

ESTIMATING HOSPITAL QUALITY BASED ON PERFORMANCE MEASURE AND HOSPITAL ACQUIRED INFECTION DATA: A BAYESIAN BIVARIATE APPROACH

DAVID J. MORTON

The Joint Commission, Oakbrook Terrace, IL, USA

RAWAN ALLOZI

University of Illinois at Chicago, Chicago, IL, USA

DULAL K. BHAUMIK*

*Biostatistics Research Center, University of Illinois at Chicago
MC912 Room 513, 1601 West Taylor Street, Chicago, Illinois 60612, USA
Email: dbhaumik@uic.edu*

SUMMARY

Hospital quality data consists of dichotomous process measurements and count outcome measures. However, current statistical approaches to evaluating hospital quality often utilize one of these measurements. Moreover, when data are collected from multiple units within a hospital, the within-hospital correlation is typically ignored. To address this, we introduce a novel methodology for determining hospital quality that incorporates both outcomes within a Bayesian framework. We illustrate our approach using performance measures obtained from the Joint Commission and hospital acquired infection measures directly from hospitals. Using our approach, we differentiate high and low performing hospitals with a high degree of certainty.

Keywords and phrases: Hospital Quality, Multivariate Outcomes, Credible Interval, Bayesian Analysis.

1 Introduction

Current practices of evaluating hospital quality are one dimensional and primarily use univariate methods. Data types analyzed are restricted to either dichotomous process measures or count outcome measures. In this paper, we introduce a new methodology to determine hospital quality that incorporates both types of data. To do this, we utilize a bivariate distribution within a Bayesian framework. Several authors have addressed this problem; however, existing literature fails to provide adequate solution in using multiple types of data to evaluate hospital quality. Austin, Lee, and Leckie (2020) profiles healthcare providers with binary indicators by comparing a latent variable item response theory model to a multivariate Bayesian random effects logistic regression model. Dunson (2000) proposes a general framework for clustered mixed outcomes. Landrum,

* Corresponding author

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

Bronskill, and Normand (2000) and Normand, Glickman, and Gatsonis (1997) present a hierarchical logistic regression model based on risk-adjusted outcome measures. Teixeira-Pinto and Normand (2008) propose a fully Bayesian latent variable model based on the approach in Landrum, Bronskill, and Normand (2000). This approach analyzes performance measure rates using yearly data and determines hospital quality of care based on multiple related measures within a therapeutic area. In O'Brien et al. (2007), the proposed approach for profiling uses composite measures within a specific disease or therapeutic area. Furthermore, Shwartz et al. (2008) compares profiling utilizing various latent variable models based on denominator weights. Sosunov et al. (2016) explores the impact of hospital size of the Center for Medicare and Medicaid Services' hospital profiling using risk-standardized mortality and readmission measures. More recent published methods of evaluating hospital care are concentrated in specific areas of care, such as acute coronary syndrome mortality (Manda et al. 2012).

Organizations such as the Center for Medicare and Medicaid Services (CMS), U.S. News and World Reports, the Leapfrog group, Consumer Reports, Healthgrades, and Becker's Hospital Review utilize performance measure data to evaluate hospitals and post online 'report cards'. As discussed, statistical techniques of evaluating hospitals by these organizations are limited to simple univariate models or rankings. Additionally, each group applies their analyses using different measures. U.S. News and World Reports creates an overall hospital score utilizing external and claims-based data across 16 specialty areas (Olmsted et al. 2017). In contrast, Leapfrog creates a standardized score based on 18 measures across four domains: Medication Safety, Inpatient Care Management, Infections and Injuries, and Maternity Care (Leapfrog Group 2018). The CMS creates five star ratings based on 57 hospital measures among 7 clinical groups: Mortality, Safety of Care, Readmission, Patient Experience, Effectiveness care, Timeliness of Care, and Efficient use of Medical Imaging (Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation 2017).

Although there are many different data sources available for analysis, current statistical methods in the literature do not adequately model these data. One unrealistic assumption is that different departments within a hospital are independent. Weighting rates of quality of care within different units by sample size nested in a hospital is a common method for creating composite scores to assess hospital quality. However, this method ignores the within department correlation of a healthcare organization. Similarly, CMS' five-star rating of healthcare organizations employs independent latent variable models for each component of care to create an overall composite, which ignores inner correlations among different areas of a hospital.

Another limitation is the types of data used in assessing hospital quality. Typically, process-based measures, such as the proportion of patients that received aspirin at arrival, are the primary data types used and overall composite rates are presented. Although processes such as achieving desired rates and providing proper therapy is critical to hospital quality, a better indicator would be based on outcomes data. When outcomes are the focus, the analysis is restricted to within a certain department or therapeutic area.

The method presented in this paper addresses both restrictions mentioned above and provides a more robust method of profiling overall quality of hospitals. Our approach combines two types

of data: process-based composite proportion data and hospital acquired infection (HAI) rate-based outcomes data. We present a Bayesian bivariate latent variable model that incorporates a binomial model for the process composite measures and a Poisson model for the HAI outcomes data. This model is a critical step forward in analyzing varying types of hospital data to create two underlying latent variables that account for inter-hospital correlations between two separate hospital initiatives: one to evaluate quality of hospital processes and another to evaluate quality of patient safety with respect to HAIs. The latent variable obtained within the Bayesian framework are combined to create a composite latent variable measure for each hospital. Only then can we rank the composite measures to determine high and low performing hospitals.

An additional advantage of our method is the ease in implementation. Since it is constructed within the Bayesian framework, updating the model as new data becomes available is a relatively simple task. This is particularly important in the ever-changing world of healthcare measurement, where new measures are developed and older measures are retired. Furthermore, the shift from chart-based measures to measures developed using electronic medical records is easily transitioned with this methodology.

The paper is organized as follows. Section 2 outlines the proposed methodology, including the bivariate model and estimating hospital quality. Section 3 includes the results of the methodology applied to hospital performance measure data and hospital acquired infections data. Section 4 compares the bivariate IRT model to a univariate IRT model. The final section presents conclusions and discussion.

2 Methods and Material

In this section, we present a multivariate model for estimating measures of hospital quality. We then combined these measures to create a univariate measure of performance for identifying low and high performing hospitals.

2.1 Multivariate Model

Hospital quality is estimated using a vector $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ik})$ following a generalized multivariate distribution based on k different aspects of quality of care for the i th hospital within the framework of a Poisson and binomial model. In this study, $k = 2$ and θ_i is assumed to be bivariate normal, where the first component is estimated based on hospital performance measure data and the second component is based on hospital safety data, i.e., HAIs.

Let y_{ij} and n_{ij} denote the numerator and denominator, respectively, for the j th measure in the i th hospital. Let $\theta_s^{(i)}$ be the latent variable representing the underlying quality measure of hospital safety. Similarly, for hospital performance measures, let $\theta_{pm}^{(i)}$ be the latent variable representing the underlying quality measure of hospital performance. Under this framework, hospital quality is represented by $\theta_i = (\theta_s^{(i)}, \theta_{pm}^{(i)})$ and exchangeable to the joint distribution. Additionally, $p(\theta_s^{(i)}, \theta_{pm}^{(i)})$ is invariant to permutations of the indices for the i th hospital. These assumptions indicate that θ_i is an independent sample from a prior distribution with unknown parameter φ . In general, if there are

k different measures of quality, the hospital specific distribution is:

$$p(\boldsymbol{\theta}_i|\varphi) = \prod_j^k p(\theta_{ij}|\varphi), \quad (2.1)$$

$$p(\boldsymbol{\theta}_i) = \int \left[\prod_i^k p(\theta_{ij}|\varphi) \right] p(\varphi) d\varphi, \quad (2.2)$$

while the joint posterior distribution is

$$p(\boldsymbol{\theta}_i, \varphi|y_{ij}) \propto p(\boldsymbol{\theta}_i, \varphi) p(y_{ij}|\boldsymbol{\theta}_i, \varphi). \quad (2.3)$$

Within this framework, a two-component bivariate regression model is assumed to estimate the bivariate estimates of hospital quality based on performance measures and hospital safety derived from HAIs. In general, the model will have the following form:

$$\begin{aligned} y_{ij} &\sim f(\mu_{ij}, n_{ij}), \\ g(\mu_{ij}) &= \alpha_j + \sigma_j \theta_i, \end{aligned} \quad (2.4)$$

where $\sigma_j > 0$ and $\boldsymbol{\theta}_i \sim \text{MVN}(0, \Sigma)$. α_j represents the baseline rate and σ_j represents the discrimination factor for the j th measure. The function $f(*)$ represents the assumed distribution with corresponding link function $g()$ which is restricted to the item response components. Based on previous sections, $f(*)$ is assumed to be a binomial distribution for the performance measure data and a Poisson model for the safety data.

For hospital safety data,

$$\begin{aligned} y_{ij} &\sim \text{Poisson}(\mu_{ij}, n_{ij}), \\ f(y_{ij}; \mu_{ij}) &= \frac{\exp(-\mu_{ij}) (\mu_{ij})^{y_{ij}}}{y_{ij}!}, \\ \log(\mu_{ij}) &= \alpha_j + \sigma_j \theta_i + \log(n_{ij}), \end{aligned} \quad (2.5)$$

where $\log(n_{ij})$ is the offset term. Moreover, for hospital performance measure data,

$$\begin{aligned} y_{ij} &\sim \text{Bin}(n_{ij}, p_{ij}), \\ f(y_{ij}; n_{ij}, p_{ij}) &= \binom{n_{ij}}{y_{ij}} p_{ij}^{y_{ij}} (1 - p_{ij})^{n_{ij} - y_{ij}}, \\ \text{logit}(p_{ij}) &= \alpha_j + \sigma_j \theta_i. \end{aligned} \quad (2.6)$$

Under this framework,

$$\theta_i = \begin{pmatrix} \theta_{i,pm} \\ \theta_{i,s} \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_i \right) \quad (2.7)$$

where Σ has a Wishart distribution with hyperparameters analogous to precision (Gelman 2006).

2.1.1 Markov Chain Monte Carlo

A full Bayesian model was employed in OpenBUGS to obtain hospital estimates using the same prior distributions on α_j and σ_j . Prior distributions of each α_j assume a relatively flat $N(0, 100)$ and a standard normal $N(0, 1)$ prior distribution on each θ_j . The prior distribution of σ_j is defined as with a half-normal distribution of σ_i , recommended by Gelman (2006), where small precision represents vague prior information and constrains the parameter to be positive. Since it is challenging to obtain closed-form expressions of the posterior distributions, Markov Chain Monte Carlo is often used in OpenBUGS to sample from the posterior distribution. Detailed descriptions and applications of the methods can be found in Gilks, Richardson, and Spiegelhalter (1996) and Ntzoufras (2013). MCMCs are a class of widely used methods that have been shown to be able to efficiently generate posterior samples. We utilize the Gibbs Sampler, which is a case of the Metropolis Hastings algorithm.

2.1.2 Metropolis Hastings

Metropolis first applied methods based on Markov chain simulations in physics (Metropolis et al. 1953). In 1970, Hastings generalized the original algorithm and developed the Metropolis-Hastings algorithm (Hastings 1970). This algorithm has served as the basis for all MCMC methods. The idea behind the algorithm is that one generates candidates from a proposal distribution and updates the sample with a probability determined by the densities of the target and proposal distributions. The theory states that, regardless of the proposal distribution selected, the Metropolis-Hastings algorithm will converge to its equilibrium distribution. Although in practice, the choice of the proposal distribution is important since poorly chosen proposals will significantly slow down convergence.

In the Bayesian framework, the posterior $f(\theta|y)$ is the target distribution. With a proposal distribution $q(\theta'|\theta)$, the Metropolis-Hastings algorithm for Bayesian inference can be summarized as the following steps:

1. Set initial values $\theta^{(0)}$.
2. For $t = 1, \dots, T$, repeat the following steps:
 - (a) Set $\theta^{(t)} = \theta^{(t-1)}$.
 - (b) Generate new candidate values for θ' from $q(\theta'|\theta)$.
 - (c) Calculate probability $\alpha = \min \left(1, \frac{f(\theta'|y)q(\theta|\theta')}{f(\theta|y)q(\theta'|\theta)} \right)$.
 - (d) Update $\theta^{(t)} = \theta$ with probability α .

2.1.3 The Gibbs Sampler

The Gibbs sampler uses the full conditional distribution as the proposal (Geman and Geman 1984). It is a special case of the single-component Metropolis-Hastings algorithm. The full conditional distribution $f(\theta_j | \boldsymbol{\theta}_{-j}, y)$ is the distribution of the j th component of $\boldsymbol{\theta}$ given current values of all other parameters and data. Such a proposal distribution results in acceptance probability of $\alpha = 1$, i.e., the chain moves accepting all iterations. Since at each step random values are generated from univariate distributions, and frequently these distributions have a known and simple form, the computation is straightforward and one can select methods from a wide variety of tools. The algorithm does become ineffective when the parameters are highly correlated or the parameter space is complicated. The algorithm can be summarized as follows:

1. Set initial values $\boldsymbol{\theta}^{(0)}$.
2. For $t = 1, \dots, T$, repeat the following steps:
 - (a) Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.
 - (b) For $j = 1, \dots, d$, update θ_j from $\theta_j \sim f(\theta_j | \boldsymbol{\theta}_{-j}, y)$.
 - (c) Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_j, \dots, \theta_d)$.

2.2 Estimating Hospital Quality

Identification of high-performing hospitals needs to incorporate the variability of the estimate, as hospitals that have lower volume or lower opportunities for providing the proper therapy are more likely to be classified as high performers based on chance. Taking into account this variability by using a predefined threshold by determining the probability that the estimate exceeds the 95th percentile as shown as follows: $P(S_i > \eta_{95}) > \gamma$, where η_{95} is the 95th percentile of S_i , the scoring measure, and γ is a predefined threshold. This guarantees that a hospital is classified as being a high-performing hospital when the true score is above η_{95} with a reasonable degree of certainty. Generally, γ is taken to be as high as 0.95. By design, this lowers the number of hospitals that are classified as high performers to be less than 5% of the total. Thus, to identify exactly 5% of the hospitals as being high performers, the lower percentile of the score is used as a cutoff. Determining η_{95} such that 5% of the hospitals have a probability exceeding γ is identical to finding the threshold in which 5% of the γ 100% credible intervals of hospital performance lie above it. We use the symmetry of the credible regions to determine the cutoff point. Note that the probability that θ_i lies above the lower bound of the $((2\gamma - 1) \times 100)\%$ credible regions is γ . For example, if $\gamma = 0.9$, then the probability that S_i exceeds the lower bound of its $((2 \cdot 0.9 - 1) \times 100)\% = 80\%$ credible intervals is 0.9 since the probability of being in the credible region is 0.8 and the probability of being above the upper bound of the credible interval is 0.1. Therefore, classification of hospitals as high performers is determined by identifying the 95th percentile of the lower bounds of the $((2\gamma - 1) \times 100)\%$ credible regions.

Classification of low-performing hospitals uses a similar approach, that is, $P(S_i < \eta_{01}) > \gamma$, where η is the 1st percentile of S_i and γ is a predefined threshold. The 1st percentile of the upper

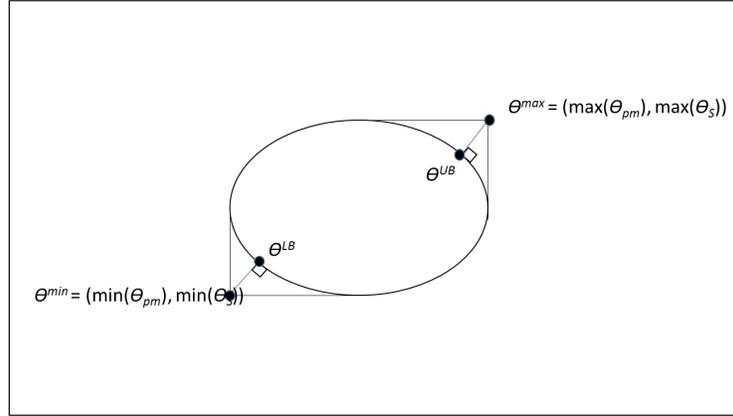


Figure 1: Identification of upper and lower bounds of θ_i .

bounds of the $((2\gamma - 1) \times 100)\%$ credible regions are calculated. Using the previous example, the probability that S_i is below the upper bound of its $((2 \cdot 0.9 - 1) \times 100)\% = 80\%$ credible regions is 0.9 since the probability of being in the credible interval is 0.8 and the probability of being below the lower bound of the credible interval is 0.1. Therefore, classifying hospitals as low performers is determined by identifying the 1st percentile of the upper bound of the $((2\gamma - 1) \times 100)\%$ credible regions.

For each hospital, $\theta = (\theta_{pm}, \theta_s)$ is estimated, where θ_s represents hospital safety quality based on the HAIs and θ_{pm} represents hospital quality based on the hospitals' ability to provide the appropriate treatment for performance measures. In general, θ can be a vector with k items depending on the number of valid estimates of hospital quality and can thus be extended to include additional hospital measures such as outcomes measures and physician measures.

For the bivariate model, the same methodology of identifying high-performing hospitals is adapted. In the bivariate case, the first step is to determine the credible region for each hospital by finding the points $\left\{ \hat{\theta}_i : \left(\hat{\theta}_i - \bar{\theta}_i \right)^T \Sigma^{-1} \left(\hat{\theta}_i - \bar{\theta}_i \right) \leq \chi_{1-\gamma}^2(k) \right\}$, where $\bar{\theta}_i$ is the centroid of the ellipse and $k = 2$ (Berger, 1980). For each hospital, define $\theta^{max} = (\max(\theta_{pm}), \max(\theta_s))$ and $\theta^{min} = (\min(\theta_{pm}), \min(\theta_s))$. The lower bound of the credible region is defined as the point $\hat{\theta}^{LB} = \operatorname{argmin} \left(\left\| \hat{\theta} - \theta^{min} \right\| \right)$, and the upper bound of the credible region is defined as $\hat{\theta}^{UB} = \operatorname{argmin} \left(\left\| \hat{\theta} - \theta^{max} \right\| \right)$, as illustrated in Figure 1. The scoring measures to determine high and low performing hospitals are a linear combination for each upper and lower bound is defined as $LB = \mathbf{w}^T \hat{\theta}^{LB}$ and $UB = \mathbf{w}^T \hat{\theta}^{UB}$, respectively, where \mathbf{w}^T is a row vector of weights that sum to 1. For this study, each estimate is weighted equally so the lower bound and upper bound are the mean of the individual terms of the lower and upper bounds. Classification then follows as previously mentioned.

2.3 Data Description

The Joint Commission is an independent, not-for-profit organization that accreditates over 22,000 hospitals and healthcare organizations and programs throughout the United States whose mission is to continuously improve healthcare. The Joint Commission collects performance measure information. A performance measure in this context consists of a numerator event, the number of times a patient population received a therapy, and a denominator event, the number of patients that were eligible to receive a therapy. Performance measure data were obtained from the Joint Commission's Quality Check website (<http://www.qualitycheck.org>), which contained yearly hospital level data in four clinical areas: Acute Myocardial Infarction (AMI), Heart Failure (HF), Pneumonia (PN) and Surgical Care Improvement Project (SCIP). AMI has seven measures: Aspirin at Arrival, Aspirin Prescribed at Discharge, angiotensin converting enzyme inhibitor (ACEI) or angiotensinogen receptor blocker (ARB) therapy for patients with left ventricular systolic dysfunction (LVSD)[ACEI or ARB for LVSD], Smoking Cessation Counseling, Beta Blocker Prescribed at Discharge, Fibrinolytic Therapy Received within 30 Minutes of Hospital Arrival, and Primary percutaneous coronary intervention (PCI) Received within 90 Minutes of Hospital Arrival. HF had four measures: Discharge Instructions, Evaluation of left ventricular systolic (LVS) Function, ACEI or ARB for LVSD, and Adult Smoking Cessation Advice/Counseling. Performance measures for pneumonia consists of six measures: Pneumococcal Vaccination, Blood Cultures Performed in the Emergency Department prior to Initial Antibiotic Received in Hospital, Adult Smoking Cessation Advice/Counseling, Initial Antibiotic Received within 6 Hours of Hospital Arrival, Initial Antibiotic Selection for community acquired pneumonia (CAP) in Immunocompetent—ICU, and Initial Antibiotic Selection for CAP in Immunocompetent— Non-ICU. Finally, there are six SCIP measures: Prophylactic Antibiotic Received within 1 Hour prior to Surgical Incision—Overall Rate, Prophylactic Antibiotic Selection for Surgical Patients—Overall Rate, Prophylactic Antibiotics Discontinued within 24 Hours after Surgery End Time—Overall Rate, Cardiac Surgery Patients with Controlled 6 a.m., Postoperative Blood Glucose, and Surgery Patients with Appropriate Hair Removal. For each hospital, a composite measure was created by summing up all the measures numerators and dividing by the sum of the measures denominators for each respective therapeutic area. All performance measures represent patient discharges from 3Q2012 to 2Q2013. Hospital Acquired Infection (HAI) measures are collected directly from hospitals through existing commercial infection control surveillance systems and electronic medical records. The HAI measures included catheter acquired urinary tract infection (CAUTI), surgical site infection (SSI): colon, SSI: hysterectomy, central-line associated blood stream infection (CLABSI), Methicillin-resistant Staphylococcus aureus (MRSA), and Clostridium difficile (C. Diff). All HAI measures represent infections reported between 7/1/2012 to 9/30/2013.

3 Results

The analysis data is composed of hospitals with reported safety and hospital performance without any constraints on the number of measures or sample sizes. There were 2,432 hospitals that reported at least one measure set for the performance measure data and at least one of the patient safety infection measures. The distribution of the data for each data grouping is presented in Table 1.

For the performance measures, each measure set had mean overall rates above 95%. SCIP had the highest overall rate with the lowest standard deviation. For the infection safety measures, MRSA and C. difficile had the lowest mean rates, and the SSI measures had the highest maximum rates of 0.33 each.

Table 1: Summary Statistics of Rates for Bivariate Data

Measures		n	Mean (SD)	Max Rate	90th Percentile	Median	10th Percentile	Min Rate
Performance	AMI	2,293	0.976 (0.0594)	1	1	0.9927	0.9412	0
	HF	2,385	0.965 (0.0652)	1	1	0.9906	0.9	0
	PN	2,374	0.968 (0.0416)	1	1	0.9773	0.9329	0.4118
	SCIP	2,337	0.984 (0.0207)	1	0.9975	0.989	0.9686	0.6667
Infection	CAUTI	2,327	0.0016 (0.0016)	0.0161	0.0038	0.00121	0	0
	SSI:Colon	2,330	0.0257 (0.0316)	0.3333	0.0612	0.01839	0	0
	SSI:Abdominal	2,256	0.0080 (0.0183)	0.3333	0.0227	0	0	0
	CLABSI	2,332	0.0010 (0.0043)	0.2	0.0022	0.0006	0	0
	C. difficile	2,408	0.0006 (0.0004)	0.0077	0.00108	0.0005	0	0
	MRSA	2,399	0.0001 (0.0001)	0.0010	0.00013	0.00002	0	0

The parameter estimates are based on two chains of 5,000 iterations allowing for a 5,000 burn-in chain. Posterior predictive checking was incorporated to determine the model fit utilizing two separate chains with different starting values in addition to incorporating Gelman-Rubin convergence statistic (Gelman and Rubin 1992). For each of the parameters for the performance measures converged much slower than the estimates of the α parameters for the safety measures. A similar pattern is also observed for the convergence of σ_i . For each estimate, the ratio for each of the two chains converges to 1 indicating agreement and convergence.

Parameter estimates with their 95% credible intervals are shown for each measure in Table 2. For the performance measures, AMI had the largest estimate of α along with the largest credible interval, while PN had the lowest α estimate. Among the safety measures, MRSA had the lowest parameter estimate of -9.797 for α while SSIs for colon surgeries had the highest α estimate of -3.71. With regard to the estimates of σ_i for the performance measures, AMI had the largest discrimination factor of 1.211 and HF had the second highest with a value of 1.107, whereas PN and SCIP had the lowest amount of discrimination with values of 0.637 and 0.863, respectively. For the safety measures, the CAUTI measure had the highest amount of discrimination of 0.728, while the amount of discrimination of the remaining safety measures ranged from 0.261 to 0.337.

Out of the 2,432 hospitals, 121 were identified as top-performing hospitals based on the method outlined using the performance measure data and the infection safety data jointly. The mean value of the estimates of θ_{pm} is 1.58, ranging from -0.39 to 2.79. The mean estimate of θ_s is 0.97 with a range of -0.69 to 3.15. The mean of the mean of the estimates of θ_{pm} and θ_s for the top-performing hospitals is 1.28, with a range of 0.806 to 2.24 and a standard deviation of 0.294. Mean rates on the four composite performance measures are 99.7% for AMI, 99.7% for HF, 98.7% for PN,

Table 2: Parameter Estimates and Credible Intervals for Bivariate Model (Full Bayes Estimate)

Measure	α	β
	(95% Credible Interval)	(95% Credible Interval)
AMI	4.851 (4.796, 4.907)	1.211 (1.161, 1.257)
HF	3.891 (3.834, 3.951)	1.107 (1.053, 1.161)
PN	3.660 (3.626, 3.695)	0.637 (0.604, 0.669)
SCIP	4.433 (4.397, 4.472)	0.863 (0.833, 0.891)
CAUTI	-6.621 (-6.665, -6.579)	0.728 (0.693, 0.767)
SSI:Colon	-3.710 (-3.744, -3.677)	0.261 (0.225, 0.298)
SSI:Abdominal	-4.914 (-4.971, -4.858)	0.267 (0.210, 0.325)
CLABSI	-7.068 (-7.104, -7.033)	0.337 (0.305, 0.370)
C. difficile	-7.445 (-7.466, -7.426)	0.319 (0.300, 0.338)
MRSA	-9.797 (-9.838, -9.757)	0.305 (0.266, 0.346)

and 99.7% for SCIP. For the safety infection measures, the observed number of infections for the high-performing hospitals is 0.59 (0.0047%) for SSI:Abdominal, 10.9 (0.003%) for C. difficile, 2.89 (0.0006%) for CAUTI, 1.94 (0.0006%) for CLABSI, 1.46 (0.0145%) for SSI:Colon, and 1.43 (0.00004%) for MRSA.

Using the methodology described, there are 24 hospitals identified in the bottom 1% of the joint distribution. The mean estimate of θ_{pm} for these low-performing hospitals is -1.29, ranging from -2.23 to -3.51, and the mean estimate for θ_s is -0.72, with a maximum value of -0.32 and a minimum value of -2.70. A contour plot of θ_{pm} and θ_s is shown in Figure 1. For the performance composite measures, the lower 1% of the hospitals identified had a mean AMI rate of 95.7% ranging from 66.7% to 100%, a mean HF rate of 90.0% ranging from 60% to 99.5%, a mean PN rate of 90.4% ranging from 60.0% to 100%, and a mean SCIP rate of 94.5% ranging from 74.2% to 98.8%. The observed number of infections and average infection rate for the safety infection measures is 3.45 (0.0147%) for SSI:Abdominal, 64.208(0.0009%) for C. difficile, 66.65 (0.005%) for CAUTI, 21.208 (0.002%) for CLABSI, 6.608 (0.048%) for SSI:Colon, and 8.28 (0.0001%) for MRSA. The density of the posterior estimates for θ_{pm} and θ_s is presented in Figure 2, which shows both posterior distributions are approximately normal with mean 0 and standard deviation of 1. Figure 2 plots the mean posterior estimates for each hospital for θ_{pm} and θ_s . The correlation between the two estimates is -0.05397 that is significantly different from 0 ($p=0.0078$). Figure 3 shows the contour plots for each of the estimates of θ_{pm} and θ_s and shows the highest density is centered around the origin. Figure 4 and Figure 5 show the estimates of θ_{pm} and θ_s compared to the observed rate for each measure along with the LOESS curve. Higher rates of performance measures are associated with higher estimates of θ_{pm} and lower number of infections are associated with higher values of θ_s .

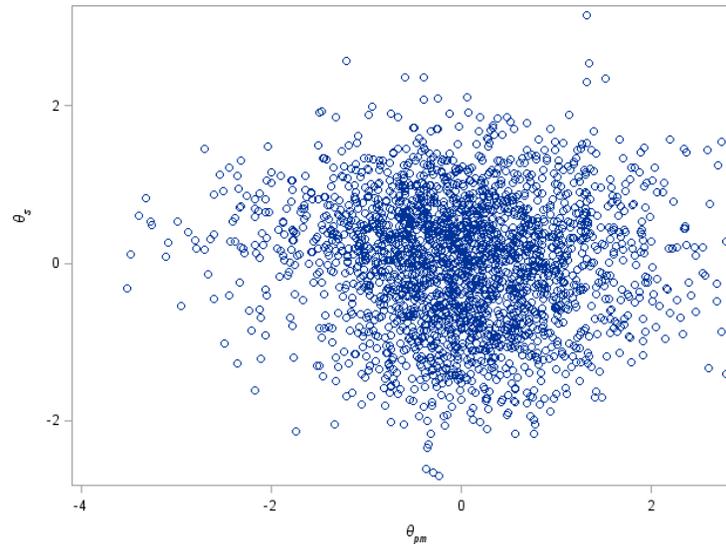


Figure 2: Plots of mean posterior estimates of θ_{pm} and θ_s for each hospital.

4 Comparison Analysis

As a comparison, we consider a single univariate outcome with the performance measure data and analyze the data without using the HAI data. We use the same methodology as Equation 2.6 except $\theta_i \sim N(0, 1)$. All assumptions and prior distribution specifications are identical. Model convergence is achieved after a burn-in of 5000 chains. Model estimates of for each performance measure α and σ are presented in Table 3 and are shown to be similar.

Of the 121 hospitals identified as high performers in the bivariate model, only 5 (4.1%) hospitals were classified in this univariate analysis matched the bivariate analysis. The other 116 (95.9%) hospitals that were identified as high performers in the univariate analysis had lower values of θ_{pm} and θ_s identified in the bivariate model with mean values of -0.02 and -0.05, respectively. Although an organization is a high performer in performance measures, as most organizations have rates close to 1, jointly modeling different areas of quality of care is important to assess overall hospital quality.

5 Discussion

Using a bivariate model to jointly model both the performance measure data and the patient safety infection data is a unique and appropriate way to model the data and determine high- and low-performing hospitals. Using this fully Bayesian method to model the performance measure data with a binomial model and the patient safety infection data with a Poisson model to determine the joint distribution of this scoring method is a robust approach and each of the observations is exchangeable. Based on credible regions, top performing hospitals were identified as having both

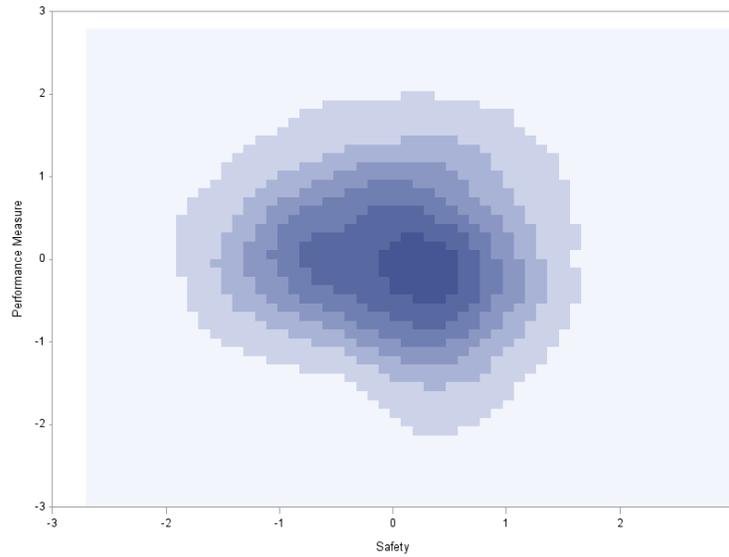


Figure 3: Contour plot of mean posterior estimates of θ_{pm} and θ_s .

Table 3: Parameter Estimates and Credible Intervals for Univariate Comparative Model (Full Bayes Estimate)

Measure	α	β
	(95% Credible Interval)	(95% Credible Interval)
AMI	4.746(4.699,4.796)	1.253 (1.214, 1.291)
HF	3.796(3.751,3.844)	1.089 (1.051, 1.128)
PN	3.605(3.573,3.638)	0.775 (0.750, 0.801)
SCIP	4.379(4.344,4.417)	0.970 (0.606, 0.660)

high measure rates and low number of infections. Identifying low-performing hospitals based on the bottom one percentile revealed organizations with low-performance measure rates and a high number of infections.

Model-based bivariate scoring methods offer more advantages than methods based on point estimates or models only using one dimension of care, as the former method borrows strength from other measure sets and dimensions. The proposed methodology accounts for hospital variation in order to make proper comparisons and categorizations of hospitals. As more data become available that represent increasing numbers of dimensions of hospital quality, the bivariate modeling approach presented in this study can be adapted to the multivariate case. Due to the exchangeability of the data, the fully Bayesian approach is preferred. This method presented accommodates the utilization of various distributions based on the data to determine an overall multivariate score which will enable a more accurate picture of hospital quality. Although we present a simple method based on

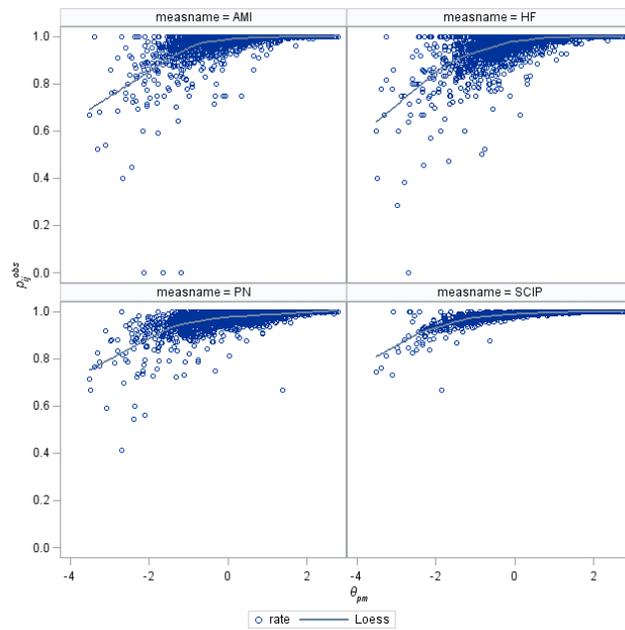


Figure 4: Plots of estimates of θ_{pm} versus observed performance measure rates.

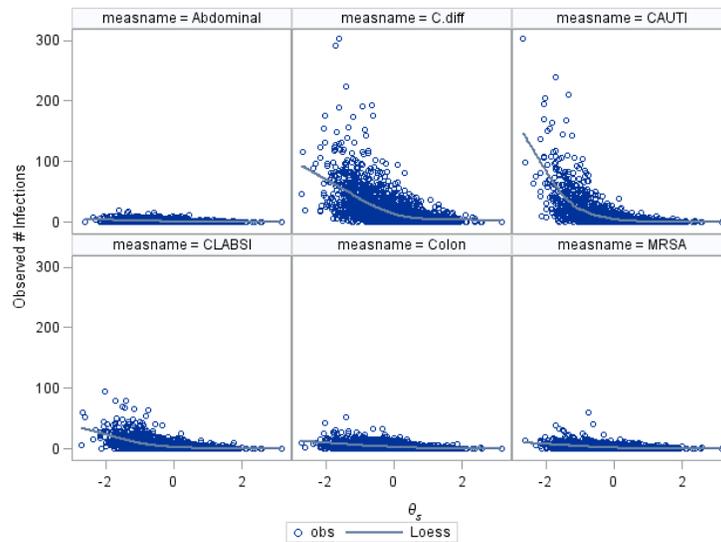


Figure 5: Plots of estimates of θ_s versus observed number of infections.

item response theory, the addition of covariates based on hospital demographics, such as bed size, volume, etc. would further help identify other relationships of quality of care.

References

- Austin, P. C., D. S. Lee, and G. Leckie. 2020. "Comparing a Multivariate Response Bayesian Random Effects Logistic Regression Model with a Latent Variable Item Response Theory Model for Provider Profiling on Multiple Binary Indicators Simultaneously." *Statistics in Medicine* 39:1390–1406.
- Dunson, D.B. 2000. "Bayesian Latent Variable Models for Clustered Mixed Outcomes." *J. R. Statist. Soc.* 62 (2): 355–366.
- Gelman, A. 2006. "Prior Distributions for Variance Parameters in Hierarchical Models." *Bayesian Analysis* 1 (3): 515–533.
- Gelman, A., and D.B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7:457–511.
- Geman, S., and D. Geman. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (6): 721–741.
- Gilks, W., S. Richardson, and D. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC.
- Hastings, W.K. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57 (1): 97–109.
- Landrum, M.B., S.E. Bronskill, and S-L. T. Normand. 2000. "Analytic Methods for Constructing Cross-Sectional Profiles of Health Care Providers." *Health Services & Outcomes Research Methodology* 1 (1): 23–47.
- Leapfrog Group. 2018. Leapfrog Value Based Purchasing Program 2017 Scoring Methodology. http://www.leapfroggroup.org/sites/default/files/Files/2017LVBPP_ScoringMethodology_020718.pdf. Accessed January 13, 2019.
- Manda, S.O., C.P. Gale, A.S. Hall, and M.S. Gilthorpe. 2012. "Statistical Profiling of Hospital Performance Using Acute Coronary Syndrome Mortality." *Cardiovascular Journal of Africa* 23 (10): 546–551.
- Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *J Chem Phys* 21 (6): 1087–1092.
- Normand, S-L. T., M. E. Glickman, and C. A. Gatsonis. 1997. "Statistical Methods for Profiling Providers of Medical Care: Issues and Applications." *Journal of the American Statistical Association* 92:803–814.
- Ntzoufras, I. 2013. "Bayesian Modeling using WinBugs." *Pharmaceutical Statistics* 12:188.

- O'Brien, S.M., E.R. DeLong, R.S. Dokholyan, F.H. Edwards, and E.D. Peterson. 2007. "Exploring the Behavior of Hospital Composite Performance Measures: An Example from Coronary Artery Bypass Surgery." *Circulation* 116 (25): 2969–2975.
- Olmsted, M.G., E. Geisen, J. Murphy, D. Bell, M. Morley, M. Stanley, and R. Powell. 2017. Methodology U.S. News & World Report 2017-18 Best Hospitals Specialty Rankings. http://static.usnews.com/documents/health/best-hospitals/BH_Methodology_2017-18.pdf. Accessed January 13, 2019.
- Shwartz, M., J. Ren, E.A. Pekoz, X. Wang, A.B. Cohen, and J.D. Restuccia. 2008. "Estimating a Composite Measure of Hospital Quality from the Hospital Compare Database: Differences When Using a Bayesian Hierarchical Latent Variable Model Versus Denominator-Based Weights." *Medical Care* 46:778-785.
- Sosunov, E.A., N.N. Egorova, H-M Lin, K. McCardle, V. Sharma, A.C. Gelijns, and A.J. Moskowitz. 2016. "The Impact of Hospital Size on CMS Hospital Profiling." *Medical Care* 54 (4): 373–379.
- Teixeira-Pinto, A., and S-L. T. Normand. 2008. "Statistical Methodology for Classifying Units on the Basis of Multiple-Related Measures." *Statistics in Medicine* 27:1329–1350.
- Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. 2017. Hospital Quality Star Rating on Hospital Compare Public Input Period: Enhancements of the Overall Hospital Quality Star Rating. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Hospital-Compare-PC-2017.pdf>. Accessed January 13, 2019.

Received: September 10, 2019

Accepted: August 1, 2020