The STS Surgeon Composite Technical Appendix

Overview

Surgeon-specific risk-adjusted operative operative mortality and major complication rates were estimated using a bivariate random-effects logistic regression model. The term bivariate refers to the fact that both operative mortality and major complications were analyzed together in a single model, not estimated one at a time in separate models. Random-effects refers to the assumption that the provider-specific parameters of interest are assumed to arise from a specified distribution defined by parameters that are also estimated in the modelling process. To adjust for case mix, we first calculated each patient's risk score for operative mortality and each patient's risk score for major complications using an existing set of STS risk models. The goal of calculating a patient risk score was to reduce the number of covariates in the hierarchical model by summarizing the predictive information from a large number of baseline covariates into a single number. Adjustment for each covariate individually in the hierarchical model would be theoretically preferable but computationally impractical due to the large number of records and covariates, and the computationally intensive nature of Bayesian hierarchical model estimation.

Calculation of Risk Scores

For each patient, risk scores for operative mortality and major complications were calculated from existing models that were specific to the individual patient's type of operation.

- For patients undergoing isolated CABG, isolated AVR, or isolated AVR + CABG, risk scores were calculated according to the published STS 2008 mortality and major complications models for isolated CABG, isolated valve, or isolated valve + CABG, respectively. To ensure high calibration for the current study cohort, coefficients of each model were re-estimated using the current 3-year study sample and current endpoint definitions.

- For patients undergoing a mitral operation without CABG, risk scores were calculated using a modified version of the published STS 2008 mortality and major complications models for isolated valve procedures. The STS 2008 models were modified to account for the inclusion of patients undergoing tricuspid repair and to provide a more detailed adjustment for endocarditis and degree of tricuspid regurgitation. Briefly, the modified models included a new variables for tricuspid repair (yes/no), a new variable for treated endocarditis (yes/no), and a more detailed adjustment for degree of tricuspid (less than moderate, moderate, severe). Coefficients of the modified models were estimated using the current 3-year study cohort and current endpoint definitions.

- For patients undergoing a mitral operation with concomitant CABG, risk scores were calculated using a modified version of the published STS 2008 mortality and major complications models for isolated valve + CABG operations. Modifications to these isolated valve + CABG models were identical to the modifications of the isolated valve models described above. Coefficients of the modified models were estimated using the current 3-year study cohort and current endpoint definitions.

A patient's mortality risk score was defined as the patient's predicted risk of operative mortality and a patient's major complication risk score was defined as the patient's predicted risk of major complication.

Hierarchical Model

For the i-th of n_j patients treated by the j-th surgeon ($j = 1, 2, \ldots, N$), let p_{1ji} be patient's predicted risk of operative mortality (i.e. mortality risk score), let p_{2ji} be the patient's predicted risk of major complications (i.e. complications risk score), let $\bar{p}_{1j} = \sum_{i=1}^{n_j} p_{1ij}/n_j$ be the average mortality risk score for surgeon j, and let $p_{2j} = \sum_{i=1}^{n_j} p_{2ij}/n_j$ be the average complications risk score for surgeon j. Let Y_{1ji} be a binary indicator of operative mortality status (0=alive, 1=dead), let Y_{2ji} be an indicator of major complications (0 = none, 1 = at least one), and let $\pi_{kji} = \Pr(Y_{kji} = 1|p_{kji})$ be the probability of the occurrence of the k-th endpoint where $k = 1$ refers to mortality and $k = 2$ refers to complications. The associations of p_{1ji} and p_{2ji} with Y_{kji} and Y_{2ji} were assumed to be described by a bivariate random effects logistic regression model with normally distributed surgeon-specific random intercept parameters. At the first level, we specified the following within-surgeon regression model:

$$\begin{array}{ll} \text{(operative mortality)} & \log\left(\frac{\pi_{1ji}}{1-\pi_{1ji}}\right) = \alpha_{1j} + x_{1ji}\beta_1 \\ \text{(major complication)} & \log\left(\frac{\pi_{2ji}}{1-\pi_{2ji}}\right) = \alpha_{2j} + x_{2ji}\beta_2 \end{array}$$

where $x_{kji} = \log(p_{kji}/(1 - p_{kji}))$, β_1 is an unknown parameter relating risk scores to mortality, β_2 is an unknown parameter relating risk scores to major complications, and α_{1j} and α_{2j} are surgeon-specific intercept parameters (random effects). Conditional on π_{1ji} and π_{2ji}, the variables Y_{1ji} and Y_{2ji} were assumed to be distributed as two independent Bernoulli variables with parameters π_{1ji} and π_{2ji}, respectively. That is:

$$\Pr(Y_{1ji} = y_{1ji}, Y_{2ji} = y_{2ji} | \pi_{1ji}, \pi_{2ji}) = \prod_{k=1}^{2} \pi_{kji}^{y_{kji}} (1 - \pi_{kji})^{1-y_{kji}}.$$

Outcomes of patients of different surgeons were assumed to be statistically independent, and outcomes of patients treated by the same surgeon were assumed to be conditionally independent given $(\alpha_{1j}, \alpha_{2j})$. The assumption that Y_{1ji} and Y_{2ji} are conditionally independent given π_{1ji} and π_{2ji} is likely to be violated in practice but was made in order to facilitate computation. Although the model assumes *conditional* independence between Y_{1ji} and Y_{2ji}, the model does not assume *marginal* independence between these two variables, as the underlying probabilities π_{1ji} and π_{2ji} depend on random effects parameters α_{1j} and α_{2j} which are assumed to be correlated, as described below.

At the second level, we specified the following between-surgeon model:

$$\begin{bmatrix} \alpha_{1j} \\ \alpha_{2j} \end{bmatrix} \overset{\text{ind}}{\sim} N\left(\begin{bmatrix} \gamma_{10} + \gamma_{11} x_{1j}^* \\ \gamma_{20} + \gamma_{21} x_{2j}^* \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right)$$

where $x_{kj}^* = \log(\bar{p}_{kj}/(1 - \bar{p}_{kj}))$, and where γ_{10}, γ_{11}, γ_{20}, γ_{21}, σ_{11}, σ_{12}, and σ_{22} are unknown parameters to be estimated from the data. In other words α_{1j} and α_{2j} were assumed to be independent draws from a bivariate normal distribution with a mean depending on the surgeon's average case mix, as reflected in x^*_{1j} and x^*_{2j}. The assumption that pairs $(\alpha_{1j}, \alpha_{2j})$ are independent is somewhat artificial as we believe that outcomes of surgeons operating at the same hospital may be correlated. Implicitly, we assume that a hospital's effect on outcomes is captured in the surgeon-level parameters $(\alpha_{1j}, \alpha_{2j})$.

Accounting for x^*_{1j} and x^*_{2j} in the surgeon-level model is intended to reduce confounding in the estimation of β_1 and β_2 in the patient-level model. As noted by various authors (e.g. Neuhaus and Kalbfleisch, Biometrics, 1998; Localio et al. Ann Intern Med, 2001; Neuhaus and McCulloch, JRSS-B, 2006), inclusion of only individual-level covariates in a hierarchical model can produce biased estimates when covariates and random effects are correlated. Inclusion of both individual-level covariates such as x_{1ji} and x_{2ji} and surgeon-level covariates such as x^*_{1j} and x^*_{2j} allows partitioning covariate effects into within-surgeon and between-surgeon components in order to produce consistent estimates of the within-surgeon model parameters β_1 and β_2. In addition, inclusion of cluster-level covariates in hierarchical models can potentially enhance the precision in the estimation of the cluster-specific parameters, which in our case are α_{1j} and α_{2j}.

Definition of Risk-Adjusted Rates

Based on this model, the j-th surgeon's risk-adjusted rates of operative mortality and major complications were defined as

$$\text{(operative mortality)} \qquad \theta_{1j} = \frac{\sum_{i=1}^{n_j} \text{expit}(\alpha_{1j} + x_{1ji}\beta_1)}{\sum_{i=1}^{n_j} \text{expit}(\text{constant}_1 + x_{1ji}\beta_1)} \times \overline{Y}_1$$

$$\text{(major complication)} \qquad \theta_{2j} = \frac{\sum_{i=1}^{n_j} \text{expit}(\alpha_{2j} + x_{2ji}\beta_2)}{\sum_{i=1}^{n_j} \text{expit}(\text{constant}_2 + x_{2ji}\beta_2)} \times \overline{Y}_2$$

where $\overline{Y}_1$ denotes the overall aggregate observed rate of operative mortality in the study sample, $\overline{Y}_2$ denotes the overall aggregate observed rate of major complication in the study sample, and constant_1 and constant_2 are chosen to represent "typical" values of α_{1j} and α_{2j}, respectively.

Definition of Composite Score

The overall composite score of the j-th surgeon was defined as

$$\theta_j = w(1 - \theta_{1j}) + (1 - w)(1 - \theta_{2j})$$

where $w = (1/\sigma_1)/(1/\sigma_1 + 1/\sigma_2)$ and σ_k denotes the standard deviation of the θ_{kj}'s across surgeons, $k = 1, 2$.

Estimation

Model parameters were estimated in a Bayesian framework by specifying a prior probability distribution for the unknown model parameters β_1, β_2, γ_{10}, γ_{11}, γ_{20}, γ_{21}, σ_{11}, σ_{12}, σ_{22}. Because our prior knowledge was limited, we specified a vague proper prior distribution that consisted of independent normal distributions for regression coefficients (β_1, β_2, γ_{10}, γ_{11}, γ_{20}, γ_{21}), and an inverse Wishart distribution for variance parameters, $\mathbf{\Sigma} = (\sigma_{11}, \sigma_{12}, \sigma_{22})$. Posterior means and credible intervals were calculated using Markov Chain Monte Carlo (MCMC) simulations as implemented in OpenBUGS version 3.2.2 software. Posterior summaries were calculated by generating 20,000 sets of simulated parameter values after a burn-in period of 1,000 MCMC iterations to ensure convergence. Adequacy of the number of MCMC iterations was assessed by the methods of Raftery and Lewis (1992) and Geweke (1991) as implemented in the CODA add-on package for R statistical software. The parameter θ_j was estimated as $\hat{\theta}_j = \sum_{l=1}^{20000} \theta_j^{(l)} \Big/ 20000$, where $\theta_j^{(l)}$ denotes the simulated values of θ_j at the l-th iteration of the MCMC procedure. A 98% Bayesian credible interval was obtained by calculating the 200th lowest and 200th highest values of θ_j across the 20,000 simulated values.

Estimation of Reliability

Reliability is conventionally defined as the proportion of variation in a measure that is due to true between-unit differences (i.e., signal) as opposed to random statistical fluctuations (i.e., noise). Equivalently, it is the squared correlation between a measurement and the true value. Accordingly, reliability was defined as the square of the Pearson correlation coefficient (ρ^2) between the set of surgeon-specific estimates $\hat{\theta}_1, \ldots, \hat{\theta}_N$ and the corresponding unknown true values $\theta_1, \ldots, \theta_N$, that is:

$$\rho^2 = \frac{\sum_{j=1}^{N} (\hat{\theta}_j - \frac{1}{N}\sum_{h=1}^{N} \hat{\theta}_h)(\theta_j - \frac{1}{N}\sum_{h=1}^{N} \theta_h)}{\sum_{j=1}^{N} (\hat{\theta}_j - \frac{1}{N}\sum_{h=1}^{N} \hat{\theta}_h)^2 \sum_{j=1}^{N} (\theta_j - \frac{1}{N}\sum_{h=1}^{N} \theta_h)^2}$$

The quantity ρ^2 was estimated by its posterior mean, namely,

$$\hat{\rho}^2 = \frac{1}{20000} \sum_{l=1}^{20000} \rho^2_{(l)}$$

where

$$\rho^2_{(l)} = \frac{\sum_{j=1}^{N} (\hat{\theta}_j - \frac{1}{N} \sum_{h=1}^{N} \hat{\theta}_h)(\theta_j^{(l)} - \frac{1}{N} \sum_{h=1}^{N} \theta_h^{(l)})}{\sum_{j=1}^{N} (\hat{\theta}_j - \frac{1}{N} \sum_{h=1}^{N} \hat{\theta}_h)^2 \sum_{j=1}^{N} (\theta_j^{(l)} - \frac{1}{N} \sum_{h=1}^{N} \theta_h^{(l)})^2}$$

with $\theta_h^{(l)}$ denoting the value of θ_j on the l-th MCMC sample $\hat{\theta}_j = \sum_{l=1}^{20000} \theta_j^{(l)}/20000$ denoting the posterior mean of θ_j. A 95% credible interval for ρ^2 was obtained by calculating the 500th smallest and 500th largest values of $\rho^2_{(l)}$ across the 20,000 MCMC samples.

References

Geweke, John. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Vol. 196. Federal Reserve Bank of Minneapolis, Research Department, 1991.

Localio, A. Russell, Berlin, Jesse A., Ten Have, Thomas R., and Kimmel, Stephen E. (2001). Adjustments for center in multicenter studies: an overview. *Annals of Internal Medicine*, **135**, 112–23.

Neuhaus, John M., and Jack D. Kalbfleisch. (1998). Between-and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, **54**, 638–645.

Neuhaus, John M., and Charles E. McCulloch. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 859–872.

Raftery, Adrian E. and Lewis, Steven. How many iterations in the Gibbs sampler? In *Bayesian Statistics 4* (J.M. Bernardo et al., editors), Oxford University Press, pp. 763-773.