multiple major problems</td>
</tr>
<tr>
<td>624</td>
<td>Neonate, birth weight &gt;2499 g, with minor abdominal procedure</td>
</tr>
<tr>
<td>625</td>
<td>Neonate, birth weight &gt;2499 g, without significant OR procedure, with multiple major problems</td>
</tr>
<tr>
<td>626</td>
<td>Neonate, birth weight &gt;2499 g, without significant OR procedure, with major problem</td>
</tr>
<tr>
<td>627</td>
<td>Neonate, birth weight &gt;2499 g, without significant OR procedure, with minor problem</td>
</tr>
<tr>
<td>628</td>
<td>Neonate, birth weight &gt;2499 g, without significant OR procedure, normal newborn</td>
</tr>
<tr>
<td>629</td>
<td>Neonate, birth weight &gt;2499 g, without significant OR procedure, with other problem</td>
</tr>
<tr>
<td>630</td>
<td>Neonate, birth weight &gt;2499 g, without significant OR procedure, with problem 635</td>
</tr>
</tbody>
</table>
The most accurate and reliable method to classify neonates. This is true whether the application is case-mix trending, utilization management and quality improvement by hospitals and physicians, comparative reporting by a data commission, prospective payment by a government agency, or price negotiations between a hospital and a payor. Each specific application has its own methodologic and policy issues, but it is critical to know how accurately patients are classified by the respective DRG systems before considering these additional issues.

Appendix 4

Technical Explanation of Explanation of Variance ($R^2$) Statistics

The most common statistical measure used to compare patient classification systems is reduction of variance ($R^2$), which measures the proportion of variation that is explained by a DRG system. $R^2$ provides a summary measure of the extent to which a DRG system is able to predict the value of a resource use or outcome variable based on the characteristics of individual patients. For a categorical variable such as DRG, $R^2$ is computed as

$$R^2 = \frac{\sum(y_i - \bar{A})^2}{\sum(y_i - \bar{A}_g)^2}$$

where $y_i$ is the value of the variable (ie, cost or length of stay) for the $i$th patient, $A$ is the average value of the variable in the database and $A_g$ is the average value of the variable in DRG $g$. The square of the difference between the actual value (ie, $y_i$) and the predicted value (ie, $A$ or $A_g$) is a measure of the variation in the data. The term

$$\sum(y_i - A)^2$$

is the amount of variation before subdividing the data into DRGs and the term

$$\sum(y_i - A_g)^2$$

is the amount of variation after subdividing the data into DRGs. The difference between these two terms is the reduction in variation resulting from the subdivision of the data into DRGs. $R^2$ is the ratio of the reduction in variation to the amount of variation before subdividing into DRGs. $R^2$ ranges between 0 and 1 and measures the fraction of variation explained by the DRGs. Thus, an $R^2$ of 0.415 would mean that subdividing the data into DRGs reduces the amount of variation in the data by 41.5%.


**ACKNOWLEDGMENTS**

This study of DRG classification systems and neonatal medicine was performed as part of a larger study of DRG classification systems and all age patients by the National Association of Children’s Hospitals and Related Institutions and 3M Health Information Systems. The research staff from 3M Health Information Systems who participated in the study design and data analysis included: Richard F. Averill, MS; Norbert I. Goldfield, MD; James C. Vertrees, PhD; Elizabeth C. Fineran, MS; and Mona Z. Zhang, MS. Project support staff from NACHRI responsible for preparation of the study manuscript and tables was Lisa J. Turner, Senior Administrative Assistant. The APR-DRG classification system and software is a proprietary product of 3M Health Information Systems. Special thanks is extended to Albert Bartoletti, MD for his instruction over the years in neonatal diagnostic conditions and classification issues.

Basic data for use in this study were supplied by the HCIA, Baltimore, MD. These data were supplied only at the request of and on the authorization of the hospitals whose data were used. Any analysis, interpretation, or conclusion based on these data is solely that of the author (John H. Muldoon, NACHRI), and HCIA specifically disclaims responsibility for any such analysis, interpretation, or conclusions based on these data.

**REFERENCES**

3. Ibid
Structure and Performance of Different DRG Classification Systems for Neonatal Medicine
John H. Muldoon
Pediatrics 1999;103;302

Updated Information & Services
including high resolution figures, can be found at:
/content/103/Supplement_E1/302.full.html

References
This article cites 1 articles, 1 of which can be accessed free at:
/content/103/Supplement_E1/302.full.html#ref-list-1

Citations
This article has been cited by 6 HighWire-hosted articles:
/content/103/Supplement_E1/302.full.html#related-urls

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
Fetus/Newborn Infant
/cgi/collection/fetus:newborn_infant_sub
Neonatology
/cgi/collection/neonatology_sub

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
/site/misc/Permissions.xhtml

Reprints
Information about ordering reprints can be found online:
/site/misc/reprints.xhtml
Structure and Performance of Different DRG Classification Systems for Neonatal Medicine
John H. Muldoon
Pediatrics 1999;103;302

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/103/Supplement_E1/302.full.html