

Severity Adjustment Fact Sheet

Severity Adjustment Model:

The evidence-informed case rate (ECR) starts when an episode is triggered, a base case rate is paid and additional allowances become available for comorbid conditions and procedures that are included as part of the ECR. Money from the PAC allowance also becomes available based on the severity and complexity of the patient and is available to providers to care for any complications that may occur. The global nature of the ECR encourages providers to coordinate care within an episode and limit duplication of services, and creates a strong incentive to eliminate waste and care defects and to improve quality.

Looking at healthcare delivery from a claims perspective for the modeling exercise, we need to understand services that are provided and claims that are received from facilities (inpatient and outpatient), radiology services, physicians, lab, pharmacy and other ancillary care providers. To construct the entire ECR, we decided to create separate risk-adjustment payments models for three components: 1) inpatient facility, 2) outpatient facility, and 3) professional & others (here all other services are aggregated together for the purposes of modeling into one bucket containing inpatient professional, outpatient professional, pharmacy, lab, radiology and other ancillary claims). The results of the three models are then brought together as a complete ECR to provide a global payment for care for the various components outlined above.

This approach removes the incentive to provide care in a given location (inpatient vs. outpatient) and lets the provider choose the type and place of services that would lead to the most cost-effective care. In addition, there may be comorbid conditions that modify the costs of care in the inpatient setting that are different than the risk modifiers in the outpatient setting, hence the need for separate models. Moreover, the inpatient facility claims have their own coding rules and a hospital confinement often comes with a bundled cost (allowed amount) on a single line item, but risk factors and services on separate line items that provide a different challenge for modeling. On the other hand, the professional claims are individually available with cost and services on a line-by-line basis and can be linked using the member ID.

The approach used for modeling involved selecting members that had a given diagnosis and then bucketing “claims” into those that were for typical care, or for potentially

avoidable complications based on the diagnosis codes and services provided. In addition, we chose to use pharmacy data to adjust for severity, since we found in the claims data, physicians do not use diagnosis codes completely, so pharmacy data helped enhance the risk-adjustment process and capture additional comorbidities.

Unit of Analysis:

The unit of analysis is the component of episode being modeled e.g. inpatient facility (stay model), outpatient facility, or professional, pharmacy and other services (PFO) model. It contains all “typical” claims that are relevant to that component of the episode. It is important to exclude claims for potentially avoidable complications (PACs) since costs related to these services should not have occurred and we do not want to adjust away or justify the occurrence of complications.

Dependent Variable:

The dependent variable is cost. We use the “allowed amount” field for costs since it usually represents the reimbursed amount to the providers along with the patient portion of the costs and does not vary with the members benefit design or utilization of services. We remove outlier ECRs by truncating the selected claims at the 1st and 99th percentile value for costs. The distribution of cost was examined, and was transformed using the log transformation (Ln) to reduce the skewness in cost data.

Independent Variables:

A separate list of independent variables is created for each ECR specific to the severity indicators and comorbidities for the condition under study. The list of “risk factors” or “cost modifiers” consists of patient demographic factors, hospital or provider characteristics, geographic area indicators, comorbidities, procedures or diagnostic tests performed, and pharmacy variables to the extent this information is available in the database. Variables are often grouped together into relatively homogenous risk categories. Frequency of occurrence of variables often determines if the variable needs to be entered alone or should be combined with other variables to ensure they are not lost from the modeling exercise. A difference in cost using univariate analysis (with or without a variable) often determines whether a given variable will be presented to the multivariable stepwise analysis. The variables selected in the inpatient model may be different from those used in the outpatient model depending on the implications of each variable to cost in various settings. The statistical modeling determines whether or not a variable should be included in the final models.

Statistical Methods

To measure the association between each categorical independent variable and cost, a bivariate analysis using the t-test (or analysis of variance (ANOVA) if the independent variable has more than 2 levels) was performed on Ln cost. Means, standard deviations and other descriptive statistics was transformed back and reported on the original scale. For the “typical” cases and services analysis, we used the ordinary least squares regression (OLS; also known as multiple linear regression) analysis for developing the ECR multivariable models.

Multiple linear regression is a statistical method for measuring associations between a set of independent variables and a continuous outcome such as cost. This method has the following advantages: (a) widely used and accepted, (b) easy to understand, (c) relatively easy to explain to non-statisticians, (d) transparent, and (e) several different model selection methods are available in SAS (e.g. Schwarz Criterion, automated variable selection procedures such as stepwise).

The linear regression models are of the following form:

$$\text{Ln } Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_px_p + \epsilon$$

Where

Ln Y is the natural log of costs for the “typical” care and services

β_0, β_1, \dots are the regression coefficients

x_1, x_2, \dots are the independent variables; these variable include demographics, and comorbidities

ϵ is the error term

This model assumes that the ϵ term is independent, normally distributed with mean 0 and constant variance.

Model Selection Criteria

The question of which model is the “best” is often not a straightforward one. Generally, the “best” model is the one that strikes a balance between being clinically plausible, fitting the data well, and yet not being overly complex (i.e. not having a large number of parameters in the model). An effective model selection process is one that combines both clinical input as well as statistical evidence (e.g. model building statistics). The process described below used a combination of clinical review and model building statistics. Variables are first examined univariately. Variables which are clinically plausible, have a minimum number of 30 cases per category, and the t-test p-value for the association between the variable and Ln cost was <0.5 are placed in the pool of candidate

independent variables. Some of these candidate variables are reclassified as “force” variables, i.e. variables that should be in all models due to their clinical importance. Multiple linear regression with the stepwise variable selection procedure is used to select from the remaining candidates. The p-values to enter and exit the model are set at 0.05. The stepwise procedure works by fitting a model with only the forced variable in the model. Then the procedure computes an F statistic for each candidate independent variable. The F statistic is the F statistic that would be obtained if only that variable were added to the model. The variable with the lowest F statistic p-value that is less than 0.05 is put in the model, and then the model is refit. New F statistics are computed for the remaining candidate variables, and again the one with the lowest F statistic p-value below 0.05 is added to the model. This process continues until no other candidates have F statistic p-values <0.05 . If, at any time during the process, the p-value for the F statistic of any variable in the model goes above 0.05, that variable is removed from the model unless the variable is a force variable.

The resulting model is reviewed for clinical plausibility. Variables are selected if they were clinically important, their coefficients are in the right direction and magnitude, and their variance inflation factor (VIF) is low suggesting no evidence of multicollinearity. The models are re-run with the selected variables. This process of model fitting, clinical review, and refitting is continued until a clinically plausible model is identified.

Regression Diagnostics:

Once a model is selected, the assumptions of the linear regression model are verified. As mentioned above, this model assumes that the errors, ϵ_i , are independent, normally distributed random variables with mean 0 and constant variance. Residual analysis was performed to verify these assumptions. If the residual analysis does not meet the linear regression assumptions, a Box-Cox transformation is performed on the data and the lambda that has the best regressions diagnostics is selected.

Bootstrap Validation:

The above model is run using stepwise selection with SLE (significance level for model entry) = 0.05 and SLS (significance level for staying in model) = 0.05 on 200 bootstrap samples, and a record is kept of how many times each variable is selected in the 200 model runs. A bootstrap sample is created by sampling with replacement from the full dataset until a sample of the same size as the full dataset is drawn. Any variables that were forced into the model above are also forced in the 200 bootstrap models. Any variable that enters into less than 160 bootstrap models (80% of the models) is dropped.

Final Model:

Variables which do not consistently perform as significant predictors in at least 75% of the 200 bootstrap validations are dropped. The final step is to run the model cleanup algorithm as in Step 4 one last time. The reason for doing this is that removal of some variables may cause perturbations in the sign of other variables, if there exists some correlation between these variables. The resulting model is the final model.

Retransformation Bias:

Because cost is log-transformed before it is modeled, there is a bias that occurs when one tries to back-transform predicted LN(ALLOW) into the original units. This results in systematic underpayment and the cumulative effect can be substantial. For this reason a Bias Correction Factor (BCF) is calculated as the ratio of the total actual costs in dollars (of all patients that went into the model) to their predicted costs. Multiplying the predicted cost by this BCF assures that there will be no systematic under or over-prediction of cost on the user's data set.

LIMITATIONS OF ANALYSIS

1. Models are only as good as the data.
2. We're using claims data from specific populations to estimate costs which will possibly be used for other populations. These cost estimates may not apply to all populations.
3. Claims data are messy with incomplete or incorrect diagnosis codes being used and there is limited clinical information that can only be obtained through patient chart reviews such as information on body weight, BP, smoking status etc.
4. Socio-economic and other factors such as educational level, family support, access to healthcare etc., which may impact costs are not available in most administrative databases but are very important for severity adjustment.
5. Laboratory data is not available to us – it has been shown to add a lot of value to administrative data for risk-adjustment purposes to assess the severity of a patient's disease.

Limitations of Clinical Guidelines

Work to date suggests clinical guidelines are necessary, but not always sufficient, to create a base ECR. Guidelines can use language that defers to a doctor's discretion, such

as, “Use treatment X, when appropriate.” This can make it difficult to determine the resources necessary to provide recommended care. Where the best available evidence or expert consensus does not provide sufficient information, empirical data must be used to fill in knowledge gaps. In these cases, data modeling will help to determine where ECRs can provide sufficient flexibility to physicians without having significant impact on the final case rate. For example, clinical guidelines are often not specific in detailing the amount of case management and follow-up required to treat a given condition. However, if the final ECR rate is not sensitive to wide variation in follow-up time, this becomes much less of an issue. Alternatively, examining cost data with high compliance rates may show a more consistent approach to case management as done by best-practice providers.

Use of Severity Adjustment Models to adjust for PAC rates

Conceptual Model:

Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, the remaining variance in PAC rates are due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window. We have developed a severity index based on patient-related factors, such as patient demographics and comorbidities. The severity-adjusted PAC counts give a fair comparison of PAC rates from population to population and helps providers determine the degree of PACs that are not related to patient-level factors but due to factors that they could control.

Methodology Overview:

A severity score is calculated for each patient based on the risk-adjustment model for professional and other services that determines the cost drivers for typical care for a given condition. Demographic variables, comorbid conditions, various types of services as well as patient-level pharmacy indicators are fed into the model. The model determines the patient-level factors that are drivers for increased financial risk. For each patient the “predicted” log coefficients from the severity adjustment model are summed to give the patient level severity score. Summing the patient level severity scores helps derive the population level average severity score. The population level severity score is exponentiated to back-transform it to dollar units and compared to the back-transformed severity score of the reference database to give the severity index. This index tells us if the population severity is higher or lower as compared to the reference database.

Adjusting the overall PAC rates by the severity index for the population helps adjust for variations in outcomes related to severity. The following examples illustrates this:

Diabetes	Medical Group 1	Medical Group 2	Medical Group 3
Number of Patients	100	50	500
Average Severity Score	9.1946	11.5845	15.6823
Reference Severity Score	9.9392	9.9392	9.9392
Severity Index	0.9251	1.1655	1.5778
PAC rates	80%	80%	80%
Severity Adjusted PAC rates	86%	69%	51%

In the above example, we are comparing three hypothetical medical groups with different population volumes and different average patient severity scores for diabetes. When compared to the reference database severity score, Medical Group 1 has lower severity while Medical Groups 2 has higher severity and Medical Group 3 has much higher severity. The PAC rates are the same in the three Medical Groups. However, when adjusted by the respective severity index, the PAC rates are different and Medical Group 3 looks the best with regards to the PAC rates.

AMI	Hospital 1	Hospital 2	Hospital 3
Number of Patients	100	50	500
Average Severity Score	5.6842	3.3486	6.5783
Reference Severity Score	6.2243	6.2243	6.2243
Severity Index	0.9132	0.5380	1.0569
PAC rates	73%	43%	85%
Severity Adjusted PAC rates	80%	80%	80%

A similar hypothetical example above looks at severity score calculations for three hospitals admitting patients for AMI. Here Hospital 1 has an average severity score close to that of the reference database, but Hospital 2 treats patients with much lower severity, and Hospital 3 treats patients with much higher severity. So even though Hospital 3 has much higher PAC rates, and Hospital 2 has much lower PAC rates, their severity-adjusted PAC rates are the same as the other hospitals.

REFERENCES:

- ¹ Prometheus Payment Inc. Available at: <http://www.hci3.org>. Accessed Aug 7, 2010.
- ² de Brantes F, Rastogi A. Evidence-informed case rates: paying for safer, more reliable care. Issue Brief (Commonwealth Fund). 2008; 40:1–13.
- ³ de Brantes, Rastogi, and Painter: Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. *Health Services Research Journal*, published online July 20 2010 DOI 10.1111/j. 1475-6773. 2010.01136.x
- ⁴ de Brantes F, Gosfield A, Emery D, Rastogi A and G. D’Andrea, “Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care”, Robert Wood Johnson Foundation Report, May 2009, <http://www.rwjf.org/pr/product.jsp?id=42555>, accessed October 2009.
- ⁵ de Brantes F, D’Andrea G, Rosenthal M. Should health care come with a warranty? *Health Affairs*. 2009; 28:w678–w687.
- ⁶ de Brantes, Rosenthal, Painter, “Building a Bridge from Fragmentation to Accountability — The Prometheus Payment Model”, *NEJM*, Sept 2009.
- ⁷ Rastogi A, Mohr BA, Williams JO, Soobader MJ, and de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. *Clin Orthop Relat Res* 2009; 467(10): 2587-2597.
- ⁸ Rathouz PJ and Rastogi A. *Linear Regression in Medical Research*. In Bailar JC and Hoaglin DC Eds. *Medical Uses of Statistics*, 3rd edition, Chapter 10, NEJM Press 2009, New Jersey: John Wiley & Sons Inc. 2009. ISBN: 978-0-470-43952-4 (cloth) and ISBN: 978-0-470-43953-1 (pbk.)
- ⁹ Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago, IL: Health Administration Press, 2003.
- ¹⁰ Chatterjee S, Hadi AS and Price B. *Regression Analysis by Example*, 3rd edition, 2000. Wiley Series in Probability and Statistics. John Wiley & Sons Inc. 2000. ISBN:0-471-31946-5.
- ¹¹ Duan, Naihua. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association*. 1983;78: 605-610.