

Estimation of Variance (2007 Data)

The Hospital Cost Transparency (HCT) project annually reports inpatient payments for APR-DRGs that, during the previous calendar year, had at least 150 observations or at least \$1 million in allowed payments and a minimum of 25 observations. Furthermore, the governing workgroup decided the variance should be estimated after stratifying the data by APR-DRG and severity of illness category (minor, moderate, major, and extreme). This means separately estimating the variance in each cell of a 4xR matrix, with R equal to the number of eligible APR-DRGs and 4 referring to the severity of illness categories.

The difficulty with this approach is that it was known in advance that a substantial number of cells would not have a sufficient number of observations to generate reliable variance estimates. Since the first year's data were ultimately published after combining the severity of illness into two categories (minor/moderate and major/extreme), the 2007 estimated variances used the same combined severity of illness categories.

In the event of small cell sizes, contingencies provided that the analyst should attempt to estimate the variance from six supplemental data sets (in order):

1. The 2007 data supplemented with all observations from the 2006 HCT data set (adjusted for inflation to 2007 dollars)
2. The combined 2006-2007 data set supplemented with all observations from the 2005 HCT data set (adjusted for inflation to 2007 dollars)
3. The 2006 Oregon Hospital Discharge Data (HDD), adjusted for inflation to 2007 dollars
4. The 2006 HDD supplemented with observations from the 2005 HDD (adjusted for inflation to 2007 dollars)
5. The 2005-2006 HDD supplemented with observations from the 2004 HDD (adjusted for inflation to 2007 dollars)
6. The 2004-2006 HDD supplemented with observations from the 2005 Nationwide Inpatient Sample (NIS), adjusted for inflation to 2007 dollars

Both the HDD and NIS data sets were previously risk-adjusted with 3M Core Grouping Software, which assigns the APR-DRG and severity of illness.

Preliminary Data Preparation

The 2007 HCT aggregate claims data were first assessed for their suitability for risk-adjustment with the 3M Core Grouping Software. This software assigns a severity of illness category based on diagnoses, procedures, length of stay, patient age, and patient discharge disposition. The risk-adjustment software will not provide a severity of illness score if either age or discharge disposition are missing, or if the principal diagnosis is either missing or invalid.

The risk-adjustment can also be influenced by secondary diagnoses and procedures. The standard inpatient discharge record contains nine diagnoses and six procedures. Internal testing revealed that deleting several diagnoses and procedures could alter the assigned severity of illness. Some claims systems do not capture all of the diagnoses

and procedures in the discharge record, making risk-adjustment of these records potentially unreliable.

Accordingly, records in the aggregate claims data were required to meet the following risk-adjustment standards:

- Age is not missing
- Length of stay is not missing
 - This requires the admit date and discharge date fields to be populated with valid dates
- Principal diagnosis is not missing
- Claims system captures at least five diagnoses
- Claims system captures at least four procedures

Additional validations were performed to assure the quality of the data and eliminate duplicate records. Details about this extensive process are available in a separate document. The final 2007 HCT aggregate claims file contained 54,358 risk-adjusted and validated records.

Records that did not meet the risk-adjustment standards and records that failed the additional validations were dropped. A separate data file of dropped records was maintained to assess selection bias (see Table 1).

Table 1: Characteristics of Submitted Claims Data

	Risk-adjusted records	Dropped records
Percent routine discharge	90.9%	86.3%
Percent female	61.6%	60.8%
Median age	37	29
Mean length of stay	3.6	4.3
Mean allowed payments	\$12,537	\$12,281

The differences between risk-adjusted records and dropped records are modest, although not negligible. The patients from dropped records tended to be a few years younger, but had slightly longer hospital stays and were less likely to be routinely discharged. The net result of this is patients from dropped records may be somewhat higher risk. The mean allowed payments for these patients were not substantially different. Thus, selecting records for risk adjustment could introduce a small amount of bias favoring patients with slightly less risk, although this does not appear to result in substantial differences in allowed payments.

Inclusion Criteria

Inclusion criteria were then applied to 2007 HCT aggregate claims data:

- Inpatient claims
 - Excluding Med-Advantage claims
 - Excluding coordination of benefit claims
 - Excluding denied claims
 - Excluding workers compensation claims
- Patient resided in Oregon

- Patient was treated as an inpatient at an acute-care hospital located in Oregon
 - Excluding VA facilities, Shriner's, and psychiatric hospitals
 - Excluding patients who discharged against medical advice
 - Excluding patients who expired

The inclusion criteria for records from the HDD and NIS were the same as the inclusion criteria for the 2007 HCT data, except that the data are limited to records where the expected primary payer is a commercial health plan.

Estimating Allowed Payments by APR-DRG

Annual hospital-specific cost-to-charge ratios were estimated for Oregon hospitals from audited financial statements. To estimate allowed payments, total payments from the HDD were multiplied by the cost-to-charge ratio and then adjusted for inflation to 2007 dollars. Separate data sets were created for calendar years 2004 to 2006, and then combined data sets were created for calendar years 2005-2006 and 2004-2006.

For the NIS, cost-to-charge ratios were obtained from CMS Payment Impact Files from 2003, 2004 (if missing from 2003), 2005 (if missing from 2003 and 2004), and 2006 (if missing from 2003 to 2005). Allowed payments were estimated with the same calculations used in the HDD estimates. The NIS data were also restricted to non-Oregon hospitals to avoid duplicating any HDD records. The NIS-supplemented data file was then created by combining the 2004-2006 HDD and NIS data sets.

The 2007 HCT aggregate claims data were stratified by APR-DRG in order to sum the allowed payments and to determine the number of observations. Records were flagged for inclusion if the APR-DRG had at least 150 observations. Records were also flagged for inclusion if the sum of allowed payments for the APR-DRG was at least \$1 million and there were at least 25 observations. Records that were not flagged were removed and maintained in a separate data file of excluded records. The APR-DRGs identified in the 2007 HCT data were then flagged in the HDD and NIS-supplemented data sets.

Estimating Variance

The 2007 HCT aggregate claims data were then stratified by APR-DRG and two combined severity of illness categories (minor/moderate and major/extreme) in order to generate cell sizes. A variance estimation flag was created to identify cells with at least 30 observations. This was repeated with the combined 2006-2007 HCT data, combined 2005-2007 HCT data, 2006 HDD, 2005-2006 HDD, 2004-2006 HDD, and the NIS-supplemented data set. The variance estimation flags were merged into the 2007 HCT aggregate claims data, the flags indicating the data set from which the variance was estimated.

With the data stratified by APR-DRG and two combined severity of illness categories, the standard deviation (SD) of the allowed payments was calculated for each cell in the 2007 HCT aggregate claims file. The SD was similarly calculated in the remaining six supplemental data sets. The SD variables from the six supplemental data sets were merged into the 2007 HCT aggregate claims file. Each record in the 2007 HCT data was assigned a SD variable based on the value in the variance estimation flag. Outliers were identified as observations with allowed payments outside +/- 4 SD from the mean.

Outliers were removed from the 2007 HCT data and transferred to a data file of excluded records.

The 2007 HCT aggregate claims file was then aggregated by APR-DRG in order to determine the number of observations after removing outliers. Records were again flagged for inclusion based on frequency (APR-DRG had at least 150 observations) or cost (APR-DRG had at least \$1 million in allowed payments and at least 25 observations). Records not flagged were removed and transferred to the data file of excluded records. The data were then stratified by APR-DRG and two combined severity of illness categories (minor/moderate and major/extreme). The allowed payments were summed for each cell and the mean allowed payments, median allowed payments, mean length of stay, and the number of observations were calculated. The results were separately tabulated based on frequency and based on allowed payments.

The data were then further stratified by hospital. Again, the allowed payments were summed for each cell and the mean allowed payments, median allowed payments, and number of observations were calculated. The results were tabulated by APR-DRG for two combined severity of illness categories (minor/moderate and major/extreme) and each hospital. Cells with fewer than two observations were reported as “0 or 1.”

In order to assess selection bias due to applying the exclusion criteria, several variables (age, gender, length of stay, allowed payments, and discharge disposition) were compared between included and excluded records (note that dropped records are not part of this analysis). In addition, these variables were compared after filtering out records involving pregnancy and childbirth, since these records represent a substantial proportion of the data set and will tend to focus on narrow ranges of age, length of stay, and allowed payments.

Data were analyzed with SAS version 9.1, SPSS version 15.0.0, 3M[®] Core Grouping Software version 10.0.1, and Microsoft Office[®] Excel 2003.

Results

In the 2006 HDD 151,178 records were identified for inclusion. In the 2005-2006 HDD 273,669 records were identified for inclusion. In the 2004-2006 HDD 409,988 records were identified for inclusion. In the NIS data set 1,240,270 records were identified for inclusion. Cost-to-charge ratios were not available for all hospitals in this NIS subsample; ultimately 818,213 records were added to the HDD, so the NIS-supplemented data set contained a total of 1,234,916 records.

Originally 142 unique APR-DRGs were identified for further analysis based on either frequency or cost. The resulting 2 x 142 matrix contained 284 cells, although 1 cell was null after generating cell sizes and five APR-DRGs were subsequently excluded after removing outliers. Of the remaining 273 cells (137 unique APR-DRGs), the variance was directly estimated from the 2007 HCT aggregate claims data for 192 cells (67.8%). The variance was directly estimated from the combined 2005-2006 HCT data for 36 cells (12.7%). The variance was directly estimated from the combined 2005-2007 HCT data for 14 cells (4.9%). The variance was directly estimated from the 2006 HDD for 21 cells (7.4%), the variance was directly estimated from the 2005-2006 HDD for 10 cells

(3.5%, and the variance was directly estimated from the combined 2004-2006 HDD for 3 cells (1.1%). Finally, the variance was directly estimated from the NIS-supplemented data for 7 cells (2.5%).

A total of 70 unique APR-DRGs had at least 150 observations. A total of 137 unique APR-DRGs had at least \$1 million in allowed payments and at least 25 observations. This represents a total of 46,704 observations and over \$558 million in allowed payments; 7654 records were excluded from further analysis. Diagnoses and procedures from six major clinical domains (obstetrics/gynecology, childbirth, digestive system, respiratory system, cardiovascular system, and musculoskeletal system) accounted for 83 unique APR-DRGs, over 80% of the observations (37,463) and over 88% of the total allowed payments (\$496 million).

APR-DRG 560 (vaginal delivery) had the highest total allowed payments and APR-DRG 640 (normal newborn) ranks had the largest number of observations. Diagnoses and procedures involving the cardiovascular system accounted for 21 of the 137 APR-DRGs while diagnoses and procedures involving the digestive system accounted for an additional 19 APR-DRGs. Diagnoses and procedures in the obstetrics/gynecology and childbirth domains accounted for 20,480 observations, or approximately 44% of the included observations.

After filtering out records involving pregnancy and childbirth, differences were generally modest when comparing included records to excluded records in the risk-adjusted aggregate claims file (see Table 2). There were modest differences in median age and mean length of stay, although this did not result in substantial differences in mean allowed payments. Overall this indicates that, after filtering out records involving pregnancy and childbirth, the included APR-DRGs do not produce an egregiously biased subset of records from the risk-adjusted aggregate claims data.

Table 2: Characteristics of Risk-Adjusted Claims Data

	Included records	Excluded records
Percent routine discharge	85.9%	84.5%
Percent female	47.8%	46.9%
Median age	53	45
Mean length of stay (days)	3.9	4.6
Minor/moderate severity of illness	84.6%	81.8%
Mean allowed payments	\$16,487	\$15,123

The magnitudes of the differences in median age, percent female, and mean allowed payments are substantially different if including pregnancy and childbirth records. It should be noted that extremely few pregnancy and childbirth records end up being excluded from the risk-adjusted aggregate claims data, so the risk of selection bias is minimal for pregnancy and childbirth records. However these records, since they are a very large subset of the aggregate claims data, can affect the assessment of selection bias in other records if not filtered out.