Estimation of Variance (2006 Data)

Background

In March 2007, the Cost Transparency Workgroup decided to focus on APR-DRGs that had at least 150 observations or at least $1 million in allowed payments and a minimum of 25 observations. Furthermore, the 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 2005 data were ultimately published after combining the severity of illness into two categories (minor/moderate and major/extreme), the 2006 estimated variances used the same combined severity of illness categories.

In the event of small cell sizes, the 2006 data was first supplemented with all observations from the 2005 data set. If any small cell sizes persisted, it was attempted to estimate the variance from the 2006 Oregon Hospital Discharge Data (HDD). In the case of small cell sizes in the 2006 HDD, this data was then supplemented with observations from the 2004 and 2005 HDD. In the case of small cell sizes in the 2004-2006 HDD, a further contingency provided that the 2004-2006 HDD would be supplemented with data from other states. Since Oregon participates in the Health Care Utilization Project, data from the 2003 Nationwide Inpatient Sample (NIS) were readily available to supplement the HDD. Both the HDD and NIS data sets were previously risk-adjusted with the APR-DRG software.

Preliminary Data Preparation

The aggregate claims data were first assessed for their suitability for risk-adjustment with the APR-DRG 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 aggregate claims file contained 39,993 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	92.0%	83.8%
Percent female	61.6%	60.8%
Median age	32	37
Mean length of stay	3.2	3.9
Mean allowed payments	$9,223	$12,596

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 older, had slightly longer hospital stays, and were less likely to be routinely discharged. The net result of this is patients from dropped records are most likely somewhat higher risk. The mean allowed payments for these patients are also moderately higher, intuitively consistent with the idea that higher risk patients utilize more resources. Thus, selecting records for risk adjustment could introduce a small amount of bias favoring patients with slightly less risk and moderately smaller allowed payments.

Methods

The inclusion criteria for records from the HDD and NIS were the same as the inclusion criteria for the aggregate claims file (primary payer is a commercial health plan excluding transfers, patients who expired, patients who left against medical advice, records that could not be risk-adjusted, and records with validation errors). The NIS data were also restricted to non-Oregon hospitals to avoid duplicating any HDD records.

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 2006 dollars. A data set was created for calendar year 2006 and for calendar years 2004 to 2006. For the NIS, cost-to-charge ratios were obtained from CMS Payment Impact Files from 2003, 2004 (if missing from 2003), and 2005 (if missing from 2003 and 2004).

Allowed payments were estimated with the same calculations used in the HDD estimates. The NIS-supplemented data file was then created by combining the 2004-2006 HDD and NIS data sets.

The 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 aggregate claims file were then flagged in the HDD and NIS-supplemented data sets.

The 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 set to identify cells with at least 30 observations. This was repeated with the 2006 HDD, 2004-2006 HDD, and the NIS-supplemented data set. The flags were merged into the aggregate claims data, so that the flag indicated if the variance was estimated from the claims data, from the 2006 HDD, from the 2004-2006 HDD, or from the NIS-supplemented data.

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 aggregate claims file. The SD was similarly calculated in the HDD and NIS-supplemented data sets. The SD variables from the HDD and NIS were merged into the aggregate claims file. Each record in the aggregate claims file 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 aggregate claims data and transferred to a data file of excluded records.

The 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 number of observations were calculated. The results were separately tabulated based on frequency and based on cost. Cells with fewer than five observations were reported as “< 5.”

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 asses 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 152,185 records were identified for inclusion. In the 2004-2006 HDD 433,410 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 1,251,623 records.

Originally 96 unique APR-DRGs were identified for further analysis based on either frequency or cost. The resulting 2 x 96 matrix contained 192 cells, although 2 cells were null after generating cell sizes and two APR-DRGs were subsequently excluded after removing outliers. Of the remaining 186 cells, the variance was directly estimated from the risk-adjusted aggregate claims data for 119 cells (64%). The variance was directly estimated from the combined 2005-2006 claims data for 28 cells (15%). The variance was directly estimated from the 2006 HDD for 30 cells (16%) and the variance was directly estimated from the combined 2004-2006 HDD for 5 cells (3%). Finally, the variance was directly estimated from the NIS-supplemented data for 4 cells (2).

A total of 55 unique APR-DRGs had at least 150 observations. A total of 92 unique APR-DRGs had at least $1 million in allowed payments and at least 25 observations. APR-DRG 560 (vaginal delivery) ranks first in both tables. Diagnoses and procedures involving the cardiovascular system accounted for 18 of the 92 APR-DRGs while diagnoses and procedures involving the digestive system accounted for an additional 16 APR-DRGs. Diagnoses and procedures in the obstetrics/gynecology domain accounted for 20,206 observations, or approximately 51% of the total 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). The only noteworthy difference was the mean length of stay (58% higher for excluded records), although this difference did not translate into higher allowed payments.

Table 2: Characteristics of Risk-Adjusted Claims Data

	Included records	Excluded records
Percent routine discharge	86.9%	85.7%
Percent female	49.9%	47.3%
Median age	52	50
Mean length of stay (days)	3.6	5.7
Minor/moderate severity of illness	88.5%	81.8%
Mean allowed payments	$12,521	$11,136

Overall this indicates that, after filtering out records involving pregnancy and childbirth, the chosen APR-DRGs do not produce an egregiously biased subset of records from the risk-adjusted aggregate claims data. The magnitudes of the differences in median age, percent female, and mean allowed payment are substantially higher 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 in this very large subset of records.