Stroke Mortality Calculation Algorithm

We estimated a logistic regression model and a hierarchical generalized linear model which accounts for the clustering of observations within hospitals. The logistic regression model links the outcome to the patient-level risk factors.[20] Let Y_{ij} denote the outcome (equal to 1 if patient dies, zero if patient lives) for the jth patient who had a stroke admission at the ith hospital; $\mathbf{Z}_{ij}$ denotes a set of risk factors based on the data. Let I denote the total number of hospitals and n_i the number of index patient stays in hospital i. We assume the outcome is related linearly to the covariates via a known linked function, h, where

$$\text{Logistic regression model:} \qquad h(Y_{ij}) = \alpha + \beta \mathbf{Z}_i \qquad (1)$$

and $\mathbf{Z}_{ij} = (Z_{1ij}, Z_{2ij}, ..., Z_{pij})$ is a set of p patient-specific covariates. In our case, h = the logit link.

To account for the natural clustering of observations within hospitals, we then estimate a hierarchical logistic regression model that links the risk factors to the same outcome and a hospital-specific random effect,

$$\text{Hierarchical logistic regression model:} \qquad h(Y_{ij}) = \alpha_i + \beta \mathbf{Z}_{ij} \qquad (2)$$
$$\alpha_i = \mu + \omega_i; \qquad \omega_i \sim N(0, \tau^2) \qquad (3)$$

where α_i represents the hospital-specific intercept, $\mathbf{Z}_{ij}$ is defined as above, μ the adjusted average outcome over all hospitals in the sample, and τ^2 the between-hospital variance component.[21] This model separates within-hospital variation from between-hospital variation. Both hierarchical logistic regression models and logistic regression models are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

We first fit the logistic regression model described in Equation (1) using the logit link. Having identified the covariates that were selected, we next fit the hierarchical logistic regression model described in Equations (2) and (3), again using the logit link function; e.g.,

$$\text{Logit } (P(Y_{ij} = 1)) = \alpha_i + \beta \, \mathbf{Z}_{ij}$$
$$\alpha_i = \mu + \omega_i, \quad \omega_i \sim N(0, \tau^2)$$

where $\mathbf{Z}_{ij}$ consisted of the covariates retained in the logistic regression model. As before, $Y_{ij} = 1$ if patient j treated at hospital i had the event; 0 otherwise.

Hospital Performance Reporting

Using the set of risk factors in the logistic regression model, we fit the hierarchical logistic regression model defined by Equations (2) - (3) and estimate the parameters $\hat{\mu}$, $\{\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_I\}$, $\hat{\beta}$ and $\hat{\tau}^2$. We calculate a standardized outcome, s_i, for each hospital by computing the ratio of the number of predicted deaths to the number of expected deaths, multiplied by the unadjusted overall mortality rate, $\bar{y}$. Specifically, we calculate

$$\text{Predicted} \qquad \hat{y}_{ij}(Z) = h^{-1}(\hat{\alpha}_i + \hat{\beta}\, \mathbf{Z}_{ij}) \qquad\qquad (4)$$

$$\text{Expected} \qquad \hat{e}_{ij}(Z) = h^{-1}(\hat{\mu} + \hat{\beta}\, \mathbf{Z}_{ij}) \qquad\qquad (5)$$

$$\hat{s}_i(Z) = \frac{\sum_{j=1}^{n_i} \hat{y}_{ij}(Z)}{\sum_{j=1}^{n_i} \hat{e}_{ij}(Z)} \times \bar{y} \qquad\qquad (6)$$

If more (fewer) "predicted" cases than "expected" cases have the outcome in a hospital, then s_i will be higher (lower) than the unadjusted average. For each hospital, we compute an interval estimate of s_i to characterize the level of uncertainty around the point estimate using bootstrapping simulations. The point estimate and interval estimate can be used to characterize and compare hospital performance (e.g., higher than expected, as expected, or lower than expected).

Creating Interval Estimates

Because the statistic described in Equation (6) is a complex function of parameter estimates, we use re-sampling and simulation techniques to derive an interval estimate. The bootstrapping simulation has the advantage of avoiding unnecessary distributional assumptions.

Calculation Algorithm

Let *I* denote the total number of hospitals in the sample. We repeat steps 1 – 4 below for b = 1,2,…B times:

1. Sample *I* hospitals with replacement.

2. Fit the hierarchical logistic regression model using all patients within each sampled hospital. We use as starting values the parameter estimates obtained by fitting the model to all hospitals. If some hospitals are selected more than once in a bootstrapped sample, we

treat them as distinct so that we have *I* random effects to estimate the variance components. At the conclusion of Step 2, we have:

a. $\hat{\beta}^{(b)}$ (the estimated regression coefficients of the risk factors).

b. The parameters governing the random effects, hospital-adjusted outcomes, distribution, $\hat{\mu}^{(b)}$ and $\hat{\tau}^{2(b)}$.

c. The set of hospital-specific intercepts and corresponding variances, $\{\hat{\alpha}_i^{(b)}, \ \text{vâr}(\alpha_i^{(b)}); \ i = 1,2,...,I\}$.

3. We generate a hospital random effect by sampling from the distribution of the hospital-specific distribution obtained in Step 2c. We approximate the distribution for each random effect by a normal distribution. Thus, we draw $\alpha_i^{(b*)} \sim N(\hat{\alpha}_i^{(b)}, \text{vâr}(\hat{\alpha}_i^{(b)}))$ for the unique set of hospitals sampled in Step 1.

4. Within each unique hospital *i* sampled in Step 1, and for each case *j* in that hospital, we calculate $\hat{y}_{ij}^{(b)}$, $\hat{e}_{ij}^{(b)}$, and $\hat{s}_i(Z)^{(b)}$ where $\hat{\beta}^{(b)}$ and $\hat{\mu}^{(b)}$ are obtained from Step 2 and $\hat{\alpha}_i^{(b*)}$ is obtained from Step 3.

Ninety-five percent interval estimates (or alternative interval estimates) for the hospital-standardized outcome can be computed by identifying the 2.5[th] and 97.5[th] percentiles of randomly half of the B estimates (or the percentiles corresponding to the alternative desired intervals).[22] (See Figure 3 below for a diagram of the analysis steps).

Figure 1. Analysis Steps

Step 1:
Compute Bivariate and Univariate
summaries
Z & Y

Step 2:
Logistic Regression Model
h(Yij) = α^A + β^{A}Zij
Obtain residuals, etc.

Step 3:
Hierarchical Logistic Regression Model
h(Yij) = α_i^A + β^{A}Z$_{ij}$
$\alpha_i^{(A)} \sim N\left(\mu_A, \tau_A^2\right)$

Step 4:

Hospital-Specific Predicted
Outcomes
$\ddot{y}_i^A(\mathbf{Z}) = \dfrac{1}{n_i} \sum_{j=1}^{n_i} h^{-1}(\ddot{\omega}_i^A + \ddot{\beta}^A \mathbf{Z}_{ij})$

Hospital-Specific Expected
Outcomes
$\hat{e}_i^A(\mathbf{Z}) = \dfrac{1}{n_i} \sum_{j=1}^{n_i} h^{-1}(\hat{\mu}_A + \hat{\beta}^A \mathbf{Z}_{ij})$

Hospital-Specific Risk-
Standardized Outcomes
$\hat{r}_i^A(Z) = \dfrac{\hat{y}_i^A(Z)}{\hat{e}_i^A(Z)}$

$\hat{s}_i^A(Z) = \dfrac{\hat{y}_i^A(Z)}{\hat{e}_i^A(Z)} \times \bar{y}$