Appendix: Models

Surgical Site Infection Risk by Procedure Volume

The Generalized Estimating Equations (GEE) models used to fit the outcome of having an SSI code as a function of annual Medicare surgical volume can be expressed as follows:

$$\text{logit}(\Pr[y_{ijt}=1]) = \frac{\Pr[y_{ijt}=1]}{1+\Pr[y_{ijt}=1]} =$$

$$\beta_0 + \beta_1\text{Low}_{ijt} + \beta_2\text{MedLow}_{ijt} + \beta_3\text{Medium}_{ijt} + \beta_4\text{MedHigh}_{ijt} + \beta_5\text{Age}_{ijt} + \beta_6\text{Gender}_{ijt} + \sum_{k=7}^{K}\beta_k + \text{Comorbidity}_{ijt}$$

where $y_{ijt}=1$ if person j in hospital i in year t had an SSI code and 0 otherwise, $\text{Low}_{ijt}=1$ if hospital i performed $1-24$ procedures on Medicare patients in year t, and 0 otherwise, $\text{MedLow}_{ijt}=1$ if hospital i performed $25-49$ procedures on Medicare patients in year t, and 0 otherwise, $\text{Medium}_{ijt}=1$ if hospital i performed $50-99$ procedures on Medicare patients in year t, and 0 otherwise, and $\text{MedHigh}_{ijt}=1$ if hospital i performed $100-199$ procedures on Medicare patients in year t, and 0 otherwise. The high volume category is omitted and is the reference. The remaining terms, in the summation, account for covariates. We used the GEE emprical or "sandwich" variance estimators to account for correlation among observations within hospital within and between years.

Surgical Site Infection Risk over Time

The Generalized Linear Mixed Models that we fit within each year can be expressed as follows:

$$\text{logit}(\Pr[y_{ij|b_i}=1]) = \frac{\Pr[y_{ij|b_i}=1]}{1+\Pr[y_{ij|b_i}=1]} = \beta_0 + \beta_1\text{Age}_{ij} + \beta_2\text{Gender}_{ij} + \sum_{k=3}^{K}\beta_k + \text{Comorbidity}_{ij} + b_i$$

where i and j index hospital and patient as above, and we assume b_i, a random intercept, is distributed normal.

The predicted values of those random intercepts, denoted $\tilde{b}_i$, are used to to rank hospitals within each year. These $\tilde{b}_i$ are what is plotted in Figure 2.

Additionally, we used the $\tilde{b}_i$ as the basis for the outcome in a final Generalized Estimating Equation model. If $\tilde{b}_i$ was in the bottom quartile for year $t+1$, next year, then $q_{i,t+1}=1$, otherwise it is 0. We similarly used the $\tilde{b}_i$ to make three categories of predictor for the $q_{i,t+1}$: $\text{wty}_{it}=1$ if $\tilde{b}_i$ for hospital i was in the bottom (*w*orst) quartile in year t (*t*his *y*ear) and $\text{wly}_{it}=1$ if $\tilde{b}_i$ for hospital i was in the bottom quartile in year $t-1$, (*l*ast *y*ear) and 0 otherwise. In addition, mty_{it} and mly_{it} were similarly defined. The best quartile indicators were omitted as reference categories.

The models for worst quartile next year as a function of only the category this year (Table 2, leftmost column) can be expressed as:

$$\text{logit}(\Pr[q_{i,t+1}=1]) = \frac{\Pr[q_{i,t+1}=1]}{1+\Pr[q_{i,t+1}=1]} = \beta_0 + \beta_1\text{wty}_{it} + \beta_2\text{mty}_{it}$$

Including the effect of procedure volume (Table 2, other columns), the model is:

$$\text{logit}(\Pr[q_{i,t+1}=1]) = \beta_0 + \beta_1\text{wty}_{it} + \beta_2\text{mty}_{it} + \beta_3\text{Low}_{ijt} + \beta_4\text{MedLow}_{ijt} + \beta_5\text{Medium}_{ijt} + \beta_6\text{MedHigh}_{ijt}$$

And finally, adding in the effect of this year and last years' quality quartiles in predicting next year (Table 3), we have:

$$\text{logit}(\Pr[q_{i,t+1}=1]) = \beta_0 + \beta_1\text{wty}_{it} + \beta_2\text{mty}_{it} + \beta_1\text{wly}_{it} + \beta_2\text{mly}_{it} + \beta_3\text{Low}_{ijt} + \beta_4\text{MedLow}_{ijt} + \beta_5\text{Medium}_{ijt} + \beta_6\text{MedHigh}_{ijt}$$

Each of the GEE models uses the independence working assumption.