SUMMARY

While risk-adjusted outcomes are often used to compare the performance of hospitals and physicians, the most appropriate functional form for the risk adjustment process is not always obvious for continuous outcomes such as costs. Semi-log models are used most often to correct skewness in cost data, but there has been limited research to determine whether the log transformation is sufficient or whether another transformation is more appropriate. This study explores the most appropriate functional form for risk-adjusting the cost of coronary artery bypass graft (CABG) surgery. Data included patients undergoing CABG surgery at four hospitals in the Midwest and were fit to a Box-Cox model with random coefficients using Markov chain Monte Carlo methods. Marginal likelihoods and Bayes factors were calculated to perform model comparison of alternative model specifications. Rankings of hospital performance were created from the simulation output and the rankings produced by Bayesian estimates were compared to rankings produced by standard models fit using classical methods. Results suggest that, for these data, the most appropriate functional form is not logarithmic, but corresponds to a Box-Cox transformation of $-1$. Furthermore, Bayes factors overwhelmingly rejected models that assumed homogenous transformations. However, the hospital ranking induced by the BCRC model was not different from the ranking produced by maximum likelihood estimates of either linear or the semi-log model.

Keywords: Risk adjustment, Box-Cox transformation, random coefficients, Markov chain Monte Carlo, coronary artery bypass graft
1 INTRODUCTION

In recent years there has been increased interest in improving accountability by evaluating and comparing the performance of health care providers. Most often this is done by ranking performance based on patient outcomes such as length of stay [1], survival [2], and cost of treatment [3]. Because health care providers treat heterogeneous populations, however, patient outcomes must be adjusted to account for severity in order to make meaningful comparisons between healthcare providers.

Outcomes are risk-adjusted by regressing the performance measure on an index of severity, or on patient-specific variables, including risk factors and comorbidities. The residuals from this regression, which have netted out patient-specific severity effects, are used to produce the ranking. Thus, the choice of model, both in terms of independent variables and the proposed functional relationship between performance and risk, is critically important. When the performance measure is a binary outcome, such as survival, there is little question that a probit or logit is the appropriate statistical model. For continuous performance measures such as cost, the most appropriate model is less obvious. The semi-log model is most often used to risk adjust costs because costs tend to be skewed, but the linear model is also used [4,5].

There has been little research into how appropriate these models are in the context of risk adjustment, whether another transformation is more appropriate, or whether the transformation is dependent upon context, for example the particular medical condition or type of healthcare provider under investigation. A recent study by Schnitzler et al. [6], who used three alternative models to risk adjust the cost of treating community acquired pneumonia at 6 hospitals and produced a rank ordering of physicians by cost, suggests that the ranking is highly model dependent. Fitting the risk adjustment model using the linear model, the semi-log model, and robust estimation they found that physicians who ranked in the highest (or lowest) 10 percent when cost was regressed on severity and patient
characteristics often ranked much lower (or higher) when the log of cost was used as the dependent variable or when a robust estimation procedure was used.

Using the most appropriate functional form for the risk-adjustment process is important to the extent that patients and healthcare providers use this information. Managed care organizations use risk-adjusted performance to select physicians for inclusion in their referral networks [7], and consumers may use this information to choose healthcare providers. If consumers and healthcare providers make decisions that are motivated by the information then inaccurate information may lead to suboptimal allocation of health care resources.

The Box-Cox model, a flexible functional form model first proposed by Box and Cox in 1962, has been widely used to study issues of functional form in economics and applied statistics [8-10]. This paper uses a Box-Cox model with random coefficients (BCRC) to risk-adjust the cost of treating patients who undergo coronary artery bypass graft (CABG) procedures. The model is fit using Markov chain Monte Carlo (MCMC) methods. Several issues are addressed using the model, including the most appropriate transformation of costs; the probability that Box-Cox transformation parameter is in the neighborhood of 0 or 1, which corresponds to the natural log transformation and no transformation, respectively; whether it can be assumed that the transformation is equal across hospitals; and how the ranking induced by risk adjustment using the BCRC model compares to the rankings produced by maximum likelihood estimation of the default models.

The paper proceeds as follows. Section 2 describes the data used in the study. Section 3 describes the model, the priors, the MCMC algorithm, model selection, and the model fitting. Results are contained in Section 4, including a summary of the marginal posterior distributions, model comparison, and hospital rankings, and Section 5 concludes with discussion.
2 DATA

Data used in this study were originally used by the Greater St. Louis Healthcare Alliance, a voluntary coalition of regional hospitals, physicians, and managed care organizations formed in 1992 with a mission of improving the quality of healthcare delivery and reducing costs. This study uses the subset of 496 CABG patients from four hospitals in the BJC Health System that was used in the risk adjustment analysis of the 1996 Comprehensive Hospital Performance Report [11].

Patients were selected by a panel of cardiothoracic surgeons from among all CABG surgery patients at four regional hospitals in the St. Louis region between January 1, 1995, and December 31, 1995. The data included a random sample of patients who were diagnosed with one or more of the following International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) principal diagnosis codes: Bypass anastomosis for heart revascularization (36.1), Hypertensive heart disease with congestive heart failure (CHF) (402), Hypertensive heart and renal disease with CHF (404), Hypertensive heart and renal disease with CHF and renal failure (404), Acute myocardial infarction (410), Other acute and subacute forms of ischemic heart disease (411), Old myocardial infarction (412), Angina pectoris (413), Other forms of chronic ischemic heart disease (414), Heart failure (428), Cardiogenic shock (785.51), Chest pain not otherwise specified (786.50), Mechanical complication of cardiac device (996.03), Other complications of internal prosthetic device (996.72), Accidental cut during catheterization (E870.6), Surgical procedure as cause of later complication (E878.2), and Cardiac catheterization as cause of later complication (E879.0).

Patients were excluded from the study if one or more of the following procedures were performed: Operations on valves and septa of heart (35), Excision of aneurysm (37.32), Resection of aorta with anastomosis (38.34), Resection of abdominal aorta with replacement (38.44), Resection of thoracic aorta with replacement (38.45), Other excision of aorta (38.64), Other repair on vessels and re-entry
operation (39.54), Cervical esophagostomy (42.11), and Other bronchial excision (321). Patients were also excluded if they were categorized under the following DRGs: Cardiac valve procedure with catheterization (104) or Cardiac valve procedure without catheterization (105).

Clinical variables, determined by the panel of cardiovascular surgeons, were abstracted from patient charts, and the cost of each patient's CABG surgery admission was obtained from the hospitals’ internal cost accounting databases, which estimated costs by a departmental ratio of cost-to-charges methodology.

**Definition of Variables**

Clinical variables used to risk-adjust the cost of CABG include blood albumin level (ALBUMIN), worsening congestive heart failure (CHF), prior open heart surgery (PRIOR), previous valve surgery (VALVE), hematocrit level (HEMATOCRIT), alveolar arterial oxygen gradient (AAO2), percutaneous transluminal coronary angioplasty (PTCA), and cardiac catheter (CATH). Serum albumin is the primary protein in the blood. Lower levels of albumin are associated with diabetes, renal insufficiency, and malnutrition [12]. Therefore, the coefficient for ALBUMIN is expected to be negative. CHF, or the inability for the heart to provide adequate blood flow to vital organs, is a serious comorbidity for patients who undergo CABG surgery.[13] The variable used in this analysis is a dummy variable that equals 1 if the patient's CHF had grown worse within 30 days prior to surgery. Patients with worsening CHF are expected to accrue higher costs. Previous CABG and valve surgery have been shown to increase risk for surgical site infections, perioperative death, and post-operative adverse outcomes [14,15]. Since these lead to higher costs, VALVE and PRIOR are expected to have positive coefficients. Hematocrit is the proportion of blood volume represented by red blood cells. Low levels have implications for oxygenation of tissues and blood viscosity. The normal range for adult males is 42-54%, and for females is 38-46%, and the variable used in this analysis is a dummy
variable that equals 1 for hematocrit below 33%. The coefficient for HEMATOCRIT is expected to be positive. AAO2, a measure of lung efficiency, is the difference between the partial pressure of oxygen in the alveoli and the partial pressure of oxygen in the arterial blood. Normal AAO2 ranges between 5 and 25 mmHg. The variable used in this analysis is a dummy variable equal to 1 for AAO2 levels greater than 30 mmHg. Higher levels are associated with decreased efficiency, therefore a positive coefficient is expected for this variable. PTCA, or balloon angioplasty, opens clogged coronary arteries via an inflated balloon in the artery. Prior PTCA may indicate higher risk since the patient subsequently underwent surgery, and is expected to have a positive sign. Finally, cardiac catheterization is a procedure that locates coronary arterial blockages. Patients who receive catheterization are expected to incur higher costs.
3 THE BOX-COX RANDOM COEFFICIENTS MODEL

The Model

The clinical and cost data were fit to the BCRC model using a Bayesian MCMC algorithm. This section describes the model, priors, the simulation algorithm, and the method used for model selection.

Let \(i=1,...,n\) index clusters and \(j=1,...,n_i\) index within-cluster observations. The BCRC model is

\[
y_{ij}^{(\lambda_i)} = x_j^T b_i + \epsilon_{ij}
\]

(1)

where the Box-Cox transformation is

\[
y_{ij}^{(\lambda)} = \frac{y_{ij}^{\lambda_i} - 1}{\lambda_i}, \quad \epsilon_{ij} \sim N(0, \sigma^2), \quad \text{and} \quad x_{ij} \text{ is a } (k \times 1) \text{ vector of covariates}[16]. \]

The collection of transformation parameters is \(\lambda = \lambda_1,...,\lambda_n\). Assume further that any heterogeneity in the constant and slope parameters across clusters is stochastic: \(\{b\}_i \sim N(\beta, D^{-1})\). \(\beta\) is the mean of the slope parameters, and the diagonal elements of the covariance matrix \(D\) gives some indication of the variation across clusters. The likelihood function for this model is

\[
f(y_i | \beta, D, \sigma^2, \lambda) = \frac{1}{(2\pi)^{\frac{n_i}{2}}} |V_i|^{\frac{1}{2}} \exp\left\{-\frac{1}{2} (y_i^{(\lambda_i)} - X_i \beta)' V_i^{-1} (y_i^{(\lambda_i)} - X_i \beta) \right\} \prod_{j=1}^{n_i} y_{ij}^{\lambda_i - 1}
\]

(2)

where \(V_i = \sigma^2 I + X_i DX_i^T\) and the product term on the right hand side is the Jacobian of the transformation. The objects of inference are \(\beta, D, \sigma^2, \) and \(\lambda\). The model allows heterogeneous transformations and unbalanced clusters, which increases the difficulty of estimation by maximum likelihood by increasing the dimension of the search space. Bayesian methods were used to fit the model because they often handle high dimensions more readily than frequentist techniques.

Priors

A Bayesian analysis proceeds by first choosing a prior distribution for the parameters of interest. Note that conditional on a value for the transformation parameter, the BCRC model is a hierarchical linear
model. Therefore, priors were chosen from conjugate families that are standard for hierarchical linear models and were assigned in two stages: first to the parameters of the model, \( \{ b \}_i, \sigma^2, \) and \( \{ \lambda \}_i, \) and second to the parameters of \( \beta \) and \( D \) \cite{17,18}. The assumption of random coefficients implies a first stage prior for \( \{ b \}_i \) of a \( k \)-variate normal distribution with mean \( \beta \) and covariance matrix \( D \). The prior for \( \sigma^2 \) is an inverse gamma distribution with parameters \( \nu/2 \) and \( \delta/2 \), and the prior for \( \{ \lambda \}_i \) is a normal distribution with mean \( \lambda_0 \) and variance \( \tau^2 \). In the second stage, the prior for \( \beta \) is a normal distribution with mean \( \beta_0 \) and covariance matrix \( B_0 \), and the prior for \( D \) is an inverse Wishart distribution with \( \eta \) degrees of freedom and scale matrix \( R \).

**MCMC Simulation**

Bayesian inference is carried out on the posterior distribution, which is the product of the likelihood function and priors. The marginal posteriors distributions are summarized by integrating the posterior distribution over parameters of interest. Of course, the posterior distribution for the BCRC model is from an unknown family of distributions, is high dimensional, and has an unknown normalizing constant. It is not, therefore, amenable to analytic integration, but it can be summarized by numerical methods using MCMC simulation. The idea behind MCMC simulation as applied to Bayesian analysis is that a Markov chain is constructed whose transition distribution converges to the posterior distribution. Then, starting from an arbitrary point, the chain is allowed to run until it converges to its stationary distribution. The initial draws, generated while the chain is in its transient stage, are discarded and the remaining draws are a (possibly correlated) sample from the posterior distribution that is used for purposes of inference \cite{19}.

The BCRC model may be simulated using a Metropolis-within-Gibbs framework, a hybrid approach that combines the Gibbs sampler and the Metropolis-Hastings algorithm, two MCMC methods that have been widely used in applied Bayesian analysis because of their ease of implementation \cite{19}. The
algorithm takes successive draws from the full conditional distributions of the parameters conditioning on previous draws from other parameters. The full conditional distributions can be shown to be a normal distribution for \( \{b_i\} \), and \( \beta \), an inverse Wishart distribution for \( D \), and an inverse gamma distribution for \( \sigma^2 \) [16]. These parameters may be drawn successively, conditioning on previous draws, in a Gibbs scheme. The full conditional distribution of \( \{\lambda_i\} \), is from an unknown family of distributions and is drawn with the Metropolis-Hastings algorithm. A candidate draw, \( \lambda_j^{(g+1)} \) is first generated by finding the mode of the full conditional distribution using the Newton-Raphson algorithm, and then taking a draw from a normal distribution centered at the mode with variance equal to the curvature of the full conditional at the mode. In the Metropolis-Hastings step, the candidate draw at iteration \( g+1 \) is accepted with probability \[
\min \left\{ \frac{\pi(\lambda_j^{(g+1)})}{\pi(\lambda_j^{(g)})}, 1 \right\} \]
If the candidate draw is rejected, then \( \lambda_j^{(g+1)} = \lambda_j^{(g)} \). Notice that candidate draws that move the Markov chain to a higher point on the posterior distribution are accepted with certainty, but candidate draws that move the Markov chain to a lower point on the distribution are not. In this way the Markov chain tends toward high probability regions of the posterior.

**Model Selection**

The goal of model selection in a Bayesian analysis is to determine the posterior odds that a particular model generated the data. This is important for this study since the primary use of the BCRC model will be to discriminate between alternative risk adjustment model specifications (e.g. linear and semi-log). Assume we are interested in comparing a finite collection of models \( M = \{M_1, \ldots, M_I\} \), where a model consists of a likelihood function and a prior: \( M_i = \{f(y \mid \Theta_i, \theta_i), \pi(\theta_i \mid M_i)\} \). Define the prior probability of the truth of model \( M_i \) as \( \Pr(M_i) = p_i \). Then by Bayes theorem, after observing the data, the posterior probability of the truth of model \( M_i \) is

\[
\Pr(M_i \mid y) \propto p_i m(y \mid M_i) \tag{3}
\]
where $m(y \mid M_i)$ is the marginal likelihood of $y$, given by

$$m(y \mid M_i) = \int f(y \mid M_i, \theta_j)\pi(\theta_j \mid M_i)d\theta_j. \quad (4)$$

The marginal likelihood is the likelihood function averaged over the prior, and the normalizing constant of the posterior distribution under model $M_i$.

After observing the data, the posterior odds in favor of model $M_i$ over model $M_j$ is given by the posterior odds ratio:

$$\frac{\Pr(M_i \mid y)}{\Pr(M_j \mid y)} = \frac{p_i}{p_j} \cdot \frac{m(y \mid M_i)}{m(y \mid M_j)}. \quad (5)$$

The ratio of marginal likelihoods on the right hand side of equation (5) is called the Bayes factor. If all models are assigned equally prior probability then the posterior odds equals the Bayes factor.

In this study I use Chib’s basic marginal likelihood identity approach to compute marginal likelihoods and posterior odds directly from the MCMC output [20]. Chib observed that by simply rearranging the posterior distribution, the marginal likelihood may be written as

$$m(y \mid M_i) = \frac{f(y \mid M_i, \theta_j^*)\pi(\theta_j^* \mid M_i)}{\pi(\theta_j \mid M_i, y)}. \quad (6)$$

Since $m(y \mid M_i)$ is not a function of $\theta$, equation (6) is an identity for any value $\theta^*$ and Chib’s estimate of the marginal likelihood can be written in log scale as

$$\ln \hat{m}(y \mid M_i) = \ln f(y \mid M_i, \theta_j^*) + \ln \pi(\theta_j^* \mid M_i) - \ln \pi(\theta_j \mid M_i, y) \quad (7)$$

Computing the marginal likelihood requires only an estimate of the ordinate of the likelihood function, the priors and the posterior, all evaluated at some $\theta^*$. Theoretically, $\theta^*$ could be any value from the support of $\theta$, but in the context of simulation it is important to choose $\theta^*$ from a high-density region of the support. In this study $\theta^*$ was chosen as the posterior mean of $\beta$, $D$, $\sigma^2$, and $\{\lambda\}$.
The functional form of the likelihood function and priors is known, and their ordinates are easily evaluated at the posterior means. Computing the ordinate for the joint posterior distribution is more difficult because the normalizing constant is unknown. Chib, however, showed that it can still be estimated by averaging over draws from the MCMC sampler [20]. Using laws of conditional probability, the posterior ordinate can be re-written as

\[
\hat{p}(\beta^*, D^*, \sigma^{2*}, \lambda^* | y) = \pi(\beta^* | y) \times \pi(D^* | \beta^*, y) \times \pi(\sigma^{2*} | \beta^*, D^*, y) \times \pi(\lambda^* | \beta^*, D^*, \sigma^{2*}, y)
\]  

(8)

and can be approximated by estimating each of the terms on the right hand side separately using draws from the MCMC sampler. The first term can be approximated by

\[
\hat{p}(\beta^* | y) = \frac{1}{G} \sum_{g=1}^{G} \pi(\beta^* | D^{(g)}, \sigma^{2(g)}, \lambda^{(g)}, b^{(g)}, y)
\]  

(9)

where \( \hat{p} \) is the full conditional distribution of \( \beta \) and \( g \) is a draw from the initial run of the MCMC sampler. The second term in equation (8) is approximated by

\[
\hat{p}(D^* | \beta^*, y) = \frac{1}{G} \sum_{g=1}^{G} \pi(D^* | \beta^*, \sigma^{2(g)}, \lambda^{(g)}, b^{(g)}, y)
\]  

(10)

where \( g \) now represents draws from a reduced run of the MCMC sampler where \( \beta \) has been fixed at \( \beta^* \). The third term is also approximated using a reduced run of the MCMC sampler by

\[
\hat{p}(\sigma^{2*} | \beta^*, D^*, y) = \frac{1}{G} \sum_{g=1}^{G} \pi(\sigma^{2*} | \beta^*, D^*, \lambda^{(g)}, b^{(g)}, y)
\]  

(11)

Because the normalizing constant of the full conditional distribution of \( \lambda \) is not known, the fourth term on the right hand side of equation (8) was approximated using kernel smoothing [21]. Ordinates were estimated as

\[
\hat{p}(\lambda^* | \beta^*, D^*, \sigma^{2*}, y) = \frac{1}{G} \sum_{g=1}^{G} \frac{1}{h_j} N\left( \lambda^*_i - \lambda^{(g)}_j \right)
\]  

(12)
where \( N \) is the Gaussian kernel, \( h_j \) is bandwidth, and \( g \) is a draw from a reduced run of the MCMC sampler with parameters set to \((\beta^*, D^*, \sigma^*, \lambda^*)\) [22-24]. Marginal likelihoods were computed to compare five different models: the unrestricted BCRC model, a restricted version of the BCRC model with \( \lambda_i = \lambda \), and three additional restricted models with \( \lambda_i \) set to 0, –0.5, and –1. Prior probabilities for the truth of these models were assumed equal.

**Model Fitting**

The sampler was run for an initial burn-in period of 1000 iterations, and an additional 5,000 draws were used for inference. Several different starting values were tested and confirmed that convergence of the Markov chain did not depend on the starting point. The model that was fit to the hospital data was:

\[
\text{COST}_{ij}^{(\lambda)} = b_{10} + b_{11} \text{ALBUMEN} + b_{12} \text{CHF} + b_{13} \text{PRIOR} + b_{14} \text{VALVE} \\
+ b_{15} \text{HEMATOCRIT} + b_{16} \text{AAO2} + b_{17} \text{PTCA} + b_{18} \text{CATH} + \varepsilon_{ij}.
\]  

(13)

Parameters for the priors were based on maximum likelihood estimates that ignored the hospital clusters. A grid search was used to estimate the model in equation (13), where the transformations were assumed constant across clusters. The likelihood function was maximized when the transformation was -.875, suggesting that a transformation close to the inverse of costs is more appropriate than a natural log transformation. Maximum likelihood estimates are reported in Table 1 and were used as hyperparameters for the priors of the BCRC model.

Priors were not chosen to reflect complete ignorance (i.e. they are not uniform distributions), because model selection is important to this study is not possible when ignorance priors are used. Marginal likelihoods are derived by integrating the likelihood over the priors, which requires finite mass in the priors. Based on the maximum likelihood results, the prior of \( \beta \) is a normal distribution with mean vector equal to the maximum likelihood estimates (.657, .085, .112, .118, .112, .083, .104, .219, -.804).
and covariance matrix with .5 on the diagonal and .25 on all off-diagonal elements. The prior for \( \{ \lambda \}_i \) is a normal distribution with mean -.875 and a variance of 3. The prior for \( \sigma^2 \) is an inverse gamma distribution with hyperparameters of .1 and .1, which imply a mean of 1 and variance of 10. The prior for \( D \) is an inverse Wishart distribution with 18 degrees of freedom (twice the number of independent variables) and a scale matrix with diagonal elements of 9 and off-diagonal elements of zero. All of the priors were centered at maximum likelihood estimates, yet had large enough variance to represent relatively vague information.
4 RESULTS

Hospital characteristics

Costs and patient characteristics for each of the four hospitals are summarized in Table 2. Note the substantial variation across the four hospitals. The lowest cost hospital incurred on average $15,616 in costs to treat a CABG patient, while the highest cost hospital incurred nearly twice that amount on average. The highest cost hospital also appears to have treated more severe cases and had the highest proportion of patients with worsening CHF, low hematocrit, and high AAO2. A histogram of costs, plotted in Figure 1, shows the typical problem with using untransformed levels of costs in a regression context: costs are highly skewed. The mode of costs is approximately $20,000, but there are several outliers with costs over $100,000.

Performance of the MCMC Sampler

The MCMC output from the model fitting was very well behaved. The first 1,000 draws of the transformation parameters, taken while the Markov chain was in its transient state, are plotted in Figure 2 and suggest that the sampler required fewer than 200 iterations to converge. Autocorrelations up to 30 lags for the 5,000 retained draws are plotted for the transformation parameters in Figure 3. While the sequence of draw for the transformation parameters is serially correlated, the correlation dies out rapidly and is only approximately .3 by lag 30. The slope coefficients and other model parameters are virtually free of serial correlation.

Posterior Analysis

The posterior distributions for all model parameters are summarized in Table 3. Because a Bayesian analysis produces probability distributions rather than point estimates, I report the mean, standard deviation and posterior probability that the coefficient is greater than zero.
The means of the marginal posterior distributions had the expected signs. Congestive heart failure (Pr(CHF>0)=0.74), prior procedures (Pr(PRIOR>0)=0.77) and associated valve procedures (Pr(VALVE>0)=0.75) all resulted in higher costs with high probability. Low hematocrit (Pr(HEMATOCRIT>0)=0.77) and high AAO2 (Pr(AAO2>0)=0.57) were also associated with higher costs. Cardiac catheterization (Pr(CATH)>0)=0.72) and PTCA (Pr(PTCA>0)=0.84) contributed to higher costs, and higher serum albumin (Pr( ALBUMIN>0)=0.30) resulted in lower costs. Note by comparing Table 3 to Table 1 the similarity between results from the BCRC model and the maximum likelihood estimates. The diagonal elements of the covariance matrix $D$, which indicate how much variation there is in the effect of the risk factors on costs, are also reported in Table 3. The fact that the variance is large relative to the posterior mean—approximately .09 on average—suggests substantial variation across hospitals.

Finally, note that in Table 3 the posterior distributions of the transformation parameters are not centered at 0 or 1, but range from a low of -1.13 to a high of -0.65. There is also a very small probability that any of the transformation parameters is greater than zero. Plots of the posterior distributions of the transformation parameters, presented in Figure 4, also indicate that the transformations are not equal across hospitals. While the distributions for the transformation for Hospitals B and D are very similar, the distributions for the other hospitals are not. For example, the probability that $\lambda_D$ is greater than the posterior mean of $\lambda_B$ is .48 and the probability that $\lambda_B$ is less than the mean of $\lambda_D$ is .698. But the probability that $\lambda_C$ is greater than the posterior mean of $\lambda_B$ is .98 and the probability that $\lambda_B$ is less than the mean of $\lambda_C$ is .97.

Model Selection

Additional evidence that the transformations are not equal across hospitals was found by model selection. Bayes factors were computed for a range of models and are reported in Table 4. Reading
across the first row shows that the data support the flexible BCRC model over all of the other models. The BCRC model was 2,099 times more likely to have generated the data than restricted BCRC model and 73,200 times more likely to have generated the data than the inverse transform model. Not surprisingly, reading in row 2 and column 5 of Table 4 we see that the inverse transform and the restricted BCRC model are virtually indistinguishable from each other. Still, the evidence is strongest for the flexible BCRC model. Furthermore, the Bayes factor is similar to the Schwartz criterion in that it penalizes higher dimensional models. The fact that the Bayes factor supports the BCRC model in spite of its additional parameters is strong evidence that the transformation differs across hospitals.

**Hospital Ranking**

*Ranking from Hierarchical Coefficients*

Since risk adjustment is often used to compare performance, I used the BCRC model to produce a rank-ordering of the hospitals. Most studies that use regression to estimate the parameters of a risk adjustment model rank the hospitals by the residuals [25,26]. Predicted costs are subtracted from observed costs, and the difference is averaged across hospitals to represent the hospital’s overuse or underuse of resources.

In Table 5 I report a comparable ranking based on the method of Goldstein and Spiegelhalter. Each hospital’s rank was computed at each iteration of the MCMC sampler, and then a mean rank was computed for each hospital [25,27]. I also report the unadjusted ranking and the rankings produced by maximum likelihood estimation of the linear and semi-log models in Table 5. The risk adjustment clearly affected the hospital ranking. Risk adjustment moved Hospital B from the position of second highest cost to lowest cost. Hospital A remained in the position of highest cost, and Hospitals C and D retained their positions relative to each other. The BCRC model produced exactly the same ranking as that produced by maximum likelihood estimation of the linear and semi-log models. This is not
entirely surprising since there are only four hospitals in the data and the means of the marginal posterior distributions were so similar to the maximum likelihood estimates. The four clusters in the data were also fairly evenly distributed across costs. We would expect to see differences in rankings where there are small margins of difference between hospitals.

*Ranking From Cluster-level Slope Coefficients*

The previous ranking, and in fact all ranking methods based on models that ignore clusters, does not account for the fact that the impact of a particular risk factor on cost may be different across hospitals, what we might call heterogeneous risk effects. For example, academic medical centers may be better able to handle patients with congestive heart failure than smaller community medical centers. Therefore we may expect the impact of CHF to be smaller at university hospitals. The fact that the diagonal elements of the covariance matrix $D$ were large relative to the size of the coefficients suggests substantial variation in slope parameters across clusters in these data.

The BCRC model can be used to produce rankings that take into account heterogeneous risk effects by ranking by predicted costs based on the hospital-specific slope parameters, $\{b\}_i$, which are simulated by the MCMC sampler as latent variables. To rank the hospitals by expected cost I estimated the average predicted cost for each hospital as

$$
\tilde{c}_i = \frac{1}{G} \sum_{g=1}^{G} (\lambda^{(g)}_i \bar{x}' b^{(g)}_i + 1) \bar{x}^{(g)}
$$

(14)

where $g$ indexes draws from the MCMC sampler and $\bar{x}$ is a vector of covariates that have been fixed at one of five levels of risk: extreme low, low, average, high and extreme high. Extreme low risk is zero for binary covariates and the sample minimum for continuous covariates. For low risk, all covariates were set to one standard deviation below the mean of the sample. Average risk is the mean.
of the sample. High risk is one standard deviation above the mean, and extreme high risk is 1 for binary covariates and the sample maximum for continuous covariates.

One obstacle was encountered in computing predicted costs using this method. The formula in equation (14) does not restrict costs to be greater than zero over the entire space of the risk factors. Because $\lambda_i$ is negative with high probability, for large enough values of $x$, the quantity $\lambda_i^{(g)} x^b_i^{(g)}$ is less than one and the solution to equation (14) requires raising a negative number to a non-integer power. This was only a problem in creating rankings for high and extreme high risk and only the iterations of the MCMC sample that yielded real predicted costs were used to create the ranking.

The ranking based on heterogeneous risk effects is presented in Table 6. Notice that the hospitals are ranked exactly as before for extreme low, low and average risk, but new rankings are obtained for high and extreme high risk. The fact that Hospital B moved from lowest cost to highest cost suggests that while Hospital B provided care for low and medium risk at a lower cost, it costs much more to care for higher risk patients. This might be expected for a small community medical center. In fact, Hospital B is the smallest community medical center in the data set. The ranking also shows that Hospital A is the highest cost hospital for all levels of risk. This might be expected for a large academic medical center, which is consistent with Hospital A.
This paper develops a Box-Cox model with random coefficients to test for the most appropriate functional form for risk adjusting the cost of treating patients undergoing CABG surgery. The model is novel in that there are very few Box-Cox models appropriate for clustered data applications [28]. Furthermore, while its potential in the context of risk adjustment has been recognized, the Box-Cox transformation has not been used previously [29]. Results suggest that the natural log transformation that is often applied to cost in order to shrink outliers may not be adequate. Posterior means of the transformations were close to –1, with very low probability in the neighborhood of zero. The results also suggest that the transformation should not be considered equal across hospitals. Not only were the posterior means for the transformation parameters varied across hospitals, Bayes factors overwhelmingly supported the most flexible BCRC model over more restricted versions.

In spite of providing a more appropriate functional form, the BCRC model produced exactly the same ranking as the linear and semi-log models fit using maximum likelihood methods. Most likely this was due to the fact there were only four hospitals in the data, but this result makes it more difficult to argue in favor of more sophisticated modeling over standard models. However, the BCRC model also provide a richer view of the ranking by allowing rankings to be produced that take into consideration the potentially heterogeneous effect of risk covariates on costs. Rankings were produced for hypothetical patients with varying degrees of risk. The hospital that ranked as lowest cost after risk adjustment remained the lowest cost hospital for hypothetical low and average risk patient, but was ranked second highest cost for a hypothetical high-risk patient. Future studies should include greater numbers of clusters and such other risk adjustment settings as physician profiling. Without additional research it is difficult to say whether general statements can be made about the most appropriate functional form for risk adjustment models of continuous outcomes.
ACKNOWLEDGEMENTS

I would like to thank Edward Greenberg, Siddhartha Chib, Robert S. Woodward, W. Claiborne Dunagan, Lawrence C. Kleinman and participants of the Penn State Health Evaluation Sciences Research Colloquium for valuable insights and suggestions.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>0.657</td>
<td>0.074</td>
<td>0.0001</td>
</tr>
<tr>
<td>CHF</td>
<td>0.085</td>
<td>0.025</td>
<td>0.0008</td>
</tr>
<tr>
<td>PRIOR</td>
<td>0.112</td>
<td>0.027</td>
<td>0.0001</td>
</tr>
<tr>
<td>VALVE</td>
<td>0.118</td>
<td>0.027</td>
<td>0.0001</td>
</tr>
<tr>
<td>HEMATOCRIT</td>
<td>0.112</td>
<td>0.023</td>
<td>0.0001</td>
</tr>
<tr>
<td>AAO2</td>
<td>0.083</td>
<td>0.024</td>
<td>0.0007</td>
</tr>
<tr>
<td>CATH</td>
<td>0.104</td>
<td>0.016</td>
<td>0.0001</td>
</tr>
<tr>
<td>PTCA</td>
<td>0.219</td>
<td>0.054</td>
<td>0.0001</td>
</tr>
<tr>
<td>ALBUMIN</td>
<td>-0.084</td>
<td>0.018</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

$R^2 = 0.32$

Table 1: Maximum likelihood estimates of a cross-sectional BC model. These estimates were used as prior values for hyperparameters.
<table>
<thead>
<tr>
<th>Variables</th>
<th>All (n=496)</th>
<th>A (n=107)</th>
<th>B (n=167)</th>
<th>C (n=58)</th>
<th>D (n=164)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COST</td>
<td>$20,338</td>
<td>$28,540</td>
<td>$20,811</td>
<td>$17,197</td>
<td>$15,616</td>
</tr>
<tr>
<td>ALBUMIN</td>
<td>3.92</td>
<td>3.93</td>
<td>3.74</td>
<td>3.99</td>
<td>4.07</td>
</tr>
<tr>
<td>CHF</td>
<td>11%</td>
<td>19%</td>
<td>10%</td>
<td>19%</td>
<td>5%</td>
</tr>
<tr>
<td>PRIOR</td>
<td>8%</td>
<td>7%</td>
<td>9%</td>
<td>12%</td>
<td>7%</td>
</tr>
<tr>
<td>VALVE</td>
<td>9%</td>
<td>11%</td>
<td>9%</td>
<td>22%</td>
<td>3%</td>
</tr>
<tr>
<td>HEMATOCRIT</td>
<td>13%</td>
<td>22%</td>
<td>10%</td>
<td>10%</td>
<td>12%</td>
</tr>
<tr>
<td>AAO2</td>
<td>11%</td>
<td>32%</td>
<td>3%</td>
<td>16%</td>
<td>4%</td>
</tr>
<tr>
<td>PTCA</td>
<td>2%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>CATH</td>
<td>61%</td>
<td>64%</td>
<td>77%</td>
<td>45%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Table 2: Characteristics of CABG patients by hospital.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Pr(β_k&gt;0)</th>
<th>D_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>0.5956</td>
<td>0.1833</td>
<td>0.9980</td>
<td>.0969</td>
</tr>
<tr>
<td>CHF</td>
<td>0.0911</td>
<td>0.151</td>
<td>0.7356</td>
<td>.0915</td>
</tr>
<tr>
<td>PRIOR</td>
<td>0.1031</td>
<td>0.1496</td>
<td>0.7600</td>
<td>.0903</td>
</tr>
<tr>
<td>VALVE</td>
<td>0.0979</td>
<td>0.1534</td>
<td>0.7460</td>
<td>.0930</td>
</tr>
<tr>
<td>HEMATOCRIT</td>
<td>0.1058</td>
<td>0.1521</td>
<td>0.7736</td>
<td>.0926</td>
</tr>
<tr>
<td>AAO2</td>
<td>0.0212</td>
<td>0.1515</td>
<td>0.5564</td>
<td>.0917</td>
</tr>
<tr>
<td>CATH</td>
<td>0.0827</td>
<td>0.1522</td>
<td>0.7178</td>
<td>.0921</td>
</tr>
<tr>
<td>PTCA</td>
<td>0.1707</td>
<td>0.1814</td>
<td>0.8434</td>
<td>.1009</td>
</tr>
<tr>
<td>ALBUMIN</td>
<td>-0.0773</td>
<td>0.1543</td>
<td>0.2958</td>
<td>.0913</td>
</tr>
</tbody>
</table>

\(\sigma^2\) 

| λ_1 | -0.9198 | 0.183 | 0.0000 |
| λ_2 | -1.1395 | 0.3236 | 0.0002 |
| λ_3 | -0.6517 | 0.2011 | 0.0020 |
| λ_4 | -1.0472 | 0.352 | 0.0060 |

Table 3: Summary of marginal posterior distributions of BCRC model based on 5,000 iterations of the MCMC sampler.
<table>
<thead>
<tr>
<th>Model</th>
<th>BCRC (-837.88)*</th>
<th>$\lambda_i = \lambda$ (-845.53)</th>
<th>$\lambda_i = 0$ (-916.64)</th>
<th>$\lambda_i = -0.5$ (-860.29)</th>
<th>$\lambda_i = -1$ (-849.08)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCRC</td>
<td>1</td>
<td>2099</td>
<td>$1.60 \times 10^{34}$</td>
<td>$5.38 \times 10^{9}$</td>
<td>$7.32 \times 10^{4}$</td>
</tr>
<tr>
<td>$\lambda_i = \lambda$</td>
<td>$4.76 \times 10^{4}$</td>
<td>1</td>
<td>$7.61 \times 10^{30}$</td>
<td>$2.56 \times 10^{6}$</td>
<td>34.85</td>
</tr>
<tr>
<td>$\lambda_i = 0$</td>
<td>$6.26 \times 10^{-35}$</td>
<td>$1.31 \times 10^{31}$</td>
<td>1</td>
<td>$3.37 \times 10^{-25}$</td>
<td>$4.58 \times 10^{-30}$</td>
</tr>
<tr>
<td>$\lambda_i = -0.5$</td>
<td>$1.86 \times 10^{-10}$</td>
<td>$3.90 \times 10^{-7}$</td>
<td>$2.97 \times 10^{24}$</td>
<td>1</td>
<td>$1.36 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\lambda_i = -1$</td>
<td>$1.37 \times 10^{-5}$</td>
<td>0.0287</td>
<td>$2.18 \times 10^{29}$</td>
<td>$7.35 \times 10^{4}$</td>
<td>1</td>
</tr>
</tbody>
</table>

*Log of marginal likelihood is in parentheses.

Table 4: Posterior odds for alternative models. Model is column 1 is the numerator, model in row 1 is denominator.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>No Adjustment</th>
<th>BCRC Model(^\dagger)</th>
<th>Linear Model(^\ddagger)</th>
<th>Semi-log Model(^\ddagger)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>(28,540)</td>
<td>(1.126)</td>
<td>(0.5767)</td>
<td>(0.2254)</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>(20,811)</td>
<td>(2.012)</td>
<td>(0.0339)</td>
<td>(0.0419)</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>(17,197)</td>
<td>(2.901)</td>
<td>(-0.2908)</td>
<td>(-0.1348)</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(15,616)</td>
<td>(3.960)</td>
<td>(-0.3393)</td>
<td>(-0.1554)</td>
</tr>
</tbody>
</table>

* Average unadjusted cost is in parentheses.
† Average rank is in parentheses.
‡ Average residual is in parentheses.

Table 5: Hospital rankings based on hierarchical slope parameters
<table>
<thead>
<tr>
<th>Risk</th>
<th>Hospital A</th>
<th>Hospital B</th>
<th>Hospital C</th>
<th>Hospital D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extreme Low Risk</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(1.15)</td>
<td>(3.99)</td>
<td>(1.84)</td>
<td>(3.00)</td>
</tr>
<tr>
<td>Low Risk</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(1.03)</td>
<td>(3.99)</td>
<td>(1.96)</td>
<td>(3.00)</td>
</tr>
<tr>
<td>Average Risk</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(3.53)</td>
<td>(2.00)</td>
<td>(3.46)</td>
</tr>
<tr>
<td>High Risk</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(1.26)</td>
<td>(1.90)</td>
<td>(2.92)</td>
<td>(3.91)</td>
</tr>
<tr>
<td>Extreme High Risk</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(1.48)</td>
<td>(1.66)</td>
<td>(3.39)</td>
<td>(3.46)</td>
</tr>
</tbody>
</table>

Average rank is in parentheses

Table 6: Predicted cost and average rank for a hypothetical patient assuming five levels of risk
Figure 1: Histogram of hospital costs
Figure 2: Convergence of AAO2
Figure 3: Autocorrelations up to lag 30 for transformation parameters
Figure 4: Marginal posterior distributions of transformation parameters
REFERENCES


