Statistical Methods for Classifying Hospital Quality

Using Hierarchical Nonlinear Mixed-Effects Models

BY

DAVID JAMES MORTON
BS, Purdue University, 1989
MS, Purdue University, 1991
MS, University of Chicago, 1998

DISSERTATION

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2015

Chicago, Illinois

Defense Committee:

Dulal K. Bhaumik, PhD, Professor of Biostatistics and Psychiatry, UIC, Chair and Adviser
Robert Gibbons, PhD, Professor of Biostatistics, University of Chicago
Ronald C. Hershow, PhD, Associate Professor of Epidemiology, UIC
Sharon Lise-Normand, PhD, Professor of Biostatistics, Harvard Medical School
Stephen Schmaltz, PhD, Director of Biostatistics, The Joint Commission

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

CHAPTER PAGE

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

FIGURE PAGE

LIST OF FIGURES (continued)

FIGURE PAGE

LIST OF ABBREVIATIONS

AMI	Acute Myocardial Infarction
BUGS	Bayesian inference Using Gibbs Sampling
CAUTI	Catheter Associated Urinary Tract Infection
CCC	Concordance Correlation Coefficient
CDC	Centers for Disease Control and Prevention
CLABSI	Central Line Associated Bloodstream Infection
CMS	Centers for Medicare and Medicaid Services
DIC	Deviance Information Criteria
EB	Empirical Bayes
FB	Full Bayes
HAI	Hospital Acquired Infections
HF	Heart Failure
IRT	Item Response Theory
MCMC	Monte Carlo Markov Chain
MRSA	Methicillin-resistant *Staphylococcus aureus*
NHSN	National Healthcare Safety Network
NPSF	National Patient Safety Foundation
NPSG	National Patient Safety Goals

LIST OF ABBREVIATIONS (continued)

PMS	Performance Measurement Systems
PN	Pneumonia
SCIP	Surgical Care Improvement Project
SSI	Surgical Site Infection
UTI	Urinary Tract Infection
WHO	World Healthcare Organization

SUMMARY

Evaluating hospital performance is crucial for improving the quality of health care. Unfortunately, there is a lack of standardized methods for such evaluation. Most organizations evaluate hospitals at the individual measure level and the types of analysis range from ranking observed rates to more sophisticated hierarchical models. Individual measures or therapies are nested within therapeutic areas of care, called measure sets, such as acute myocardial infarction and heart failure. Moreover, evaluation of hospital quality is concentrated on these performance process measures and there has currently been no attempt to evaluate hospital quality of care based on hospital safety data.

This study presents three methods using nonlinear mixed-effects models to evaluate hospitals using composite measures based on hospital performance measures that are defined as proportion measures. A proportion measure is a measure where the numerator is a subset of the denominator and the ratio can be expressed as a rate or proportion as opposed to a continuous measure such as time from entering the emergency room until admission to the hospital. The first approach utilizes a composite performance measure based on yearly as well as longitudinal data and uses a hierarchical mixed-effects logistic regression model. The parameters are estimated by both Empirical and Full Bayes methods. Measure of agreement between the estimates of the parameters obtained by the Full Bayes and Empirical Bayes models uses the concordance correlation coefficient (CCC).

The second method incorporates a hierarchical mixed-effects Poisson regression model to each type of data, i.e., hospital performance data and hospital patient safety data. Estimates obtained from the model utilizing the hospital performance data will be compared. For each data type, estimates are obtained by using an Empirical Bayes method and a Full Bayesian model and are compared with each other.

The final method uses a bivariate mixed-effects hierarchical model that incorporates the correlation between hospital performance measure data and hospital safety data. An overall measure of quality is constructed based on the bivariate latent variable to form a quality of care measure.

In conclusion, each of these methods is a reasonable way to measure hospital quality of care by utilizing aggregated publically available data. The arena of hospital performance measures is constantly changing and some measures used in this study are no longer active measures that are being collected. There are also new measures being developed that represent other types of measures, such as outcomes measures, along with measures being developed in other areas of hospitals care. Although the analysis presented here represents a snapshot of the measures currently available, the methods developed in this study are flexible to incorporate new measures and measure sets as they become available.

1. INTRODUCTION

There are many types of data available and there is a constant need to consolidate data into a meaningful quantity that will help consumers to make educated decisions. The area of interest for this study is the area of quality of care in hospitals and healthcare organizations. With a plethora of data available on the internet provided by many government and private organizations, there is a need to consolidate data within a multitude of areas that will serve as a guide for consumers to help determine where to find the best hospitals, not only for certain medical conditions but overall. Additionally, new statistical methods incorporating and consolidating different data types will help hospitals to identify areas in need of quality improvement.

1.1 Performance Measures

Evaluation of hospital performance is an ongoing practice as hospitals strive to provide the best possible care for their patients in addition to meeting financial incentive programs. Effective implementation of evidence-based performance measures has become a critical success factor for healthcare providers to demonstrate meaningful improvements in clinical quality. Since the quality of hospital care is multidimensional and not directly measurable, multiple hospital performance measures are evaluated to ascertain quality of care. The quality of hospital care based on hospital performance measures has been focused

primarily on the individual therapy or process measure within a clinical area or measure set but not at the set level. Due to the lack of availability of more enriched data sets, current literature has been restricted to yearly "snapshots" of the data. Although access to quarterly performance data has been available since January 2007 from the Joint Commission, there is no analytical study evaluating hospital-level performance utilizing these quarterly data. Proper implementation of quarterly data will help evaluate the quality of hospital care not only at a static point in time, but also longitudinally.

Financial incentives are available to hospitals that participate in various quality improvement initiatives. Insurers have implemented incentive programs for hospitals to achieve better performance [Kuhn, 2011]. For example, in 2006, the Centers for Medicare and Medicaid Services (CMS) proposed a market basket update of 2.1% for those hospitals that participate in reporting hospital quality data for their Annual Payment Update program whereas those hospitals that do not participate receive 2.0% less [CMS, 2011]. The incentive by the Joint Commission in reporting these quality improvement measures is part of their accreditation requirement. Reporting the data is not enough, as the CMS has instituted a value-based purchase plan that links their payment plan to the quality of clinical care where providers are rewarded for their delivery of high-quality clinical care consistent with their vision of providing "the right care for every person every time" [CMS, 2012]. In 2010, the CMS has increased the incentive rate even more for better quality in the four clinical areas of acute myocardial infarction (AMI), pneumonia (PN), heart failure (HF), and the surgical care improvement project (SCIP).

Furthermore, the CMS has plans for instituting additional payment incentives for rewarding hospitals that show improvement over time.

The first national program that required hospitals to report nonstandardized performance measure was the Joint Commission's ORYX initiative that started in 1998. In 2002, hospitals were required to submit performance measure data to the Joint Commission as part of the ORYX initiative. In 2005, the Joint Commission and the CMS aligned together in defining, collecting and publishing these data. Currently, there are 57 performance measures collected by the Joint Commission with 31 of them being publicly reported. These publicly reported measures, which include the measures that define the AMI, HF, PN, and SCIP measure sets, have been endorsed by the National Quality Forum and adopted by the Hospital Quality Alliance and are used by various stakeholders to demonstrate quality of care to the public, purchasers, payers, and others [Casey, 2010; Williams et al., 2006].

Profiling hospitals is constantly under scrutiny as hospital report cards and ratings are presented to the public. The *US News and World Report* and *Consumer Reports* publish hospital rankings to define America's "top" hospitals. Healthgrades, an independent healthcare ratings organization, publishes "America's 50 Best Hospitals" in addition to presenting a distinguished hospital award for quality excellence. The CMS reports their findings on the Hospital Compare website (http://www.hospitalcompare.gov); whereas the Joint Commission's Quality Check reports their version of quality performance ratings (http://www.qualitycheck.org). Recently, the Joint Commission has defined a subset of their performance measures to be accountability measures that have the greatest impact on patient outcomes [Joint Commission, 2011]. With

these accountability measures, the Joint Commission instituted a hospital recognition program that rewards hospitals with an overall composite rate greater than 95%, including those accountability measures with fewer than 30 denominator cases. Additionally, hospitals must achieve a rate greater than 95% for each of the accountability measures for those measures that have at least 30 denominator cases [Joint Commission, 2011]. It is possible for a hospital to have a composite rate of greater than 95% but have no individual measures that have greater than 30 cases. A hospital may have a rate of 95% on a subset of accountability measures with more than 30 cases but have a composite rate below 95%. For example, suppose there are four accountability measures submitted by a hospital and that three of the measures have rates of 29/30. Suppose the last measure has a denominator of 29 with rate of 0% (0/29). As this denominator is less than 30, it is not included in the measure level criteria. The overall composite rate will be 87/119, which is 73%, therefore this hospital will not be identified as a top performer by the Joint Commission.

Hospitals are multifaceted in terms of patient care and services. Measures for a therapeutic area are used to assess quality of hospital care within that particular area. The composite measure based on multiple therapeutic areas is important to evaluate the overall performance of an individual hospital. Though the overall average composite score for each measure set weighted by the number of patients is a popular statistic used by insurers and pay-for-performance providers for financial gains, this composite score ignores the correlation between measures and measure sets. Ignoring these correlations causes incorrect inference due to the improper estimation of the variance structure. Furthermore, adjustments for covariates,

missing data, and small sample sizes cannot be addressed with raw composite scores. Normand et al. [1997] present a hierarchical logistic regression model based on risk-adjusted outcome measures. O'Brien et al., [2007] present profiling utilizing composite measures. Teixeira-Pinto and Normand [2008] proposed a fully Bayesian latent variable model based on Landrum et al. [2000] for analyzing the performance measure rates using yearly data and determining hospital quality of care based on multiple related measures. Chassin [2010] has shown that trends in these hospital performance measures have been steadily increasing, on the average, over time, although, no formal statistical techniques are employed.

1.2 **Safety Measures**

Patient safety is one of many elements that help to determine hospital quality of care in addition to performance measures, outcome measures, and patient satisfaction. Each point in the process of patient care, from admission to discharge, presents a degree of inherent unsafety. Patient safety is of high concern for many healthcare organizations, many of which focus primarily on issues of patient safety. Those include the Joint Commission, the CMS, the National Patient Safety Foundation (NPSF), the World Health Organization (WHO), and the Centers for Disease Control and Prevention (CDC), to name a few. Although it is widely known that patient safety is a fundamental principle of healthcare that is vital for the success of healthcare institutions, there are some differences in which each organization defines patient safety.

The CDC has developed the National Healthcare Safety Network (NHSN), which is devoted primarily to provide healthcare organizations the means of tracking hospital acquired infections (HAI) in order to identify areas of improvement. Although the NHSN started with only 300 hospitals decades ago, the number of healthcare organizations has increased to more than 11,000 medical facilities and is still expected to continue growth in the future. Within the NHSN, the CDC tracks five surveillance modules to help healthcare organizations monitor various aspects of patient safety. Events associated with devices, procedures, and antimicrobial agents are the primary focus in the following modules: Device-associated module, procedure-associated module, antimicrobial and resistance module, multidrug resistance organism and *Clostridium difficile* infection module, and vaccination module.

The CDC estimates that 41,000 central line-associated bloodstream infections (CLABSI) occur within US hospitals each year [CDC, 2011]. These preventable infections increase hospital stay and cost in addition to adding additional risk of mortality. Proper techniques for inserting the central line and proper management can prevent CLABSI. These techniques are addressed in the CDC's "Healthcare Infection Control Practices Advisory Committee (CDC/HIPAC) Guidelines for the Prevention of Intravascular Catheter-Related Infections, 2011" [O'Grady, 2011].

Urinary tract infections (UTIs) represent more than 15% of infections reported by acute care hospitals [Hellinger et al., 2011]. These types of infections are the second most common type of HAI, tied with PN. These infections are mostly caused by instrumentation of the urinary tract. Each year, more than 13,000 deaths are associated with UTIs [Klevins et al., 2012].

Complications stemming from catheter-associated urinary tract infections (CAUTI) are cystitis, pyelonephritis, gram-negative bacteremia, prostatitis, epididymitis, and orchitis in males, endocarditis, vertebral osteomyelitis, septic arthritis, endophthalmitis, and meningitis. These complications cause increased hospitalizations and costs in addition to mortality. Prevention of CAUTI is outlined in the CDC/HICPAC document "Guideline for Prevention of Catheter-associated Urinary Tract Infection" [Gould et al., 2010].

Approximately 16 million operative procedures were performed in the United States in 2010 as reported from the CDC in the "Data from the Hospital Discharge Survey." In 2012, Magill et al. report surgical site infections (SSI) to be the most common HAI accounting for 31% of all HAIs among hospitalized patients [Yi et al., 2011]. The NHSN data for 2006–2008 (16,147 SSIs following 849,659 operative procedures) showed an overall SSI rate of 1.9% [Awad, 2012]. While advances have been made in infection control practices, including improved operating room ventilation, sterilization methods, barriers, surgical technique, and availability of antimicrobial prophylaxis, SSIs remain a substantial cause of morbidity and an associated mortality rate of 3% has been attributed to them. Of this, 75% of the mortality rate has been directly related to the SSI [Kleven et al., 2007].

The WHO defines patient safety as the reduction of risk of unnecessary harm associated with healthcare to an acceptable minimum. The WHO defines areas of patient safety to be HAIs, medication errors, unsafe surgery, clinical handovers, and injection safety, and reports that infections acquired within hospitals worldwide affect approximately 1.4 million people at any given point of time. Additionally, the WHO is developing programs to address patient safety

issues that include the “High 5’s” project that identifies solutions to five patient safety issues in five countries for five years.

In 1997, the NPSF was founded to act on behalf of the advancement of patient safety worldwide. According to the NPSF, patient safety is “the prevention of healthcare errors, and the elimination and mitigation of patient injury caused by healthcare errors.” Healthcare errors cause an unintended outcome from a defect in care to the patient and these errors can be caused by any member of the healthcare team at any point of contact to the patient. They consider the following areas with regard to patient safety: wrong site surgery, medication errors, HAIs, falls, readmissions, and diagnostic errors.

In 2006, as part of the Affordable Care Act, the CMS published their final rule for implementation of Provider Preventable Condition, which includes hospital acquired conditions resulting from foreign objects retained after surgery, air embolism, blood incompatibility, stage III and IV pressure ulcers, falls and traumas, CAUTIs, SSIs, and deep vein thrombosis. Failure to follow this rule requires states to implement nonpayment policies to those organizations that do not establish Provider Preventable Conditions.

Additionally, the CMS is focusing their efforts on preventable readmissions in the areas of HF, heart attack, and PN. It is estimated that for every five Medicare patients admitted into the hospital, one patient is readmitted within 30 days of discharge for the same diagnosis. This represents approximately 2.6 million seniors at a cost of more than $26 billion dollars each year. In 2012, the CMS proposed a rule to reduce preventable readmissions by 20% as compared to the 2010 baseline.

The Joint Commission, a national not-for-profit accreditation organization, has established accreditation standards, through which healthcare organizations can measure, assess, and improve performance to focus on providing safe and high-quality patient care. Within these standards are the National Patient Safety Goals (NPSG), which evaluate hospital safety in the following areas: correct identification of patients; improvement of staff communication; safe use of medications; infection prevention; identification of patient safety risks; and prevention of surgical errors. Within the area of infection prevention the Joint Commission's standards are concerned with hand hygiene, difficult-to-treat infections, central line infections, infections after surgery, and UTIs.

Although each organization has slightly different definitions and areas of concentration that define patient safety, the most common element among the groups is HAIs. These are infections acquired by patients during the time of treatment or hospitalization. A patient may receive an infection in a variety of different settings ranging from hospital acute care or surgery units to outpatient clinics to long-term care facilities. Infections can be caused by, but are not limited to the use of medical devices such as ventilators and catheters, complications following surgical procedures, antibiotic overuse, or transmission between patients and healthcare providers.

Among the top ten leading causes of death in the United States, HAIs result in approximately 1.7 million infections and 99,0000 associated deaths in 2002 [Kleven et al., 2007]. They are a particularly significant cause of morbidity and mortality in hospitals with a nearly ten-fold increase in hospitals stays for methicillin-resistant *Staphylococcus aureus* (MRSA)

infections since 1995 [Elixhauser and Steiner, 2007]. Three-quarters of all HAIs are composed of four main areas: UTI, SSIs, Blood stream infections, and PN. Currently the CDC reports that those organizations that are collecting HAI data and monitoring their facilities progress are showing reductions in HAIs [Malpiedi et al., 2013]. The CMS estimates that one in every 20 patients will develop an infection as a result of their hospital care, resulting in $28 to $33 billion dollars annually of preventable healthcare expenditures. [CMS National Action Plan, 2009]

1.3 **Statement of the Problem**

Organizations such as the CMS, the Joint Commission, and Healthgrades have developed rating systems to evaluate how a hospital performs at a measure level, and only the Joint Commission has attempted to evaluate hospitals at a measure set or therapeutic area level. None of the organizations have developed a methodology that evaluates hospital quality across therapeutic areas, such as AMI, HF, PN, and SCIP. Determining which types of hospitals perform better or worse than other types of hospitals has not been determined across various therapeutic areas. Additionally, evaluation of hospital quality efforts has not been studied longitudinally. To date, there have been no publically reported papers or score cards statistically analyzing hospital performance measures over time to determine trends within a hospital to evaluate an individual hospital's quality improvement efforts.

Missing from the research are methods that evaluate hospital quality using hospital safety data. Statistical methods to analyze safety data have not yet been formalized. Furthermore, there

have been no attempts in combining the hospital safety and performance data to create an overall hospital index of quality. Currently, evaluation of hospital quality has been analyzed using univariate methods and not multivariate methods. This study explores new statistical techniques by which to evaluate hospitals using univariate and bivariate methods.

1.4 **Purpose of the Study**

The purpose of this study is to develop statistical methods to evaluate hospital quality using multiple measure sets for static yearly data based on composite measures of each therapeutic area or measure set. Furthermore, I will develop a robust statistical technique for evaluating hospital improvement efforts longitudinally utilizing quarterly data. Additionally, I will evaluate hospital quality using Bayesian methods that utilize hospital performance measures and hospital safety measures and I will develop an overall hospital quality score based on a joint model. This model will incorporate multivariate methods to evaluate hospital quality of care.

This study is organized in three parts. The first part assumes a hierarchical mixed-effects logistic regression model that identifies the top and bottom performing hospitals based on a latent variable using composite measures calculated from hospital performance measure data within a static year time frame. Additionally, I present a hierarchical mixed-effects longitudinal logistic regression model to identify those hospitals that show increasing or decreasing trends in quality of patient care over time. The parameters of each static and longitudinal regression model are estimated using a Full Bayes (FB) approach and an Empirical Bayes (EB) approach. Each

approach is compared using Linn's CCC to determine the agreement of the parameter estimates between methods.

The second part of the study uses two hierarchical mixed-effects Poisson regression models to evaluate hospital quality. The first model uses static yearly composite measures based on hospital performance measures. Parameter estimates of the latent variable from this model are compared to the estimates obtained from the hierarchical mixed-effects logistic regression model obtained in the first part of this study using Linn's CCC. The second model uses static yearly hospital safety data. Estimates of the latent variable of both hierarchical mixed-effects Poisson regression models are compared using Linn's CCC to determine if hospital quality can be evaluated by using either hospital performance measure data or hospital safety data.

For the final part of this study, a dual purpose hierarchical mixed-effects model will be used incorporating a bivariate latent variable that incorporates both composite hospital performance measure yearly data and yearly safety data. Bivariate methods on the latent variable will be used to evaluate top-performing hospitals and to evaluate the correlation between both types of data.

In this study, I focus on modeling these two aspects of patient safety and hospital performance measures and employ Bayesian techniques involving latent variables to study the underlying quality of patient safety and performance measures within an organization. Utilizing the bivariate model helps to determine an overall estimate of hospital performance based on both hospital performance measure data and hospital safety data and identifying the correlations between hospital performance measures and HAIs.

1.5 **Significance of the Problem**

For this critically important problem, patients or consumers of healthcare have a choice on which hospitals or healthcare organizations they can visit. Even if patients are limited in their choice of hospitals due to their healthcare insurance plan, they should have proper data to make an educated choice on which hospital is best for them based on many factors such as disease, therapy, or insurance plan. As it exists today, there are so many different types of data and statistics reported that it is very difficult for the consumer to evaluate which hospital is the best for them.

These hierarchical nonlinear mixed-effects statistical methods will have great impact on the hospital community. Deriving methods that are easy to implement will greatly help the consumer of hospital services to determine which organization is best for him or her. Having a single quantifiable number or rating will have direct impact on where patients choose to go for their hospital services.

This work will also have great impact on the hospitals themselves. Chassin et al., [2010] has shown that measuring hospital quality invokes hospital improvement efforts. Once deficiencies are discovered, hospitals and healthcare organizations will administer efforts in order to improve their performance and outcomes.

2. BAYES METHODS

Bayesian statistics have been developing quickly over the past two decades. Bayesian theory differs from the classical statistical theory, the frequentist approach, in that the former applies prior distributions on the random effects whereas the latter does not assume a prior distribution on the random effects. These prior distributions are combined with the data likelihood to obtain the posterior distribution of the parameters of interest. Statistical inferences are based on the posterior distributions of the parameters of interest.

The fundamental Bayesian approach is to specify a model to the observed data $\boldsymbol{y} = (y_1, y_2, .., y_n)$ given unknown vector of parameters $\boldsymbol{\theta}$ in the form of a probability distribution $f(\boldsymbol{y}|\boldsymbol{\theta})$, which is the data likelihood given the parameter vector. I assume that $\boldsymbol{\theta}$ is a quantity of interest with a prior distribution $p(\boldsymbol{\theta})$ and base all inferences concerning $\boldsymbol{\theta}$ on the posterior distribution denoted by

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{h(\boldsymbol{y})} \propto f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{2.1}$$

Equation 2.1 is known as Bayes' Theorem where $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior distribution of the parameter θ that differs by a constant, $h(\boldsymbol{y})$, from the combination of the likelihood and the prior distribution, $f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$. Thus, after observing the vector of data $\boldsymbol{y}$, one can employ this rationale to combine the data information with the prior distribution to obtain the posterior distribution in order to make Bayesian inferences.

The introduction of a prior distribution introduces a subjective point of view to the Bayesian approach. However, Bayesian methods failed to become widely used data analysis tools until recently due to difficulties involved in the computation of the posterior distribution. Prior to 1990, a generalized approach was lacking for all problems although asymptotic results were available for specific problems that had conjugate prior and posterior distributions.

Gelfand et al. [1990] implemented the Markov Chain Monte Carlo (MCMC) methods in the framework of Bayesian computation, that Bayesian methods regained the attention of the statistical community. Since then, Bayesian statistics have become fashionable and many related works appeared in the statistical literature. During the early 1990s, studies focused on implementing MCMC methods for different models including hierarchical and mixed-effects models [Gelman and Rubin, 1992; Gelfand et al., 1992; Gilks and Wild, 1992; Dellaportas and Smith, 1993]. Algorithms that allowed model averaging, selection, and exploration [Green, 1995; Dellaportas and Forster, 1999; Dellaportas et al., 2002; Sisson, 2005; Hans et al., 2007] continued to develop.

These advances are an example of the rapid development of Bayesian statistics. However, Bayesian analysis did not gain wide popularity until the appearance of standard Bayesian analysis software based on MCMC. The first versions of Bayesian inference Using Gibbs Sampling (BUGS) software appeared in the late 1990s. In BUGS, the user only needs to specify the structural of the statistical model. The software then generates samples from the posterior distribution of the specified model using one of the MCMC methods, the Gibbs sampler. Since its first appearance, the BUGS software has become a popular tool in Bayesian methods and

demonstrated its value in the implementation of Bayesian models in numerous scientific fields. WinBUGS [Lunn et al., 2000] is a stand-alone software package developed by the BUGS project team (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml). It features a graphical user interface and on-line monitoring and convergence diagnostics and can be called from other statistical software (such as from R using R2WinBUGS). The current version, WinBUGS 1.4.3, is available via the WinBUGS project Webpage (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml). Currently, there are no plans to further develop WinBUGS, but there is another version called OpenBUGS that is an open-source version of the WinBUGS package where all future work on the BUGS project will be developed. OpenBUGS can be accessed at http://www.openbugs.net.

2.1 **Markov Chain Monte Carlo Method**

This section provides a brief introduction to the stochastic simulation methods for posterior sampling that are widely used in Bayesian analysis. The focus is on MCMC and, especially, the Gibbs sampler, which is the algorithm behind the computations in OpenBUGS. Detailed descriptions and applications of the methods can be found in Ntzoufras, [2009], Gilks et al. [1996], Givens and Hoeting [2005], and Gamerman and Lopes [2006]. Properly obtaining a sample from the posterior distribution has always been the main issue for Bayesian statisticians. With the exception of a few rare situations (usually when priors and posteriors are conjugate), it is difficult to obtain closed-form expressions of the posterior distributions, and one needs to rely

on numerical methods to obtain posterior samples. The MCMCs are a class of widely used methods that have been shown to be able to efficiently generate posterior samples.

The Monte Carlo method is a computational technique that provides numerical solutions to problems using repeated random sampling. Monte Carlo methods compute probabilities heuristically as if they are actually happening by running stochastic simulations many times. A good example is approximation of the value of π: Using the idea of the Monte Carlo method, one can randomly draw uniform samples from the unit square and count the number of times samples are in the unit circle. That number divided by the sample size and multiplied by four is a good approximation of the value of π.

Although traditional Monte Carlo methods are popular in many optimization and numerical integration problems where it is impossible to obtain a closed-form expression or infeasible to apply a deterministic algorithm, these methods mostly deal with unidimensional distributions. Monte Carlo methods cannot be used in all cases, especially when the problem involves sampling from posterior distributions, where traditional Monte Carlo methods show their limitations [Givens and Hoeting, 2005]. In contrast, simulation methods based on Markov chains, namely MCMC, are much more flexible and thus can deal with many problems, such as posterior sampling, that are beyond the traditional Monte Carlo methods.

A Markov chain requires that the distribution of the random variable $\boldsymbol{\theta}$ at time $t + 1$ given previous values of $\boldsymbol{\theta}$ at all previous times only depends on the $\boldsymbol{\theta}$ value at time t, and this distribution is independent of time t. That is

$$f\left(\boldsymbol{\theta}^{(t+1)} \,\middle|\, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(t)}\right) = f\left(\boldsymbol{\theta}^{(t+1)} \middle| \boldsymbol{\theta}^{(t)}\right). \tag{2.2}$$

When the Markov chain is irreducible, aperiodic, and positive-recurrent, as $t \rightarrow \infty$ the distribution of $\boldsymbol{\theta}^{(t)}$ converges to its equilibrium distribution, and this is independent of the initial values of the chain $\boldsymbol{\theta}^{(1)}$ [Gilks et al., 1996]. Metropolis et al. [1953] first introduced MCMC into physics in 1953. Hastings generalization of the Metropolis algorithm came in 1970 [Hastings, 1970], and then Gemen and Gemen developed the Gibbs sampler in 1984 [Geman and Geman, 1984]. By the late 1980s, MCMC methods were implemented in Bayesian methods [Tanner and Wong, 1987; Gelfand and Smith, 1990; Gelfand et al., 1990] and soon became the major computational tool in Bayesian inference. The main idea of MCMC methods is to construct a Markov chain whose equilibrium distribution is the target distribution (i.e., the chain converges to the target distribution). This separates MCMC from other common simulation methods that sample directly from the target distribution. Another distinction is that, in contrast to independent samples obtained from common simulations, samples from MCMC are dependent since they are from Markov chains. For the details of MCMC methods see Gilks et al. [1996].

In the Bayesian analysis context, the posterior distribution is the target and Markov chains need to be constructed so they converge to the posterior distribution. Additionally, these Markov chains must easily generate samples from the conditional distribution $f(\boldsymbol{\theta}^{(t+1)|}\boldsymbol{\theta}^{(t)})$. With these requirements satisfied, the following steps generate samples from the posterior distribution from which inferences can be drawn:

1. Set an initial value $\boldsymbol{\theta}^{(0)}$ for the Markov chain(s).
2. Generate N values (iterations) until the chain(s) reaches equilibrium.
3. Perform convergence diagnostics by examining the trace, density, and auto-correlation plots. If the diagnostics fail, generate more samples.
4. Discard the first M observations (burn-in) and use the remainin $N-M$ values as a posterior sample.
5. Obtain the summaries of the posterior sample, such as the mean, median, standard deviation, quantiles, and correlations. Make Bayesian inferences using these posterior summaries.

2.2 **Markov Chain Monte Carlo Algorithms**

Many MCMC algorithms have been developed since the introduction of the initial Metropolis algorithm in 1953 [Metropolis et al., 1953]. The Metropolis-Hastings algorithm, which Hastings [Hastings, 1970] developed based on the original Metropolis algorithm, and the Gibbs sampler [Geman and Geman, 1984], are the two most popular methods. Recent advances include the slice Gibbs sampler [Higdon, 1998; Damien et al., 1990; Neal, 2003], the reversible jump MCMC [Green, 1995], and perfect sampling [Propp and Wilson, 1996; Merller, 1999]. These variants are all based on the original algorithms and most of them deal with specific problems. In this section I provide brief introductions to the two major MCMC methods, the Metropolis-Hastings algorithms, and the Gibbs sampler. For details about MCMC methods, see

Gilks et al. [1996], Robert and Casella [2004], Givens and Hoeting [2005], and Gamerman and Lopes [2006].

2.3 **The Metropolis-Hastings Algorithm**

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. In practice, however, the choice of the proposal distribution is important since poorly chosen proposals will significantly slow down convergence.

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

1. Set initial values $\boldsymbol{\theta}^{(0)}$.
2. For t = 1, ... , T repeat the following steps
   (a) Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.
   (b) Generate new candidate values for $\boldsymbol{\theta}'$ from $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$.

(c) Calculate probability $\alpha = min(1, \frac{f(\boldsymbol{\theta}'|y)q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{f(\boldsymbol{\theta}|y)q(\boldsymbol{\theta}'|\boldsymbol{\theta})})$.

(d) Update $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}'$ with probability α.

Since the normalizing constant $f(y)$ in $f(\boldsymbol{\theta}|y)$ cancels out, the probability α can be re-written as

$$\alpha = min(1, \frac{f(y|\boldsymbol{\theta}')f(\boldsymbol{\theta}')q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{f(y|\boldsymbol{\theta})f(\boldsymbol{\theta})q(\boldsymbol{\theta}'|\boldsymbol{\theta})})$$

There are some special cases of the Metropolis-Hastings algorithm. Random-walk Metropolis uses a special proposal $q(\boldsymbol{\theta}'|\boldsymbol{\theta}) = q(|\boldsymbol{\theta}' - \boldsymbol{\theta}|)$ instead of the symmetric proposal in the original algorithm, resulting in an acceptance probability $\alpha = min(1, \frac{f(\boldsymbol{\theta}'|y)}{f(\boldsymbol{\theta}|y)})$ that depends only on the posterior distribution. In the independence sampler, the proposal distribution does not depend on the previous state of the chain. This algorithm is efficient when the proposal is a good approximation to the posterior distribution. And in contrast to the random-walk Metropolis, where the optimal acceptance rate is around 0.25, the acceptance rate for independence sample must be high enough for the algorithm to be efficient. Component-wise Metropolis-Hastings, which is also called Metropolis within Gibbs, is an algorithm where the parameter vector is divided into subvectors that are updated sequentially using original Metropolis-Hastings procedures. The advantage of this algorithm is that it involves sampling from distributions of lower dimensions that are usually straightforward and computationally less intensive.

2.4 **The Gibbs Sampler**

Geman and Geman developed the Gibbs sampler [Geman and Geman, 1984], which uses the full conditional distribution as the proposal. The Gibbs sampler is usually cited as a separated MCMC technique, due to its popularity, even though it is only a special case of the single-component Metropolis-Hastings algorithm. The full conditional distribution $f(\theta_j \mid \boldsymbol{\theta}_{\setminus j}, \boldsymbol{y})$ is the distribution of the jth 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, ... , T repeat the following steps

   (a) Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.

   (b) For j = 1, ... , d update $\boldsymbol{\theta_j}$ from $\boldsymbol{\theta_j} \sim f(\boldsymbol{\theta_j} \mid \boldsymbol{\theta}_{\setminus \mathbf{j}}, \boldsymbol{y})$.

   (c) Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$ where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_j, \ldots, \theta_d)$.

2.5 **Empirical Bayes Estimation**

In this section, the basic ideas and methods of the EB approach, without loss of generality, will be considered but can be extended to the models presented. For reasons of simplicity, in this section I only consider models with fixed and random effects. Estimation of model parameters with fixed covariates may be similarly conducted by including the corresponding covariate parameters in the coefficient vector $\boldsymbol{\beta}$. The mixed-effects model for the i^{th} hospital:

$$\boldsymbol{y_i} = \boldsymbol{X_i\beta} + \boldsymbol{Z_i\gamma_i} + \boldsymbol{\varepsilon_i}, \tag{2.3}$$

where $\boldsymbol{y_i}$ is a vector of the outcome for the i^{th} hospital, $\boldsymbol{\beta}$ is a vector for the fixed-parameters, $\boldsymbol{\gamma_i}$ is a vector of random effects for the i^{th} hospital, $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ are design matrices of the fixed and random effects for the i^{th} hospital , and $\boldsymbol{\varepsilon_i}$ is a vector of errors.

Then the observation $\boldsymbol{y_i}$ and random effects $\boldsymbol{\gamma_i}$ have joint normal distribution

$$\begin{bmatrix} y_i \\ \gamma_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix}, \begin{bmatrix} Z_i\Sigma_\gamma Z_i^T + \Sigma_i & Z_i\Sigma_\gamma \\ \Sigma_\gamma Z_i^T & \Sigma_\gamma \end{bmatrix}\right), \tag{2.4}$$

where the covariance matrix of the random effects is Σ_γ and Σ_i is a block diagonal matrix of the error covariance matrices of all links.

Estimation of the model parameters is based on an expectation-maximization algorithm. In the E step, given the data and utilizing the current values of the other parameters, the "expected *a posteriori*" or EB estimates of the random effects as well as the conditional variances of the random effects are computed. In the M step, the maximum marginal likelihood (MML) estimates of the regression coefficients, error variances, and the variances of the random effects are obtained, given the current values of the random effects. The algorithm iterates between the EB and MML estimates until convergence.

Two stochastic processes are taken into account when making EB inference, one for the data $f(\boldsymbol{y_i}|\boldsymbol{\gamma_i})$, and the other one for the random effects $g(\boldsymbol{\gamma_i})$. The posterior distribution of the random effects given data, $g(\boldsymbol{\gamma_i}|\boldsymbol{y_i})$, contains all information about $\boldsymbol{\gamma_i}$ available in $\boldsymbol{y_i}$. Bayes theorem states:

$$g(\boldsymbol{\gamma_i}|\boldsymbol{y_i}) = \frac{f(\boldsymbol{y_i}|\boldsymbol{\gamma_i})g(\boldsymbol{\gamma_i})}{h(\boldsymbol{y_i})} \tag{2.4}$$

In the EB approach, estimates of the posterior mean $\bar{\boldsymbol{\gamma}}_{\boldsymbol{i}}$ and the covariance matrix $\bar{\Sigma}_{\gamma|}\boldsymbol{y_i}$ of the random effects are calculated. Since $\boldsymbol{\gamma_i}$ and $\boldsymbol{y_i}$ are jointly normally distributed (2.4), the conditional distribution can be written as

$$\boldsymbol{\gamma_i}|\boldsymbol{y}_i \sim N(\Sigma_\gamma \boldsymbol{Z_i^T}\left(\boldsymbol{Z_i}\Sigma_\gamma \boldsymbol{Z_i^T} + \Sigma_j\right)^{-1}(\boldsymbol{y}_j - \boldsymbol{X_i}\boldsymbol{\beta}),$$
$$\Sigma_\gamma - \Sigma_\gamma \boldsymbol{Z_i^T}(\boldsymbol{Z_i}\Sigma_\gamma \boldsymbol{Z_i^T} + \Sigma_i)^{-1}\boldsymbol{Z_i}\Sigma_\gamma). \tag{2.5}$$

The posterior mean provides the EB estimate of $\boldsymbol{\gamma_i}$ and the posterior covariance matrix provides the EB estimate of the uncertainty about $\boldsymbol{\gamma_i}$. Hence, the estimates are calculated as follows:

$$\widetilde{\boldsymbol{\gamma}}_i = \Sigma_\gamma \boldsymbol{Z_i^T}\left(\boldsymbol{Z_i}\Sigma_\gamma \boldsymbol{Z_i^T} + \Sigma_j\right)^{-1}(\boldsymbol{y}_j - \boldsymbol{X_i\beta})$$

$$= \Sigma_\gamma[\Sigma_\gamma + (\boldsymbol{Z_i^T}\,\Sigma_i^{-1}\,\boldsymbol{Z_i})^{-1}]^{-1}(\boldsymbol{Z_i^T}\,\Sigma_i^{-1}\boldsymbol{Z_i})^{-1}\boldsymbol{Z_i^T}\Sigma_i^{-1}(\boldsymbol{y_i} - \boldsymbol{X_i\beta})$$

$$= \boldsymbol{R}(\boldsymbol{Z_i^T}\,\boldsymbol{Z_i})^{-1}\boldsymbol{Z_i^T}(\boldsymbol{y_i} - \boldsymbol{X_i\beta})$$

and

$$\tilde{\Sigma}\boldsymbol{\gamma}|\boldsymbol{y_i} = \Sigma_\gamma - \Sigma_\gamma \boldsymbol{Z_i^T}\left(\boldsymbol{Z_i}\Sigma_\gamma \boldsymbol{Z_i^T} + \Sigma_i\right)^{-1}\boldsymbol{Z_i}\Sigma_\gamma\Big)$$

$$= (\boldsymbol{Z_i^T}\Sigma_i^{-1}\,\boldsymbol{Z_i} + \Sigma_\gamma^{-1})^{-1}$$

$$= (\boldsymbol{I} - \boldsymbol{R})\Sigma_\gamma\,,$$

where $\boldsymbol{R} = \Sigma_\gamma[\Sigma_\gamma + (\boldsymbol{Z_i^T}\,\Sigma_i^{-1}\boldsymbol{Z_i})^{-1}]^{-1}$.

3. DATA

3.1 Hospital Performance Measure Data

Hospital performance measure data focuses on binary process accountability measures within four measures sets based in the clinical areas of AMI, HF, PN, and SCIP. Process measures are those measures that determine whether or not the specific clinical therapy within a clinical area has been provided to the patient. Measures and therapies will be used interchangeably throughout this dissertation. Accountability measures are those measures that meet four criteria that produce the greatest impact on patient outcomes when hospitals demonstrate improvement. The first criteria for an accountability measure is that strong scientific research demonstrates that performing the evidence-based process of care improves health outcomes. The second criterion is proximity, in that performing the care process is closely connected to patient outcomes. The measure must accurately assess whether or not the care has actually been provided for the third criteria. The final criterion is that implementing the measure will have little or no chance of incurring adverse consequences. Table I shows the process measures that were active from January 1, 2010 to December 31, 2010. There are six process measures in AMI; four process measures in HF; PN includes six process measures, and SCIP includes five process measures, three which span seven surgical procedures: blood vessel surgery, colorectal surgeries, coronary artery bypass graft surgeries, hip replacement surgeries, hysterectomies, knee replacement surgeries, and open heart surgeries. Scientific evidence shows that providing care within each of these individual processes represents the best practice in the

treatment of each clinical area with higher rates representing better care. The Joint Commission challenges hospitals to achieve an observed target rate of 95% or higher for each of these measures.

The data were collected from Joint Commission accredited hospitals that were participating in the Joint Commission's ORYX initiative. Hospitals submitted aggregated monthly data for the measures between July 31, 2011 and July 31, 2013. The data submitted were subject to edit checks to ensure quality and consistency. Issues identified in data quality require hospitals to retransmit data to ensure data accuracy and quality.

Performance measure data were obtained from the Joint Commission's Quality Check website (http://www.qualitycheck.org), which contains patient data aggregated quarterly and yearly at the hospital level but not at the patient level. Each Joint Commission accredited hospital submits patient level data to an intermediary organization called a Performance Measurement System (PMS). Approximately 50 PMS organizations service all Joint Commission accredited hospitals. Each PMS applies a measure specification algorithm to the patient level data based on the ICD-9-CM codes and determines whether the patient meets the measure requirements. The PMS then aggregates the number of patients eligible for each measure and the number of patients that received the therapy for each measure and sends monthly aggregate data to the Joint Commission.

TABLE I

ACTIVE HOSPITAL PROCESS MEASURES FROM 3Q12-2Q13

Measure Set	Measure ID	Measure Name
AMI	AMI-1[a]	Aspirin at Arrival
	AMI-2[a]	Aspirin Prescribed at Discharge
	AMI-3[a]	ACEI or ARB for LVSD
	AMI-4	Smoking Cessation Counseling
	AMI-5[a]	Beta Blocker Prescribed at Discharge
	AMI-7[a]	Fibrinolytic Therapy Received within 30 Minutes of Hospital Arrival
	AMI-8[a]	Primary PCI Received within 90 Minutes of Hospital Arrival
HF	HF-1	Discharge Instructions
	HF-2	Evaluation of LVS Function
	HF-3[a]	ACEI or ARB for LVSD
	HF-4	Adult Smoking Cessation Advice/Counseling
PN	PN-2	Pneumococcal Vaccination
	PN-3b[a]	Blood Cultures Performed in the Emergency Department prior to Initial Antibiotic Received in Hospital
	PN-4	Adult Smoking Cessation Advice/Counseling
	PN-5c[a]	Initial Antibiotic Received within 6 Hours of Hospital Arrival
	PN-6a[a]	Initial Antibiotic Selection for CAP in Immunocompetent—ICU
	PN-6b[a]	Initial Antibiotic Selection for CAP in Immunocompetent—Non-ICU
SCIP	SCIP-Inf-1a[a]	Prophylactic Antibiotic Received within 1 Hour prior to Surgical Incision—Overall Rate
	SCIP-Inf-2a[a]	Prophylactic Antibiotic Selection for Surgical Patients—Overall Rate
	SCIP-Inf-3a[a]	Prophylactic Antibiotics Discontinued within 24 Hours after Surgery End Time—Overall Rate
	SCIP-Inf-4[a]	Cardiac Surgery Patients with Controlled 6 a.m. Postoperative Blood Glucose
	SCIP-Inf-6[a]	Surgery Patients with Appropriate Hair Removal

a—an accountability measure

3.2 **Hospital Performance Composite Measures**

I consider an overall yearly composite score for each of the four clinical areas (AMI, HF, PN, and SCIP) as defined by the opportunity model. Data are obtained quarterly and are analyzed yearly by combining all four quarters of data for patient discharges in the 12-month period from 3rd quarter of 2012 to 2nd quarter of 2013. For each process measure, a patient that meets the measure criteria has the opportunity to receive the appropriate measure defined therapy. For example, for the AMI measure aspirin at arrival, if a patient arrives at the hospital with a diagnosis of AMI and the patient is eligible to receive the therapy of taking an aspirin, the measure identifies if the patient did or did not receive the aspirin therapy. Thus, the composite measure in each clinical area shows the percentage of time that the appropriate therapies were provided.

Computation of a composite measure in each clinical area is as follows. Let y_{ijk} be a binomial variable that represents the numbers of patients that received the k^{th} therapy in the j^{th} clinical area at the i^{th} hospital. Let n_{ijk} be the number of eligible patients to receive the k^{th} therapy in the j^{th} clinical area at the i^{th} hospital. Let J_i be the number of clinical areas submitted by the i^{th} hospital ranging from 1 to 4 depending on the number of measure sets submitted. For example, if a hospital chose to submit only measures in AMI and PN, then J_i will be 2. Let the number of therapies in the j^{th} clinical area denoted by K_{ij} be 6 for AMI, 1 for HF, 4 for PN, and 5 for SCIP in the i^{th} hospital. Let $y_{i..} = \sum_j^{J_i} \sum_k^{K_{ij}} y_{ijk}$ be the total number of times appropriate therapies were provided in all the clinical areas in the i^{th} hospital. Similarly, let $n_{i..} = \sum_j^{J_i} \sum_k^{K_{ij}} n_{ijk}$ be the total

number of opportunities to provide appropriate therapies in all the clinical areas in the i^{th} hospital. The observed overall composite rate of providing appropriate therapies in all clinical areas in the i^{th} hospital is defined as $p_i^{obs} = y_{i...}/n_{i.}$.

3.3 **Safety Data**

The CDC collects in their NHSN and provides the data to the CMS for public dissemination on the CMS Hospital Compare website (http://www.medicare.gov/hospitalcompare/search.html).

The HAI measures are collected directly from hospitals through existing commercial infection control surveillance systems and electronic medical records. This system ensures timely data collection that also minimized data collection efforts that allows healthcare organizations to focus their attention on preventing HAIs. The HAI measures included along with the reporting time period are summarized in Table II.

TABLE II

HOSPITAL HEALTHCARE ACQUIRED INFECTION MEASURES

HAI Measure	Reporting Period
CAUTI	7/1/2012–9/30/2013
SSI:colon	7/1/2012–9/30/2013
SSI: hysterectomy	7/1/2012–9/30/2013
CLABSI	7/1/2012–9/30/2013
MRSA	7/1/2012–9/30/2013
C. difficile	7/1/2012–9/30/2013

A CAUTI, as defined by the CDC, is a UTI where the indwelling urinary catheter was in place for more than two calendar days on the date of the event, with the day of device placement being Day 1 and an indwelling urinary catheter was in place on the date of the event for the day before. If an indwelling urinary catheter was in place for more than two calendar days and then removed, the UTI criteria must be fully met on the day of discontinuation or the next day. If a patient in transferred from one location to another and an infection is present within two calendar days, then the infection is associated with the transferring location. Device and patient days are used for denominators. Indwelling urinary catheter days, which are the number of patients with an indwelling urinary catheter device, are collected at the same time each day. The daily counts are summed and are used for the total for each month. When denominator data are available from electronic databases, these sources may be used as long as the counts are not substantially different (+/- 5%) from manually collected counts, validated for a minimum of three months.

An SSI must occur within 30 days after any NHSN operative procedure, involves only skin and subcutaneous tissue of the incision, and the patient must have at least one of the following: purulent drainage from the superficial incision; organisms isolated from an aseptically obtained culture of fluid or tissue from the superficial incision; superficial incision that is deliberately opened by a surgeon and is culture-positive or not cultured and the patient has at least one of the following signs or symptoms (pain or tenderness, localized swelling, redness, or heat). A culture-negative finding does not meet this criteria; diagnosis of a superficial incisional SSI by the surgeon or attending physician or other designee. Denominator data are those patients that have had the procedures of colon surgery and hysterectomy, respectively.

A central venous catheter, known as a central line, is a tube or catheter that is placed in a large vein in the chest, neck, or groin in order to provide medication or fluids to patients or to collect blood for medical tests. Central lines are used to access a major vein that is close to the heart and can remain in a patient for a long period of time (weeks or months) and can be the source of serious infections. A CLABSI is a very serious infection that occurs when either a bacteria or virus enters the bloodstream through the central line. Strict protocols must be adhered to when inserting the line in order to maintain sterility of the central line and prevent CLABSI. Stringent infection control practices must be followed each time a healthcare provider checks the central line or changes the dressing around the central line. Patients who get a CLABSI often show signs of fever and redness or soreness around the central line. Preventative measures to reduce CLABSI include performing hand hygiene, applying appropriate skin antiseptic, ensuring the skin preparation agent has dried completely prior to insertion of the central line, using all maximal sterile barrier precautions (sterile gloves, sterile gown, cap, mask, and large sterile drape), following all recommended practices, washing hands or using alcohol-based hand rub before and after touching the central line once in place, and finally, removing the central line as soon as it is unnecessary.

Methicillin-resistant *Staphylococcus aureus*, or MRSA, is a type of staph bacteria that is resistant to many antibiotics. In a hospital or nursing home, MRSA can cause severe problems such as bloodstream infections, PN, and SSIs. The spread of MRSA is typically caused by direct contact with an infected wound or contaminated hands of healthcare providers. People may also carry MRSA without having signs of infection and spread the bacteria to others and potentially

cause infection. Studies show that when healthcare providers follow CDC guidelines, then MRSA can be completely prevented.

Clostridium difficile (C. difficile) is a bacterium that causes colitis, an inflammation in the colon. Those at risk of acquiring *C. difficile* include the elderly and patients that have other illnesses or conditions requiring long-term use of antibiotics. This bacterium is found in the feces and is transmitted if a patient touches items or surfaces that are contaminated and then touch their mouths or mucous membranes. Hand contact is the most common way that healthcare providers spread these bacteria to patients or other surfaces. The *C. difficile* bacterium can survive for long periods of time on surfaces. Symptoms include watery diarrhea (at least three bowel movements per day for two or more days), fever, loss of appetite, nausea, and abdominal pain and tenderness.

The HAI measures are obtained from the NHSN data collection tool run by the CDC. Hospitals enter the required data into the data collection tool. Patients included in the HAI measures come from acute care hospitals and are adult, pediatric, neonatal, Medicare, and non-Medicare patients. These measures show how infections are contracted in a hospital during the course of medical treatment, and are important because these infections can be prevented if the hospital follows established safe care guidelines.

4. HIERARCHICAL MIXED-EFFECTS LOGISTIC REGRESSION MODELS

4.1 Binomial Regression Model

Both the Joint Commission and the CMS have adopted the use of the opportunity model to define their composite measures in order to simplify multidimensional measures into a one-dimensional score to summarize the overall hospital quality in a therapeutic area. This opportunity model was developed in 1998 by the Hospital Core Performance Measure Project, and was based on the assumption that whenever a patient meets the criteria for a measure population there exists an opportunity to provide the appropriate evidence-based intervention [CMS, 2005]. This model is currently used by the Joint Commission's Quality Check for each hospital reported clinical area based on a moving 12-month time period to represent the most current practices of the hospital as determined by the most recent transmission of data. Additionally, the Joint Commission has recently instituted a hospital recognition program that utilizes an overall hospital composite measure using the opportunity model by combining all available accountability measures for a given calendar year.

4.2 Static Year Analysis

A mixed-effects or latent variable model is used for the observed measure rates to analyze the composite measure rate. Teixeira-Pinto and Normand suggest a two-parameter Normal-Ogive [1997] model based on Landrum et al. [2000] profiling healthcare providers, which replicates

item response theory (IRT) models used in psychological and educational testing [van der Linden et al., 1997; Alagumalai et al., 2005] based on a single clinical area. Cohen [2003] shows an education institution's overall average proficiency is modeled utilizing a clustered IRT model where the latent variable is clustered with students within grades within different classrooms or schools. This clustered IRT model is extended to composite measures among multiple clinical areas within a hospital to get an overall measure of hospital quality while still retaining the unidimensional IRT model.

In this model, the latent variable θ_i represents the quality of care in the i^{th} hospital where larger values of θ_i are associated with a higher quality of care. Let α_j be the fixed effect representing the overall mean rate for the j^{th} clinical area on the logit scale; σ_j be a positive measure set-specific discrimination weight fixed effect. Let p_{ij} be the probability that the proper therapy was given within the j^{th} clinical area for the i^{th} hospital. I propose the following model:

$$logit(p_{ij}) = g_{ij}(\alpha, \sigma, \beta, \theta), \tag{4.1}$$

where

$$g_{ij}(\alpha, \sigma, \beta, \theta) = \alpha_j + \sigma_j * \theta_i$$

and $\sigma_j > 0$, $\theta_i \sim f(*)$, where $f(*)$ is a prior probability distribution.

Hence

$$\Pr(y_{ij} = 1) = p_{ij} = \frac{\exp(g_{ij}(\alpha,\sigma,\beta,\theta))}{1+\exp(g_{ij}(\alpha,\sigma,\beta,\theta))}. \tag{4.2}$$

Composite measures should have larger values of σ_j for those therapeutic areas that are less homogeneous among the hospitals. I will consider two separate prior distributions for θ_i and compare the results between both estimates of θ_i. The first distribution *f(*)* will be a standard normal distribution, i.e., $\theta_i \sim N(0,1)$. In this instance, the latent score, θ_i, ranges from $-\infty$ to ∞, where large negative values are indicative of poor quality of care, values of zero are indicative of average quality of care, and large positive values are indicative of high quality of care. Although assuming normality on the latent variable may be too restrictive, it solves the problem of the within-hospital correlation between the therapeutic areas. The second distribution *f(*)* is a rectangular distribution, that is, $\theta_i \sim$ Uniform(-a, a). The value of "a" will be determined by exploration with the corresponding data. With this second prior distribution, I assume the latent score to have equal probability across all where larger values of θ_i indicate superior hospital quality and lower values of θ_i suggest lower hospital quality. Due to the lack of patient-level data, the within patient variability cannot be estimated as only aggregate data are publically available at this point in time. It is not known whether a particular patient was eligible for one or more therapies within a measure set, or clinical area.

I adopt an EB approach using model 4.1 utilizing marginal maximum likelihood methods to estimate model parameters. This approach is directly applied within the framework of PROC NLMIXED in SAS version 9.3 where EB estimates of the random effects are obtained in addition to maximum likelihood estimates of the fixed effects. Gaussian-Hermite quadrature is used for numerical integration of the likelihood over the random effects with nonadaptive scaling used for the quadrature. This means that the quadrature points are centered around the EB

estimates for the random effects but the current random effects variance matrix is used as the scale matrix. The optimization technique used within the scope of the EB model is Newton-Raphson. All parameters within the EB method are assumed to be normally distributed with the additional constraint of $\sigma_i > 0$.

I also adopt an FB approach with model 4.1 using the OpenBUGS package. The assumption of a half-normal distribution of σ_i, recommended by Gelman [2006] with small precision that represents vague prior information and constrains the parameter to be positive. For the other parameters, prior distributions are chosen to be N(0,100). Posterior means and posterior variances of θ_i are estimates of the quality of care and the variance. 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.

4.3 **Agreement between Predicted Measures**

To compare the estimates $\hat{\theta}_i$ obtained from each method and an overall observed rate, p_i^{obs}, I incorporate Lin's [1989] CCC. In this case, the CCC measures the agreement between $\hat{\theta}_i's$ obtained from each method as opposed to the Pearson correlation coefficient, which measures their relationship. The CCC is the Pearson correlation coefficient multiplied by a bias correction factor that measures the accuracy, or amount of deviation from a 45º line. Thus, the CCC takes values between -1 and 1 where the value of one indicates a high correlation that is achieved

when the line between the observed and predicted values (also between different predicted values) passes through the origin making a 45º angle with the horizontal line. The CCC is a consistent estimator and it follows asymptotically a normal distribution [Lin, 1992]. The computation of CCC is simple and can be easily implemented.

4.4 Classification of Hospitals

Laird and Lewis [1989], Lockwood et al. [2002], and Shen and Louis [1998] propose various methods and estimators to rank hospitals, and Austin et al. [2001] show how each of these methods or ranking hospitals differs. Classification of hospitals is based on the estimate of quality of care score, S_i, derived from the posterior distribution of θ_i, as opposed to ranking hospitals.

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. I take 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 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, in order 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 γ100%-credible intervals of hospital performance lie above it. I use the symmetry of the credible intervals to determine the cutoff point. Note that the probability that θ_i lies above the lower bound of the ((2γ -1)x100)% credible intervals is γ. For example, if γ = 0.9, then the probability that S_i exceeds the lower bound of its ((2 * 0.9 -1)x100)%=80% credible intervals is 0.9 since the probability of being in the credible interval 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 ((2γ -1)x100)% credible intervals.

Classification of low-performing hospitals uses a similar approach except I determine $P(\theta_i < \eta_{01}) > \gamma$ where η_{01} is the 1st percentile of θ_i and γ is a predefined threshold. Thus, I determine the 1st percentile of the upper bounds of the ((2γ -1)x100)% credible intervals. Using the previous example, probability that S_i is below the upper bound of its ((2 * 0.9 -1)x100)%=80% credible intervals 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 ((2γ -1)x100)% credible intervals. Each classification of either high or low performers based on S_i is compared to classification based on the Joint Commission methodology.

4.5 **Static Year Results**

In this section I discuss the results for the static year 3Q2012 to 2Q2013 utilizing all available data. There were a total of 2,957 hospitals that submitted data in at least one of the four measure sets. A majority of the hospitals had data in all four measure sets. The mean number of measure sets selected was 3.8 and the 25th percentile was 4. The majority of hospitals (2,480, 83.9%) submitted data in all four of the measure sets. Eleven percent of the hospitals had data in only three measure sets, and 156 (5.3%) hospitals only submitted data in two or fewer measure sets (See Figure 1).

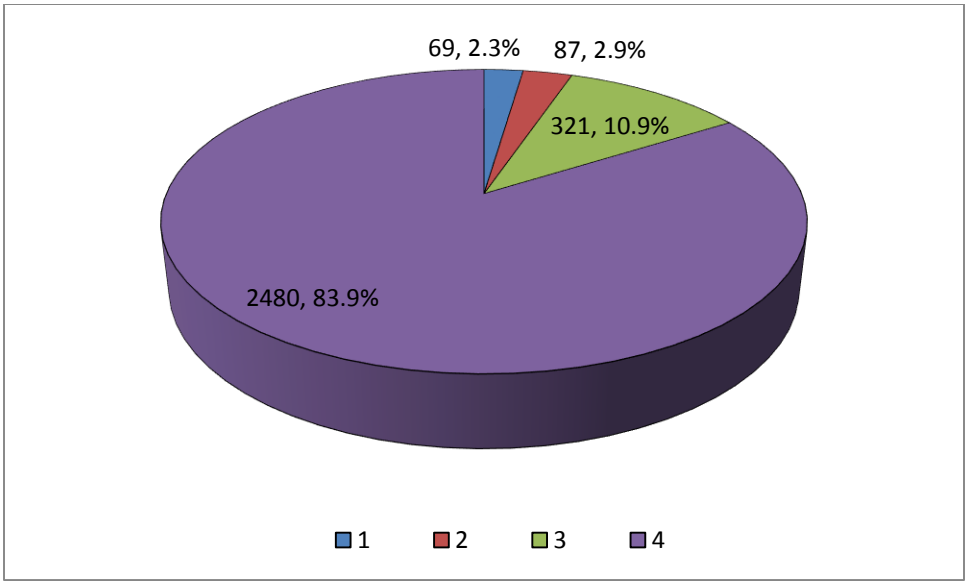

Figure 1. Distribution of the number of measure sets selected (N,%)
1=AMI, 2=HF, 3=PN, 4=SCIP

Table III shows the individual accountability measure rates. For AMI, four of the six measures had average rates that exceeded 95% and two had rates below 90%. Although AMI-7a (See Table III) had the lowest average rate of 51.3%, only 222 (13.7%) organizations provide fibrinolytic therapy. Among those hospitals that provide fibrinolytic therapy, the mean number of patients eligible to receive the therapy is 3.5 with a median number of eligible patients being 1. Measures in the other sets are above 90% with the exception of PN-6a that have a mean rate of 77.21%. Similar to AMI-7a, PN-6a has low denominator count with the mean number of patients eligible for this measure being 13.5 with a median of 11. As shown, rates within each measure can span the entire interval from 0% to 100% indicating that not all hospitals are performing at the desired threshold of 95%.

For the AMI measure set, there were 2,633 hospitals that submitted data. The average number of denominator cases (sum of all AMI denominator cases for each hospital divided by 2,633) for AMI was 1,240.9 with a median number of denominator cases being 699. There were 230 (8.7%) hospitals with less than 30 patients for the AMI composite measures.

There were 2,843 hospitals that submitted HF composite data. There were a mean number of 138.1 denominator cases among the hospitals with a median of 106 patients submitted for analysis. Heart Failure had the highest percentage of hospitals submitting data with less than 30 cases with 483 (17.0%).

TABLE III

DISTRIBUTION OF MEASURES OF A STATIC YEAR (3Q2012–2Q2013)

Measure ID	N	Mean of Rates	Max	Q3	Median	Q1	Min
AMI-1	2,632	0.98373	1	1	1	0.99134	0
AMI-2	2,595	0.98077	1	1	1	0.99091	0
AMI-3	2,602	0.97905	1	1	1	0.98913	0
AMI-5	1,371	0.93003	1	1	0.97059	0.92308	0
AMI-7a	191	0.5319	1	1	0.66667	0	0
AMI-8a	2,595	0.94821	1	1	0.99342	0.96667	0
HF-3	2,843	0.96205	1	1	0.99115	0.95402	0
PN-3b	2,554	0.91411	1	1	1	0.875	0
PN-6a	2,862	0.96239	1	1	0.9759	0.95238	0.33333
PN-6b	2,724	0.97725	1	1	1	0.975	0
SCIP-Inf-1a	2,773	0.98151	1	1	0.99211	0.98206	0
SCIP-Inf-2a	2,773	0.98484	1	0.99919	0.9932	0.9848	0
SCIP-Inf-3a	2,773	0.97213	1	0.99451	0.98411	0.96855	0
SCIP-Inf-4	1,023	0.96257	1	0.98726	0.97273	0.9505	0.59873
SCIP-Inf-6	2,784	0.99804	1	1	1	1	0.33333

The number of hospitals submitting PN data that were used in the PN composite measure was 2,865. The average number of patients per hospital that contributed to the PN composite

measure was 258.6 with a median of 220 patients. There were 64 (2.2%) hospitals that submitted less than 30 patients over the one year time frame for this composite measure.

For the SCIP composite measure, there were 2,785 hospitals that submitted data contributing to the composite. This measure set had the lowest number of hospitals that submitted less than 30 denominator cases for the yearly time span with 26 (0.9%) hospitals. The mean number of patients included in the denominator was 3,391.1 with a median of 2,627 patients. The SCIP measure set had the highest volume of patients among all four measure sets included in this analysis.

TABLE IV

OBSERVED MEASURE SET COMPOSITE AND OVERALL COMPOSITE RATES OF STATIC YEAR (3Q2012–2Q2013)

Measure Set	N	Mean of Rates	Standard Deviation	Max	90th Percentile	Median	10th Percentile	Min
AMI	2,633	0.9720	0.0279	1.0000	1.0000	0.9924	0.9333	0.0000
HF	2,843	0.9620	0.0508	1.0000	1.0000	0.9912	0.8888	0.0000
PN	2,865	0.9630	0.0412	1.0000	1.0000	0.9758	0.9206	0.0000
SCIP	2,785	0.9824	0.0193	1.0000	0.9977	0.9885	0.9660	0.5000
Overall Composite	2,957	0.9770	0.0359	1.0000	0.9965	0.9866	0.9544	0.5366

The accountability measures are combined to form a composite score for each of the measure sets and combined to form an overall composite score, as defined in section 3.2, for the hospital shown in Table IV. For all four measure sets, the average of the hospital rates is above

95%. By comparing Table III and Table IV there is a loss in the variability between the individual measures when a composite measure is calculated because those individual measures with low rates are averaged out when computing a composite measure. The minimum composite score for a hospital in the SCIP observed composite measure is 50%, whereas the individual measures had minimum rates of as low as 0%. This is also seen in the overall composite measure where the minimum observed hospital composite measure has increased to 53.66%, potentially over-inflating the true rate and losing the ability to detect areas of improvement at the measure level.

TABLE V

PEARSON CORRELATIONS BETWEEN MEASURE SET RATES

Pearson Correlation Coefficients, Prob > |r| under H0: Rho=0 N				
	AMI	HF	PN	SCIP
AMI	1.00 <.0001 2,632			
HF	0.2830 <.0001 2,620	1.00 <.0001 2,841		
PN	0.3705 <.0001 2,602	0.3528 <.0001 2,809	1.00 <.0001 2,865	
SCIP	0.3679 <.0001 2,517	0.3531 <.0001 2,678	0.4433 <.0001 2,694	1.00 <.0001 2,784

Within each measure set, the Pearson Correlation coefficients are computed for the rates between each of the accountability measures. For AMI, the correlation coefficients range from 0.18561 to 0.52435, whereas the coefficients ranged from 0.2698 to 0.40344 for PN. The correlation coefficients for SCIP ranged from 0.16247 to 0.46595. No correlations are computed for HF because there is only one accountability measure within that measure set. Table V shows a similar trend of correlations coefficients between the rates each of the four measure sets. The highest correlation is between SCIP and PN with a correlation coefficient of 0.4433 and the lowest correlation is between AMI and HF with a correlation coefficient of 0.2830. This suggests that there is very little correlation between clinical areas, even though they are statistically significant.

The FB analysis of model 4.1 was performed in OpenBUGS. Convergence diagnostics calculated within OpenBUGS to determine the Gelman-Rubin convergence statistic (<1.1) justifies the burn-in choice. Figures 2 and 3, as an example, graphically show convergence of α_i and σ_i for the model using a standard normal prior distribution of θ_i. Both chains in the figures converge after 2,000 samples are taken. Similarly, for the model using the rectangular distribution for θ_i, the convergence of σ_i occurred after approximately 1,000 more samples, as seen in Figure 4, than using the standard normal prior for θ_i, As seen in these figures, the chain that started close to the posterior distribution converged quicker than the chain with initial values further away from the final posterior estimate.

For each model with different assumptions of the prior distribution of θ_i(standard normal and rectangular), the last 5,000 posterior samples of each chain were used for analysis. Each

model had a low ratio of Monte Carlo error to posterior standard deviation for each model parameter. Figure 5 shows the posterior density for α_i with the standard normal prior for θ_i whereas Figure 6 shows the posterior density for α_i with the rectangular prior for θ_i. Both densities are similar with estimates from the model using the standard normal prior for θ_i being slightly higher (see Table VII) than the estimates from the model using the rectangular distribution prior for θ_i with the percent change from the former model ranging from 7% to 9%. The DIC for this normal prior was 89,070.0 and the rectangular prior DIC was 89,586.1.

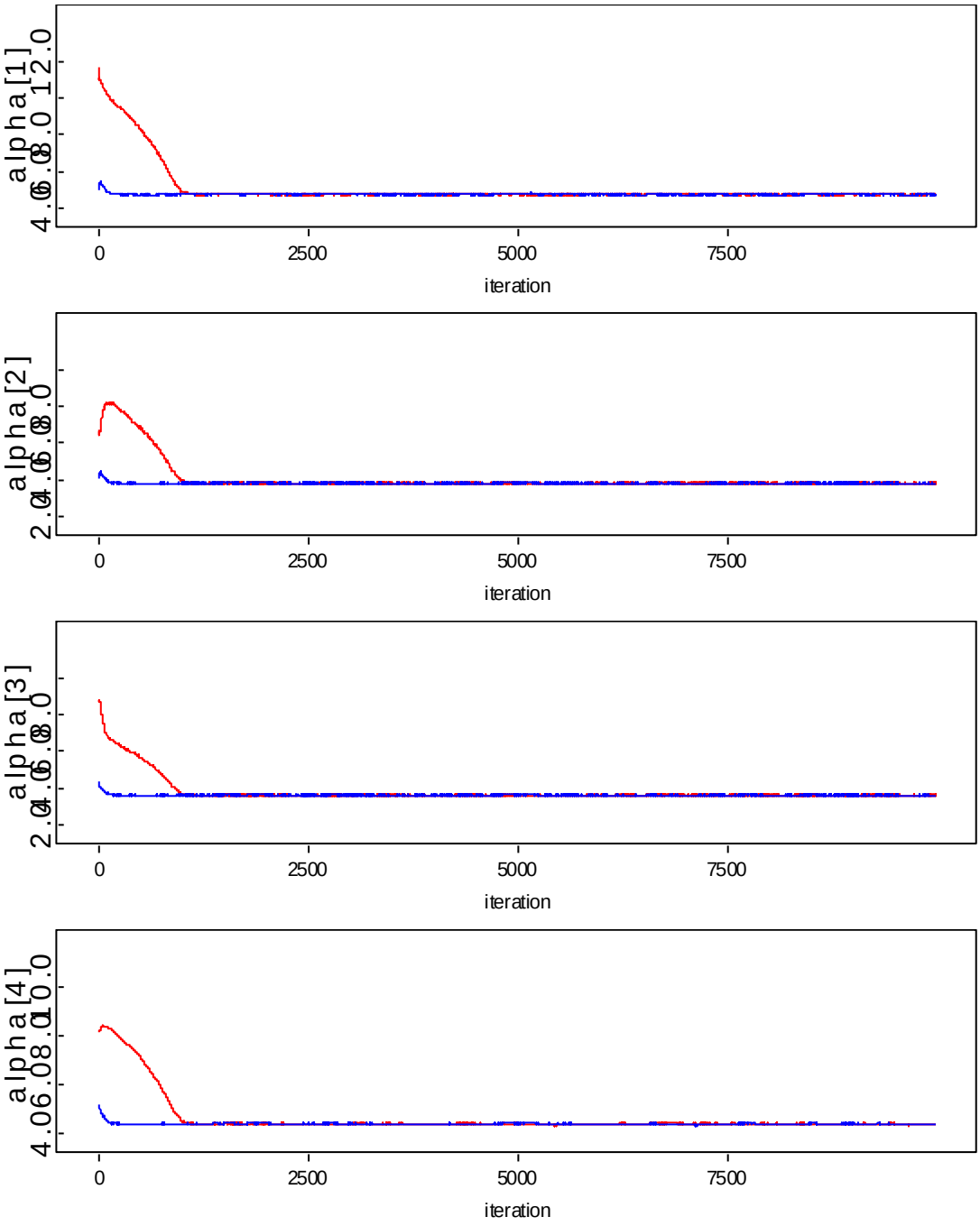

Figure 2. Diagnostic plots of α_i with Normal prior of θ_i.
1=AMI, 2=HF, 3=PN, 4=SCIP

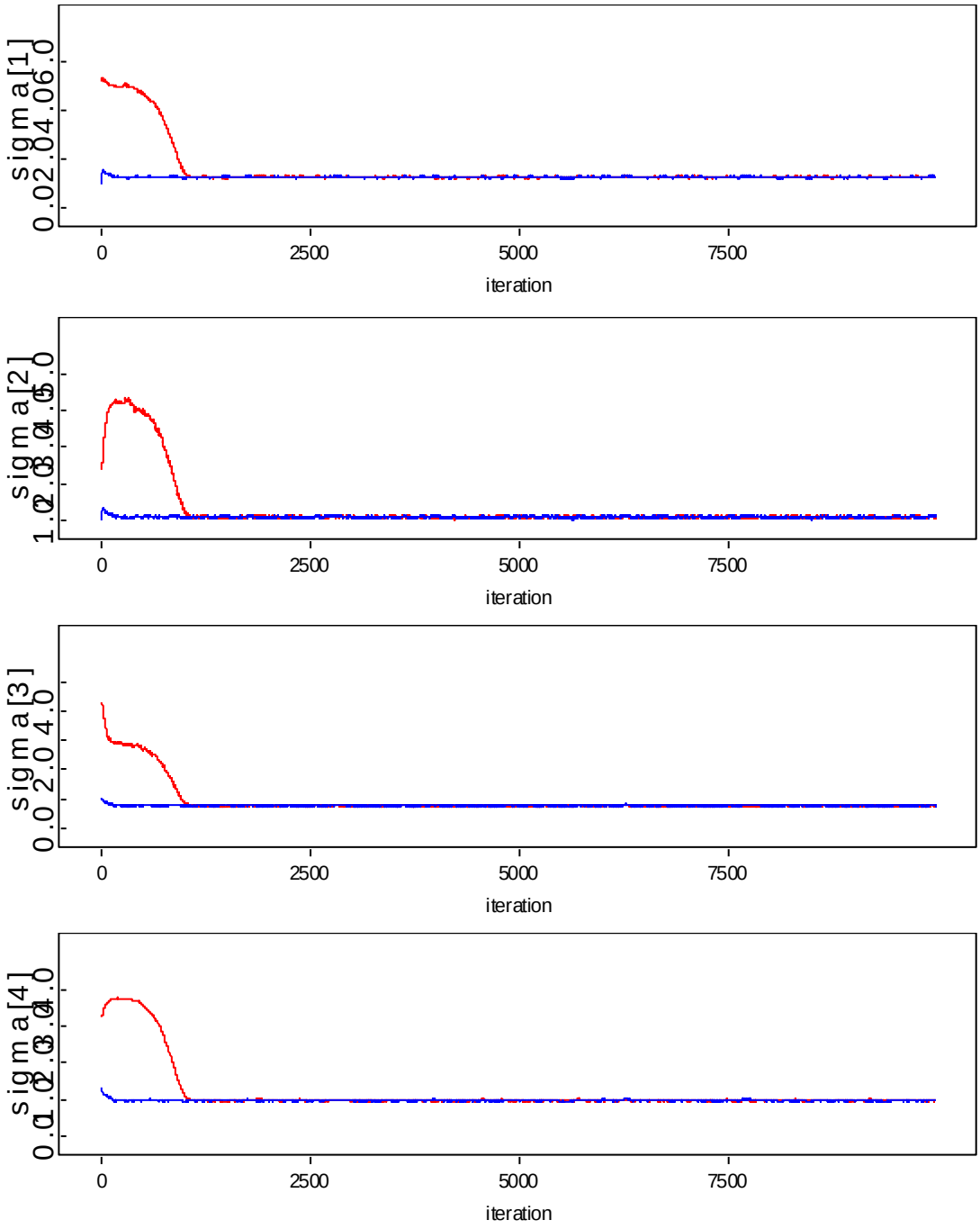

Figure 3. Diagnostic plots of σ_i with Normal prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

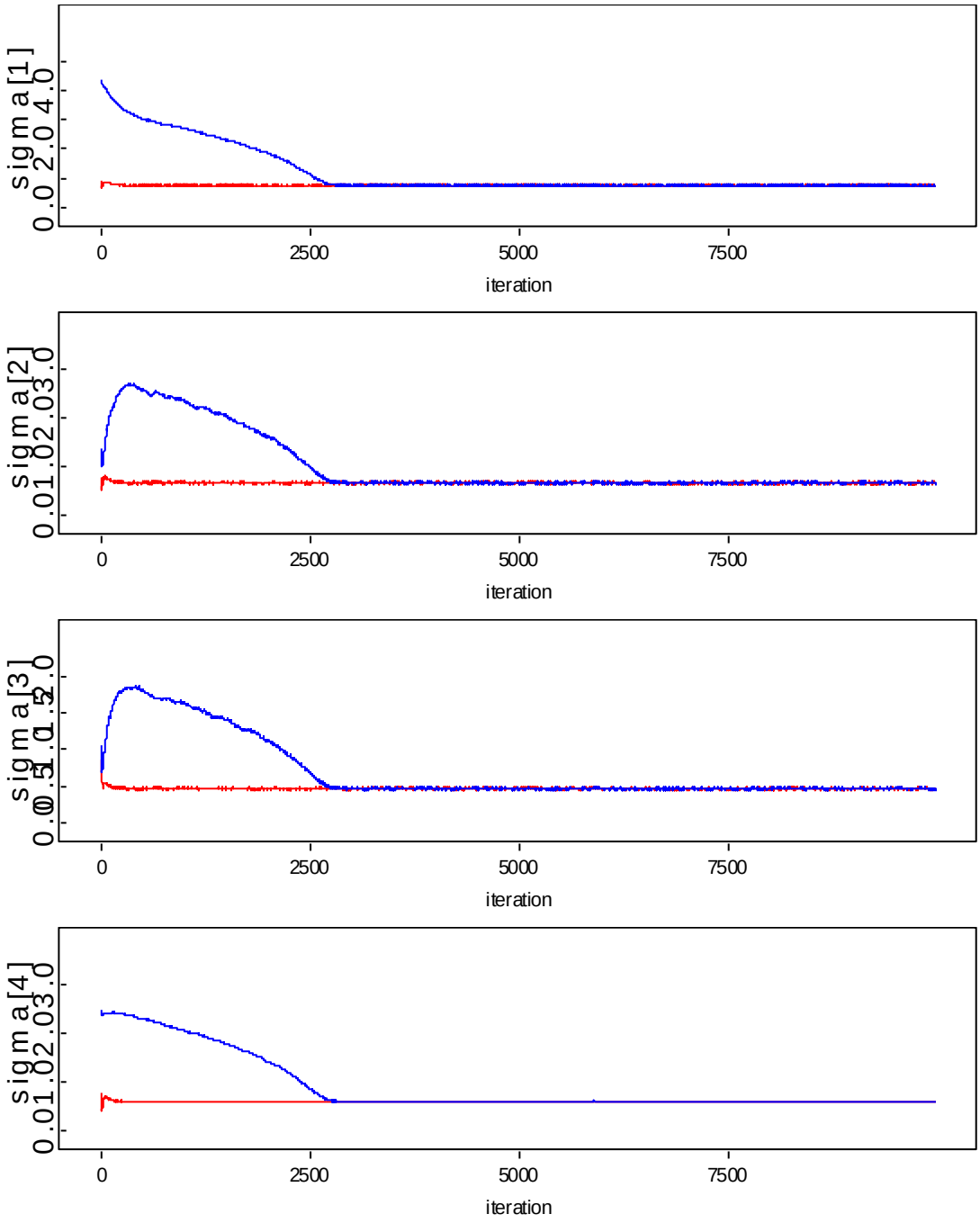

Figure 4. Diagnostic plots of σ_i with rectangular prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

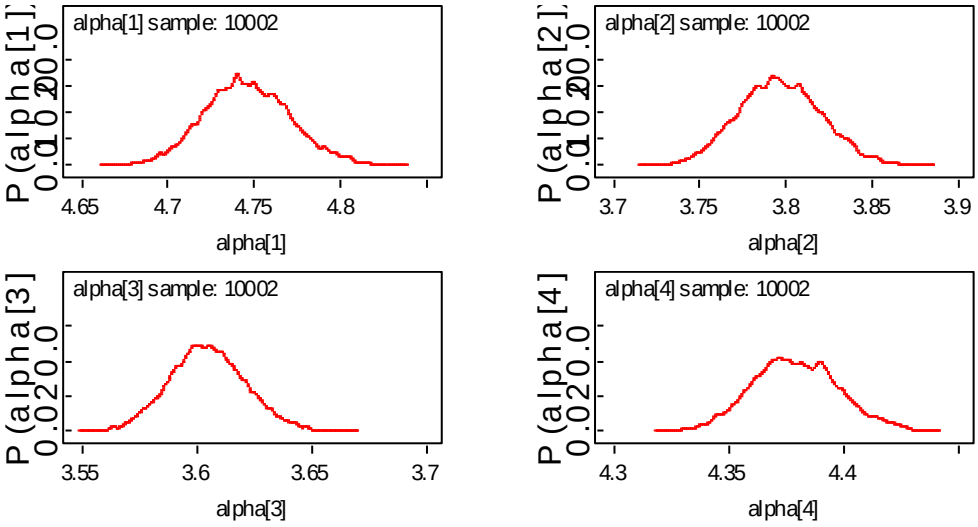

Figure 5. Density plots of α_i with Normal prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

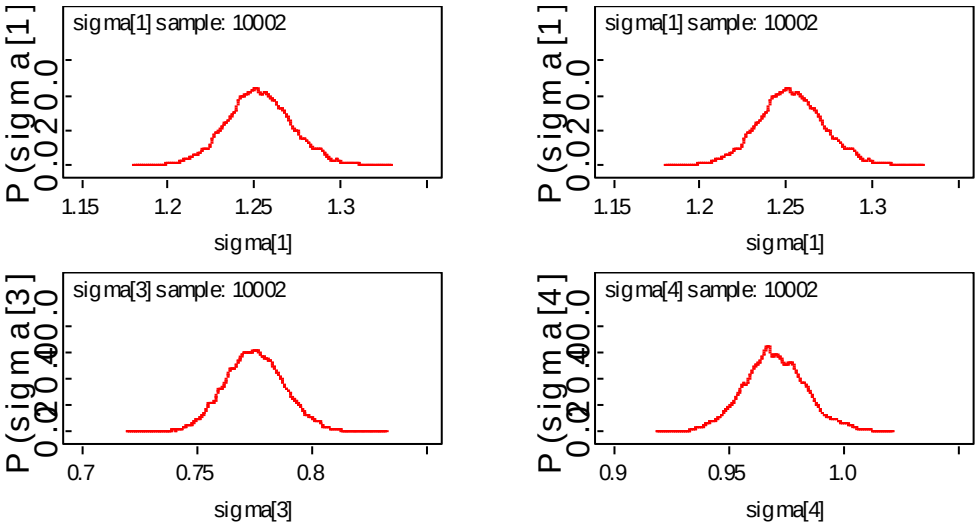

Figure 6. Density plots of σ_i with Normal prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

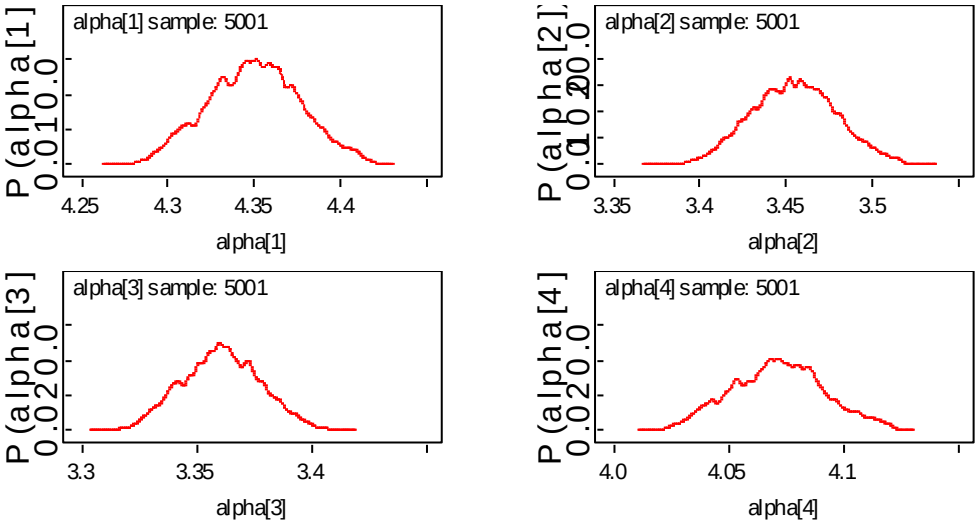

Figure 7. Density plots of σ_i with Normal prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

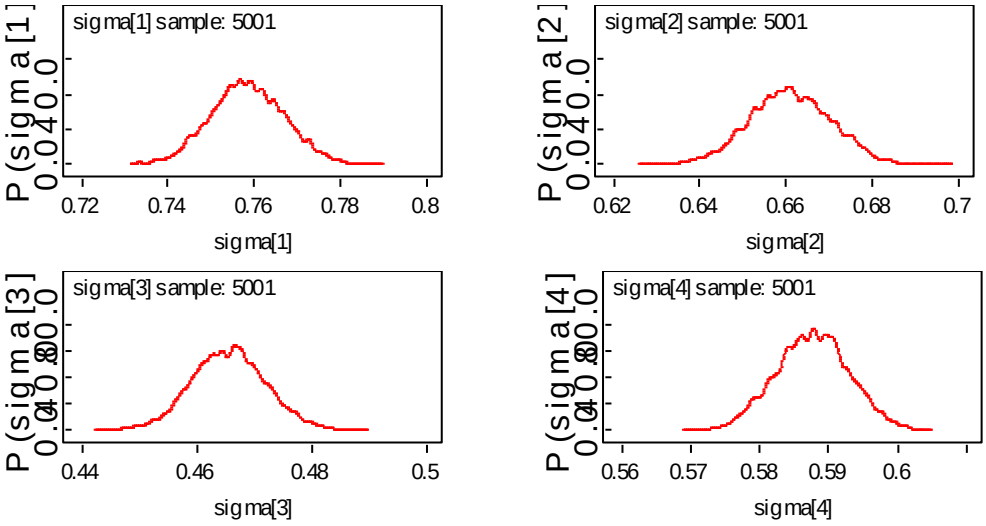

Figure 8. Density plots of σ_i with rectangular prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

TABLE VI

PARAMETER ESTIMATION USING FULL BAYES MODEL

BASED FOR STATIC DATA

Parameter	Full Bayes Estimate (95% Credible Interval) Normal Prior θ_i	Full Bayes Estimate (95% Credible Interval) Rectangular Prior θ_i
α_{AMI}	4.7460 (4.6990, 4.7960)	4.3450 (4.2960, 4.3940)
α_{HF}	3.7960 (3.7510, 3.8440)	3.4510 (3.4060, 3.4950)
α_{PN}	3.6050 (3.5730, 3.6380)	3.3580 (3.3270, 3.3880)
α_{SCIP}	4.3790 (4.3440, 4.4170)	4.0680 (4.0310, 4.1040)
σ_{AMI}	1.2530 (1.2140, 1.2910)	0.7577 (0.7421, 0.7740)
σ_{HF}	1.0890 (1.0510, 1.1280)	0.6603 (0.6414, 0.6795)
σ_{PN}	0.7751 (0.7501, 0.8014)	0.4651 (0.4529, 0.4778)
σ_{SCIP}	0.9699 (0.6058, 0.6602)	0.5871 (0.5775, 0.5972)

Estimates of σ_i are presented in Table VI and the graphs of the posterior distributions for each therapeutic area are presented in Figures 6 and 8 for each model with different assumptions for the prior distribution of θ_i. As seen with estimates of α_i, the estimates of σ_i are lower for the model with the rectangular prior distribution for θ_i than the estimates obtained from the model using the standard normal prior distribution of θ_i. The percent change of the estimates obtained in the model with the rectangular distribution and the estimates obtained from the model with the normal prior of θ_i is consistently 39% for each of the therapeutic areas. Regardless of the model, the amount of discrimination is consistent among the models, i.e., AMI has the highest amount

of discrimination with HF having the second amount of discrimination followed by SCIP and PN has the lowest amount of discrimination.

TABLE VII

PARAMETER ESTIMATION USING EMPIRICAL BAYES

MODEL FOR STATIC DATA

Parameter	Empirical Bayes Estimate (95% CI) Normal Prior θ_i
α_{AMI}	4.7289 (4.6897, 4.7681)
α_{HF}	3.7843 (3.7444, 3.8243)
α_{PN}	3.5993 (3.5717, 3.6269)
α_{SCIP}	4.3700 (4.3406, 4.3994)
σ_{AMI}	0.9881 (0.9660, 1.0102)
σ_{HF}	0.8699 (0.8443, 0.8955)
σ_{PN}	0.6276 (0.6102, 0.6450)
σ_{SCIP}	0.7720 (0.7569, 0.7871)

Results of the EB analysis are presented in Table VII, which includes estimates and 95% confidence intervals. Estimates of α_i obtained from the EB analysis are similar to the estimates found from FB analysis with normal prior for θ_i. Similar to the FB analysis, the estimates from the EB estimates for σ_i follow the same pattern with AMI having the largest estimate along with

the largest discrimination between the four measure sets with all measure-specific estimates being statistically significant based on the respective confidence intervals.

Classification of hospitals is based on the posterior distribution of θ_i. Figure 9 shows the density plots of for each estimate of θ_i. The assumption that θ_i has a standard normal prior yielded a lower variance than the estimate of θ_i with a rectangular prior distribution which had the largest variance of all the estimates. The EB estimate of θ_i was in between the other estimates.

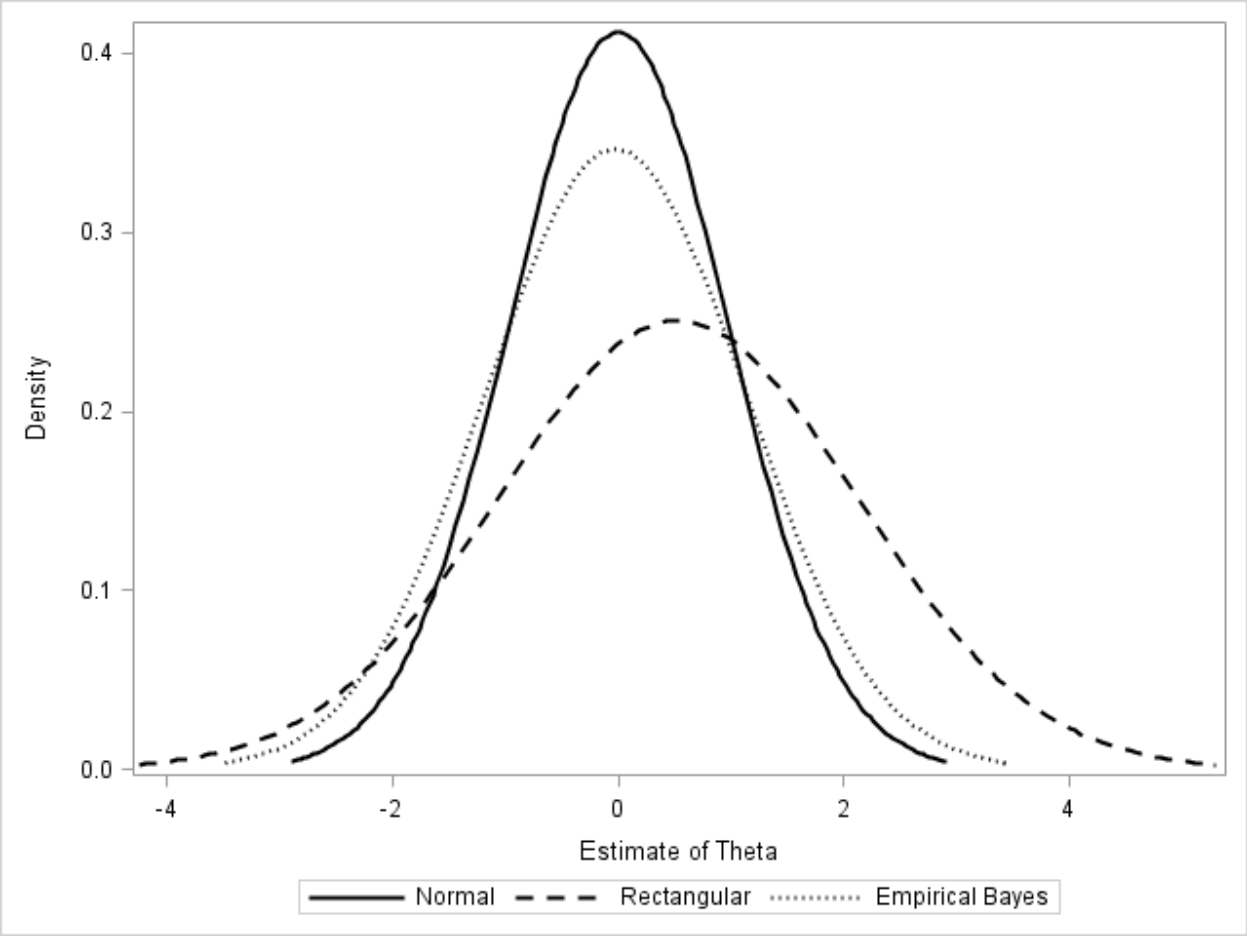

Figure 9. Density plot of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}$ for each prior distribution.

Using estimates of θ_i from the FB models to identify the top 5% of performing hospitals yielded 147 high-performing hospitals. The model assuming a normal prior distribution of θ_i yielded estimates that ranged from 1.47 to 3.21. The observed values of p_i ranged from 99.69% to 100% and a mean overall observed rate of 99.87%. Denominator sizes for each of these observed measures ranged from 682 to 32,628 with a mean number of denominator cases being 5,516. The lengths of the credible intervals obtained from the model with the normal prior

distribution assumption were statistically significantly smaller ($p<.05$) than the lengths of the credible intervals obtained from the model using the rectangular distribution as the prior distribution for θ_i.

TABLE VIII

CCC OF $\widehat{\theta_i}$

$\widehat{\theta_i}$	Full Bayes $\widehat{\theta_i}$ Prior=N(0,1)	Full Bayes $\widehat{\theta_i}$ Prior=U(-4,4)	Empirical Bayes $\widehat{\theta_i}$ Prior=N(0,1)
Full Bayes $\widehat{\theta_i}$ Prior=N(0,1)	1.00		
Full Bayes $\widehat{\theta_i}$ Prior=U(-4,4)	0.6832	1.00	
Empirical Bayes $\widehat{\theta_i}$	0.9642	0.8866	1.00

Estimates, $\hat{\theta}_i$, are obtained from model 4.1 from each FB model utilizing the standard normal prior distribution and the rectangular distribution. Results of calculating Lin's CCC between each FB and EB methods are displayed in Table VIII. The EB estimate and FB estimate using the standard normal prior have the highest CCC of 0.9642. Both estimates obtained from the FB model had the lowest CCC. Figure 10 shows plots comparing each estimate of θ_i. These plots show that the EB estimates were most similar when compared with the FB model with the standard normal prior of $\theta_{i,,}$ which is most likely due to the fact that the EB model automatically assumes a normal prior distribution for the random effects in SAS. Comparing both models that

assumed a normal prior distribution to the model that assumed a rectangular distribution shows that the estimates are similar in the center of the distribution, but get skewed at the tails.

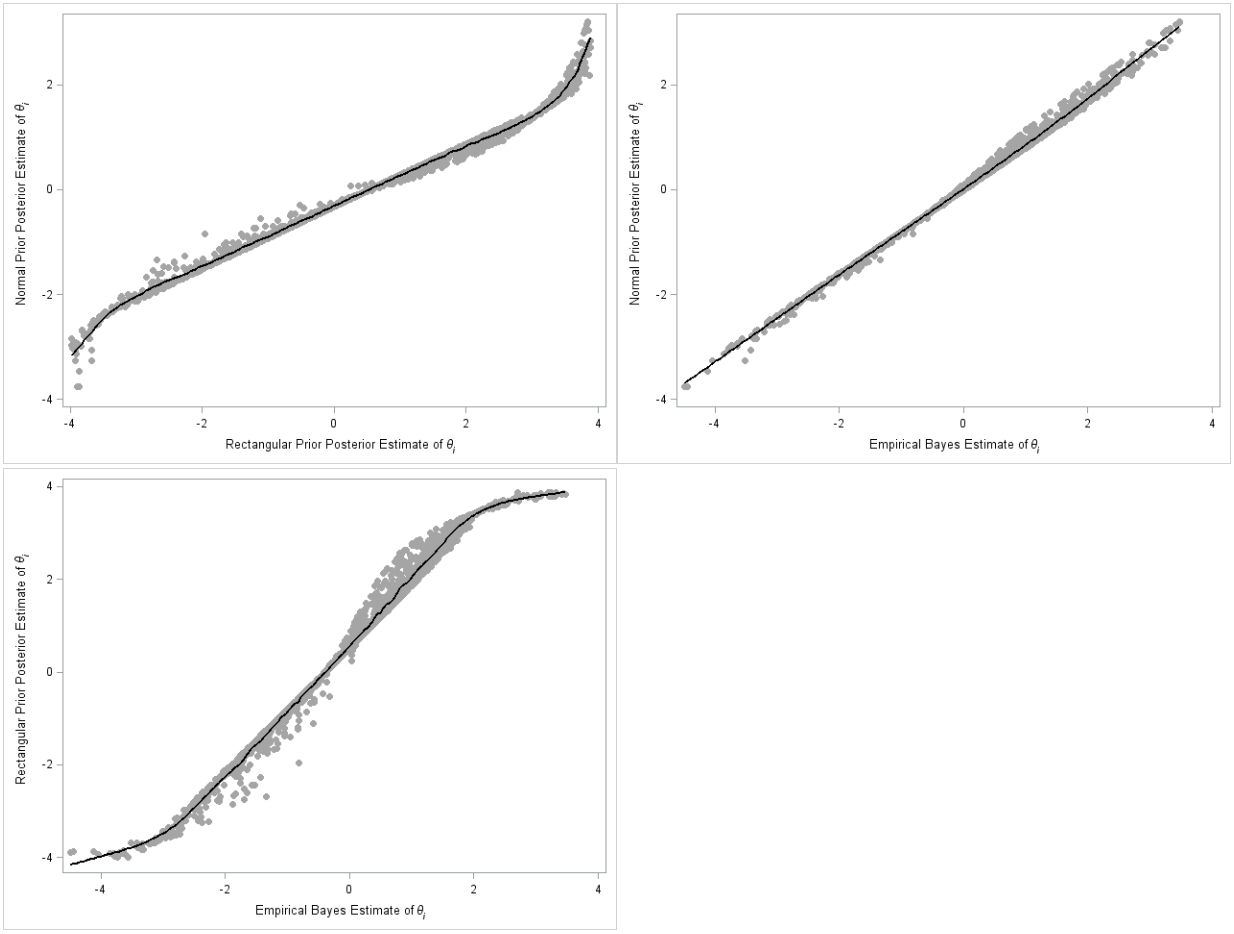

Figure 10. Plots of estimates of $\boldsymbol{\theta_i}$.

Each method classifies the top 5% of hospitals, i.e., 147 hospitals, as top-performing hospitals. When comparing the FB models with different priors, 144 of the same hospitals out of the 147 that were identified in the top 5%. There were three different hospitals that did not match the analysis depending on the prior distribution assumed for θ_i which totaled to six unique hospitals. The differences between the six hospitals was negligible when comparing the overall observed composite rates, p_i. The observed composite rates of the three different hospitals identified as top performers from the model with the standard normal were 99.69%, 99.71%, and 99.76%. The observed overall composite rates of the discrepant hospitals identified from the model with the rectangular distribution were 99.76%, 99.83%, and 99.85%. Figure 11 shows plots of each estimate of θ_i versus the overall observed composite rate p_i along with the Loess line that fits the best locally weighted smoothed curve. In each of the plots, hospitals with low composite rates are associated with lower values of the estimate of θ_i.

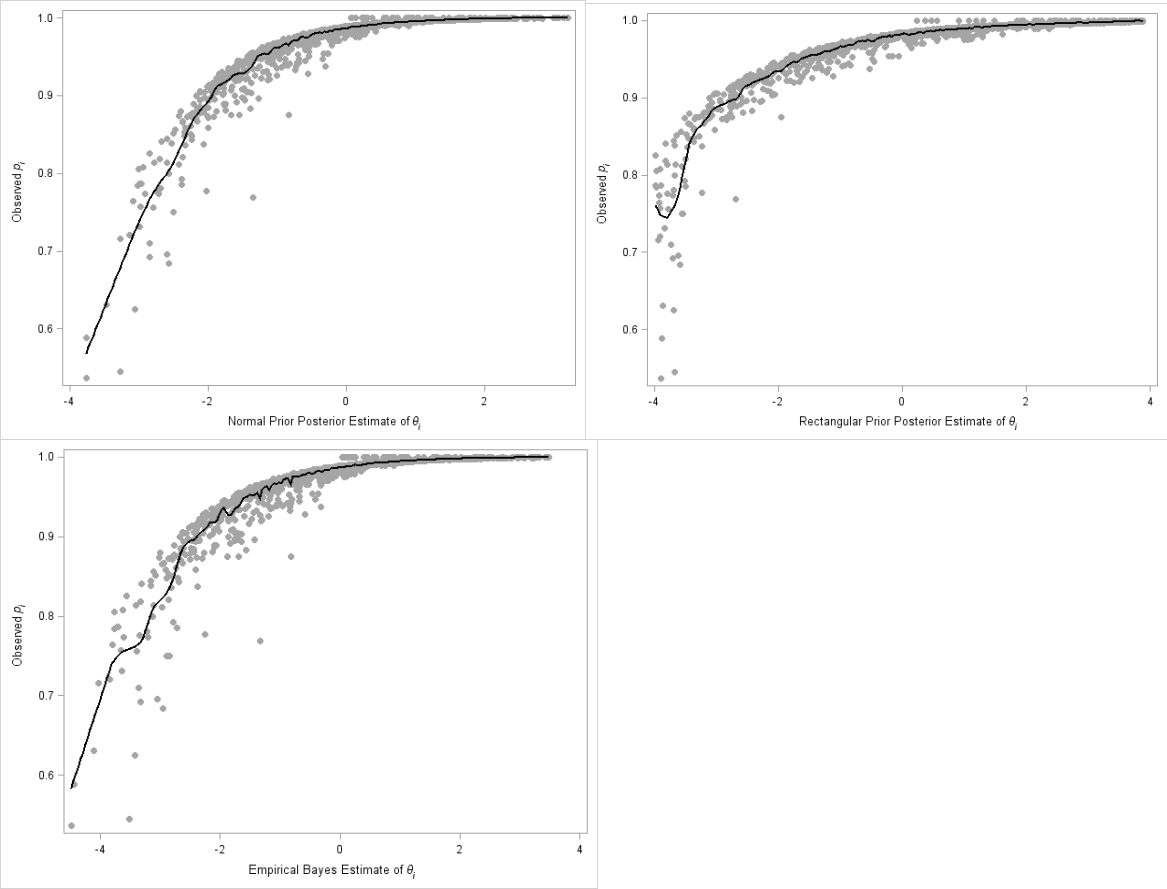

Figure 11. Plots of estimates of $\boldsymbol{\theta_i}$ versus observed rate $\boldsymbol{P_i}$ with Loess line.

Using the method to determine the hospitals performing in the lower 1% based on the estimates of θ_i from the FB model assuming the normal prior distribution yields 29 total hospitals. The mean overall observed rate, p_i, of these 29 hospitals was 74.78% with a standard deviation of 8.75%, a minimum rate of 53.66% and a maximum rate of 85.68%. The 95th percentile of these 29 hospitals was 84.42% indicating that the majority of these hospitals would fall below the 85% rule based on Joint Commission standards for poor-performing hospitals.

Low-performing hospitals identified using the FB model with the rectangular distribution also had low overall observed rates. The mean rate of the 29 hospitals identified by this model was 76.44% with a standard deviation of 7.65%. The 95th percentile of these low-performing hospitals was 85.20% and the 90th percentile was 84.42%. The minimum and maximum rates in this group were the same as found in the model with the normal prior distribution.

There were two hospitals identified as low performers in the model with the normal prior distribution that were not identified by the model using the rectangular prior distribution. Similarly there were two hospitals identified as low performers in the model that used the rectangular prior distribution of θ_i that were not identified in the model assuming a normal prior distribution of θ_i. Each of these four hospitals had rates below 90%. The hospital rates of the low-performing hospitals identified in the model with a normal prior that were not identified in the model with the rectangular prior were 54.54% and 62.50% with denominators of 22 and 32, respectively. Conversely, the overall observed rates of the low-performing hospitals identified in the model with the rectangular distribution but not identified in the model employing the normal prior distribution were 85.2% (n_i=446) and 80.0%(n_i=120).

4.6 **Results Using Complete Data**

Complete data are defined as those organizations that have submitted data in each of the four measure sets and have at least 30 denominator cases in each of the four composite measures. This section shows the results of the model after using complete data within the same framework

of model 4.1. This will show the sensitivity of using models to assess hospital quality with all types of data. The data used for this analysis consist of the same hospital performance measure data used in the previous section for each of the four measure sets or therapeutic areas of AMI, HF, PN, and SCIP. Composite measures are calculated in exactly the same manner as in the previous section.

There were 2,113 hospitals that submitted data for all four measure sets and had at least 30 denominator cases in each of the four therapeutic areas. Table IX shows the descriptive statistics for the measures used in the analysis. The rates for the complete data are similar to the rates seen in section 4.1. The AMI measure had the highest mean rate among the 2,133 hospitals with a rate of 98.16%, while HF and PN had similar rates of 97.00% and 97.06%, respectively. The major difference in the rates of the complete set when compared to the data without restrictions is the bottom part of the distribution. The minimum rates of hospitals using all the data for AMI, HF, and PN were 0.0 compared to the complete data that had rates of 54.05%.

TABLE IX

OBSERVED MEASURE SET COMPOSITE AND OVERALL COMPOSITE RATES OF STATIC YEAR (3Q2012–2Q2013)—COMPLETE DATA

Measure Set	N	Mean of Rates	Standard Deviation	Max	90th Percentile	Median	10th Percentile	Min
AMI	2,133	0.9816	0.0328	1.0000	1.0000	0.9921	0.9545	0.5405
HF	2,133	0.9700	0.0468	1.0000	1.0000	0.9884	0.9139	0.4736
PN	2,133	0.9706	0.0338	1.0000	1.0000	0.9779	0.9380	0.5961
SCIP	2,133	0.9843	0.0193	1.0000	0.9969	0.9889	0.9693	0.7333
Overall Composite	2,133	0.9831	0.0190	1.0000	0.9964	0.9879	0.9670	0.7742

The correlations of each of the measure sets are displayed in Table X. Although all correlation coefficients are significantly different from zero, the measure set composite rates are not highly correlated with the highest correlation coefficient being 0.5195 between the PN and SCIP measure sets. The lowest correlation is between the HF and PN therapeutic areas with a correlation coefficient value of 0.3727.

TABLE X

PEARSON CORRELATION COEFFICIENTS

Pearson Correlation Coefficients, N = 2,113 Prob > |r| under H0: Rho=0				
	AMI	HF	PN	SCIP
AMI	1			
HF	0.49058 <.0001	1		
PN	0.41759 <.0001	0.37267 <.0001	1	
SCIP	0.51155 <.0001	0.4049 <.0001	0.51951 <.0001	1

Estimating convergence of model 4.1 with the complete data is achieved graphically. Figures 12 and 13 show the convergence of α_i and σ_i, respectively, using a normal prior distribution of θ_i. Convergence of α_i occurs for each measure set prior to 2,500 draws from the posterior distribution. Similarly for σ_i, convergence also occurs prior to 2,500 iterations of model 4.1. Density plots of the posterior distribution of each α_i and σ_i, are show in Figures 15 and 16. The DIC for the model with the normal prior is 71,815.6 whereas the DIC for the rectangular prior model is slightly lower at 69,102.5, indicating that the rectangular prior fits the data slightly better than the model using the normal prior distribution.

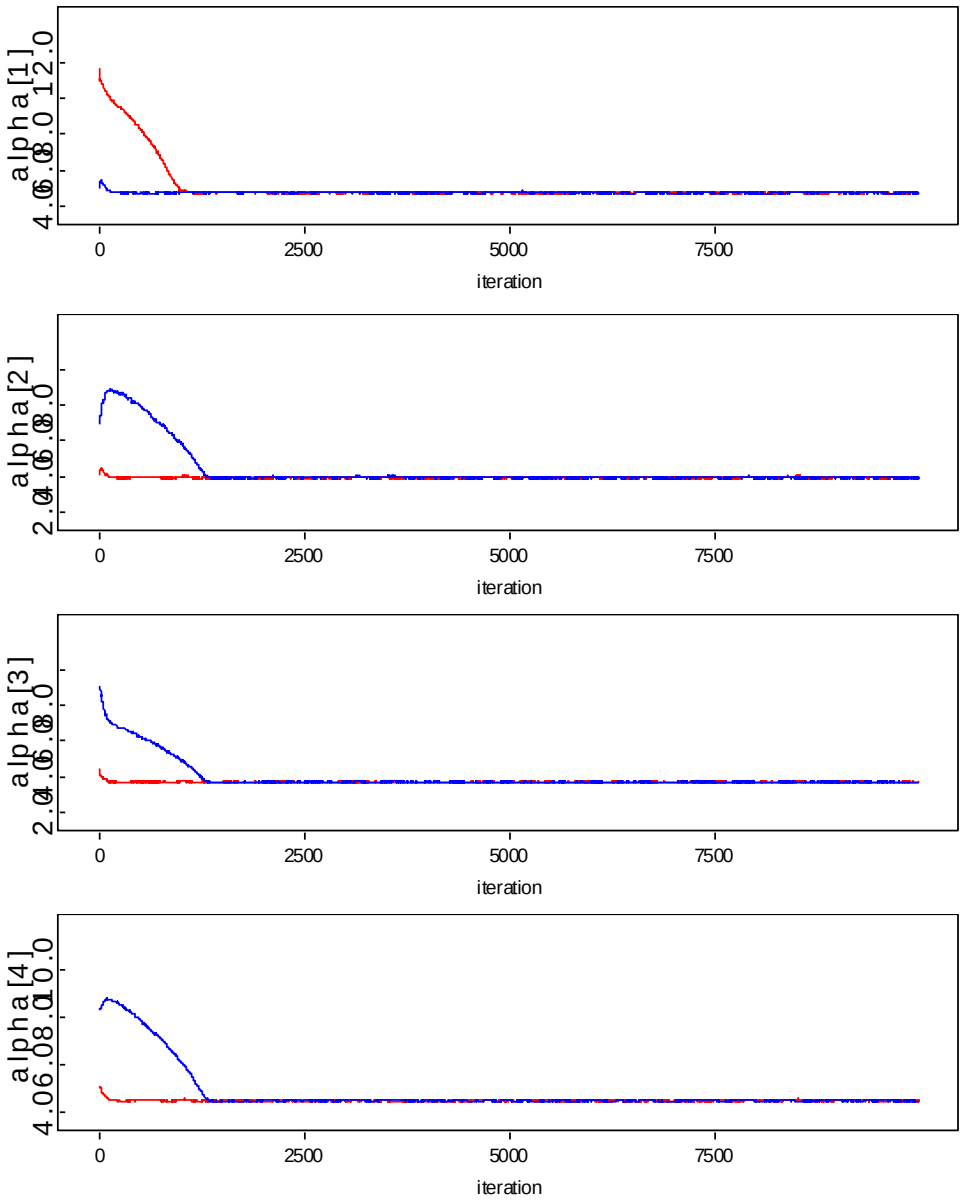

Figure 12. Diagnostic plots of α_i with Normal prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

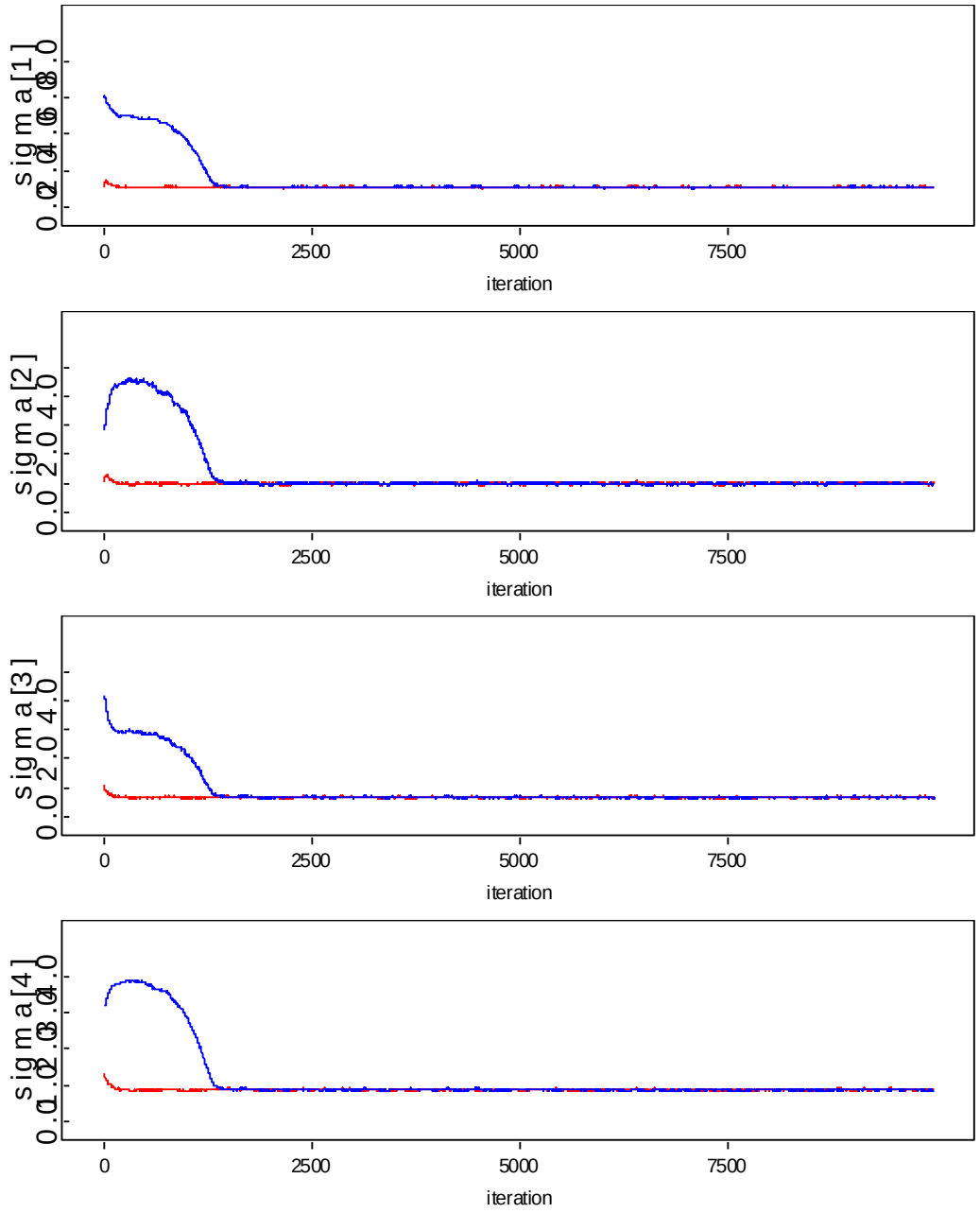

Figure 13. Diagnostic plots of σ_i with Normal prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

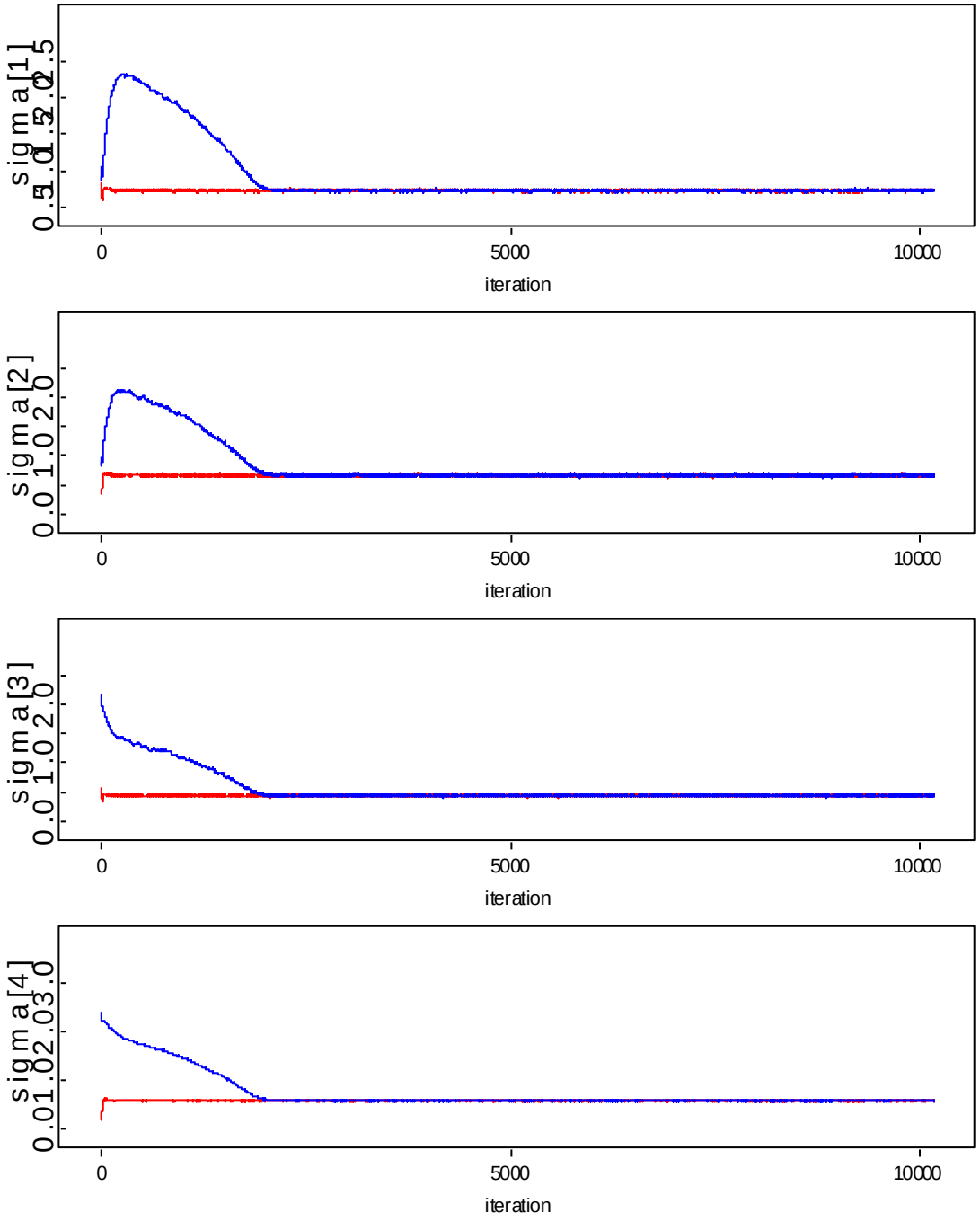

Figure 14. Diagnostic plots of σ_i with rectangular prior of $\boldsymbol{\theta_i}$.
1=AMI, 2=HF, 3=PN, 4=SCIP

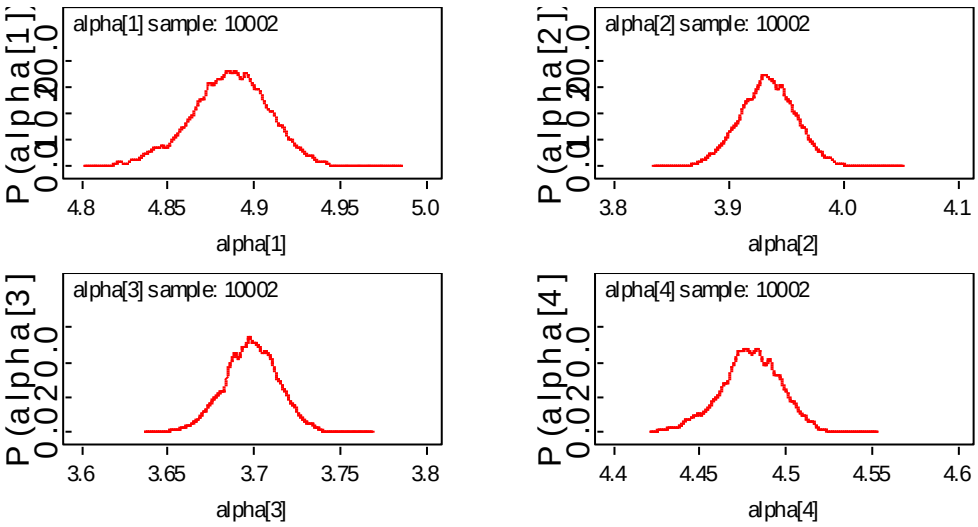

Figure 15. Density plots of α_i with Normal prior of $\boldsymbol{\theta_i}$—complete data.
1=AMI, 2=HF, 3=PN, 4=SCIP

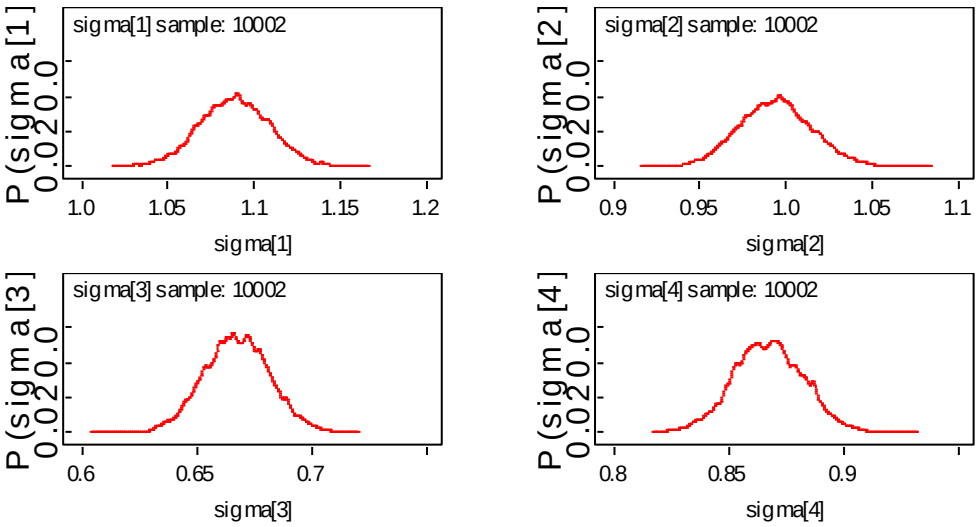

Figure 16. Density plots of σ_i with Normal prior of $\boldsymbol{\theta_i}$—complete data.
1=AMI, 2=HF, 3=PN, 4=SCIP

TABLE XI

PARAMETER ESTIMATION USING FULL BAYES MODEL

BASED ON COMPLETE STATIC DATA

Parameter	Full Bayes Estimate (95% Credible Interval) Normal Prior θ_i	Full Bayes Estimate (95% Credible Interval) Rectangular Prior θ_i
α_{AMI}	4.8859 (4.8379, 4.9293)	4.4173 (4.3609, 4.4711)
α_{HF}	3.9339 (3.8850, 3.9811)	3.5089 (3.45297, 3.5614)
α_{PN}	3.6982 (3.6666, 3.7287)	3.4116 (3.37509, 3.4462)
α_{SCIP}	4.4793 (4.4420, 4.5125)	4.1039 (4.05989, 4.1455)
σ_{AMI}	1.0885 (1.0487, 1.1282)	0.7253 (0.7083, 0.7422)
σ_{HF}	0.9938 (0.9536, 1.0369)	0.6625 (0.6422, 0.6827)
σ_{PN}	0.6672 (0.6391, 0.6953)	0.4377 (0.4238, 0.4518)
CIP	0.8672 (0.8372, 0.8968)	0.5781 (0.5668, 0.5894)

Estimates of the model parameters are based on 5,000 simulations after convergence is achieved for both chains and are shown in Table XI for the model assuming a standard normal prior distribution of θ_i and the model assuming a rectangular (-4,4) distribution of θ_i. Using a uniform (-4,4) distribution was used because it encompassed the same values of θ_i obtained from the normal prior. Also, using a range greater than (-4,4) had issues with convergence of the estimates. As seen with the full data set, the direction of discrimination based on σ_i is in the same direction with AMI having the highest amount of discrimination and PN having the lowest amount of discrimination based on the model with the standard normal prior. Estimates of α_i follow a similar trend with the largest of estimates being with AMI, then the next highest

estimate being for SCIP, with HF and PN having similar lower estimates. The estimates of α_i and σ_i based on the model with the assumption of a rectangular distribution show similar patterns although the estimates obtained are slightly lower than the model assuming a normal prior distribution of θ_i. Parameter estimates from the EB model are presented in Table XII. All estimates are similar to those found in the FB models. The estimates for α_i are very similar to those found in the FB model with the assumption of a normal prior distribution of θ_i. Estimates of σ_i fall between the estimates of the FB models although the trend of discrimination is the same.

TABLE XII

PARAMETER ESTIMATION USING EMPIRICAL BAYES MODEL FOR STATIC DATA—COMPLETE DATA

Parameter	Empirical Bayes Estimate (95% CI) Normal Prior
α_{AMI}	4.8672 (4.8263, 4.9082)
α_{HF}	3.9189 (3.875, 3.9629)
α_{PN}	3.6919 (3.6628, 3.721)
α_{SCIP}	4.4678 (4.4364, 4.4992)
σ_{AMI}	0.8814 (0.859, 0.9039)
σ_{HF}	0.8096 (0.7828, 0.8364)
σ_{PN}	0.5488 (0.5302, 0.5674)
σ_{SCIP}	0.7064 (0.6903, 0.7224)

Estimates of θ_i are obtained and a sample of posterior density plots of θ_i for four randomly selected hospitals is presented in Figures 17 and 18. The posterior densities for these hospitals are approximately normally distributed centered around the mean of each estimate. For hospital 1 in this sample, the mean value of $\widehat{\theta_i}$ is 0.5098, the mean value of $\widehat{\theta_i}$for hospital 2 is 0.229, the estimate for hospital 3 is 0.0223 and the mean value of the estimate is -0.3189. Estimates of θ_i with the rectangular prior distribution for the same hospitals are higher with values 1.4749, 0.9991, 0.6172, and 0.1492 for hospitals 1 through 4.

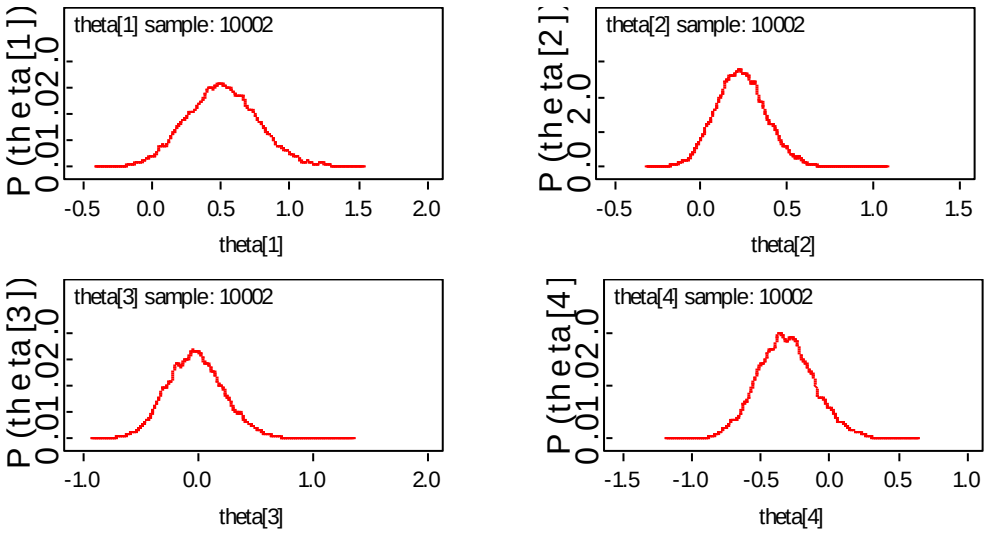

Figure 17. Density plots of $\widehat{\boldsymbol{\theta}_i}$ with Normal prior of $\boldsymbol{\theta}_i$ for hospitals 1,2,3,4.

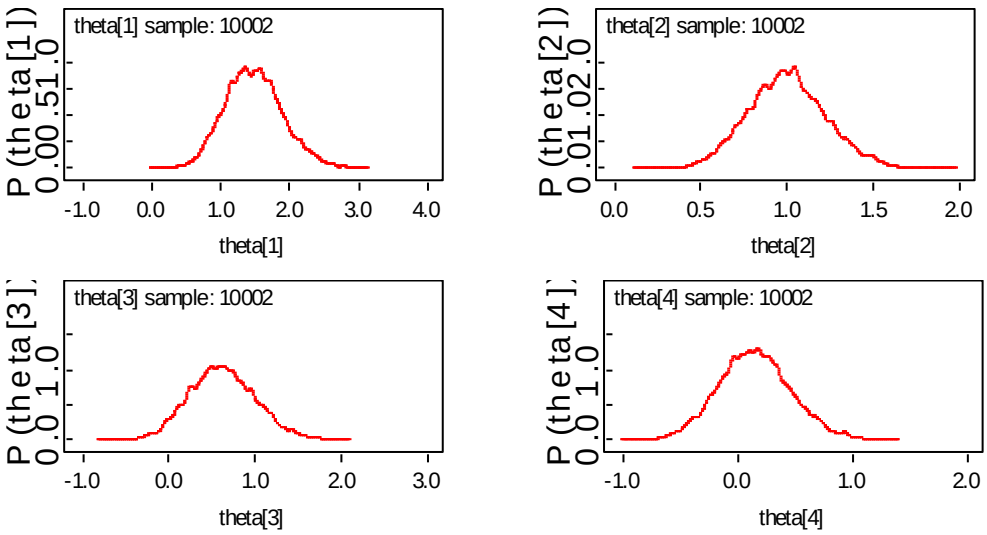

Figure 18. Density plots of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}$ with rectangular prior of $\boldsymbol{\theta}_{\boldsymbol{i}}$ for hospitals 1,2,3,4.

Evaluation of hospital quality is based on the estimates of θ_i obtained from each of the two FB models. The top-performing hospitals are based on the upper 5% of the posterior distribution of θ_i. The density plots of $\widehat{\theta}_i$ obtained in each of the three models are graphically shown in Figure 19 and these figures show each to be approximately normally distributed. Both models that assumed a normal prior had a mean value of $\widehat{\theta}_i$ to be close to 0 with standard deviations close to 1 and the mean of $\widehat{\theta}_i$ obtained by the model with the rectangular distribution was close to 0.6 with a standard deviation of 1.5. The CCC between the $\widehat{\theta}_i$ obtained from both FB models with normal and rectangular prior was 0.7908. The CCC of $\widehat{\theta}_i$obtained from the EB model and the FB model with the normal prior assumption was 0.9694 and the corresponding CCC of $\widehat{\theta}_i$ between the EB and the FB model with the rectangular prior was 0.9349.

Each model identified 105 top-performing hospitals and both models had identified 101 of the same hospitals in the upper 5% of the posterior distribution. There were four hospitals identified in each FB model (normal and rectangular prior) that differed. The mean observed value of p_i of the top-performing hospitals identified in the model with the normal prior distribution was 0.9987 ranging from 0.9971 to 1.00 with $\widehat{\theta_i}$ ranging from 1.59 to 3.32. From the model with the rectangular prior distribution, the mean observed value of p_i was 0.9987 ranging from 0.9975 to 1.00 with $\widehat{\theta_i}$ ranging from 3.27 to 3.88.

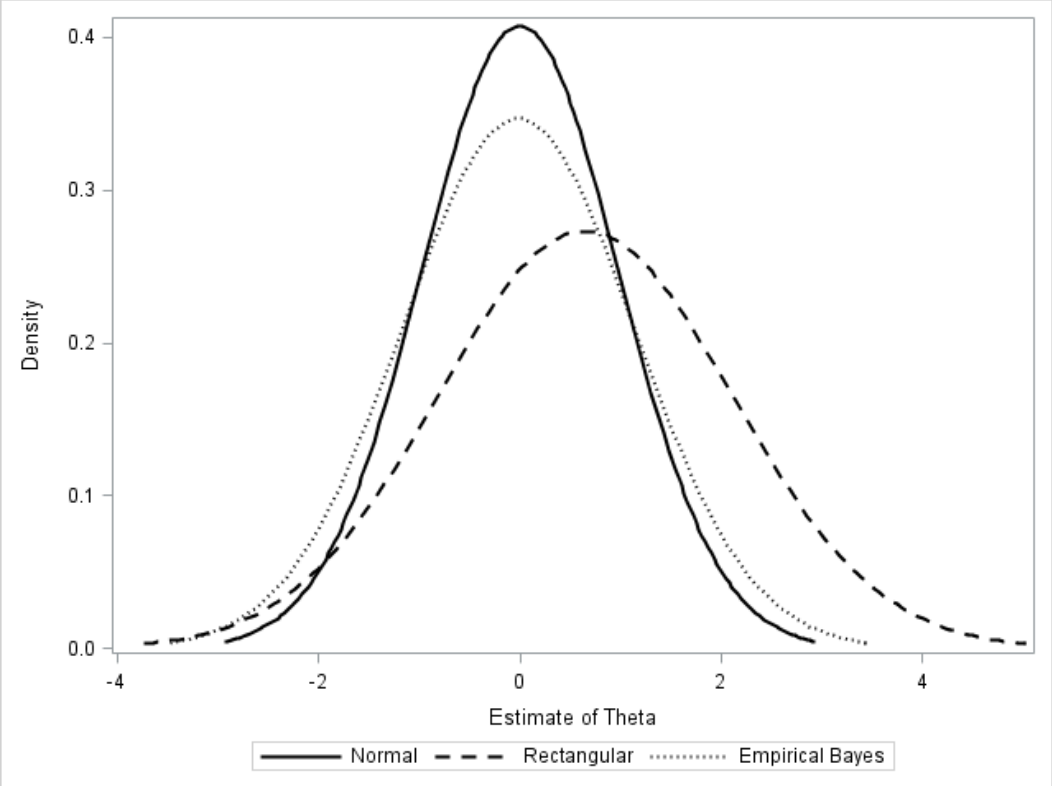

Figure 19. Density plot of $\widehat{\boldsymbol{\theta}_{\boldsymbol{i}}}$ for each prior distribution.

There were four hospitals identified as top performers from each FB model with differing prior distributions that did not match making a total of eight differing hospitals. Each of these eight hospitals had observed values of π close to the lower range of observed π identified above. Similarly, the values of $\widehat{\theta_i}$are on the boarder of the lower end of the top performers identified above.

Low-performing hospitals were identified in a similar way. There were 21 hospitals identified as low performers based on $\widehat{\theta_i}$. The mean of $\widehat{\theta_i}$ obtained from the rectangular prior distribution was -3.54 ranging from (-2.85, -3.99) with observed rate, p_i, having a mean value of 0.8574 ranging from (0.7742, 0.9141). From the model with the normal prior distribution assumption, $\widehat{\theta_i}$as a mean of -2.84 ranging from -2.33 to -3.51 with a mean observed p_i of 0.8574 distributed between 0.7742 and 0.9141. In this instance, the same hospitals were identified as low performers regardless of which prior distribution was assumed.

4.7 **Sensitivity of Models and Model Selection**

This section identifies the sensitivity of the models based on the identification of top-performing hospitals and estimates obtained from both (normal and rectangular prior) FB models. As shown, the parameter estimates of α_i are similar in each model. Estimates of σ_i are lower in the model with the complete data indicating a lower amount of discrimination. This is due to the nature of having complete data and not needing to account for missing data, which is expected. In comparing the identification of high- and low-performing hospitals, regardless of

which assumption on the prior distribution was made, the exact same hospitals identified in the complete data (section 4.2) were also identified as either top- or low-performing hospitals in the analysis using all the available data (section 4.1).

Apart from evaluating models based on identification of top-performing hospitals there are differences for each of the different approaches. Under the FB approach, the assumption of the prior distribution for the latent variable has different effects on the model. Using a non-normal distribution that has a wider variance affects the model fit. Larger variances on the prior increases the goodness of fit statistics based on the deviance function, which has an effect of potentially misclassifying hospitals, especially when evaluating poor-performing hospitals.

Using the EB approach is similar to the FB approach with the assumption of the standard normal distribution for the prior of the latent variable. As of this writing, SAS is only capable of using the standard normal distribution on the prior in the calculation of the EB estimates. Although the EB confidence intervals for the parameter estimates were similar to the FB credible intervals and the identification of top-performing hospitals is similar, the interpretation of the EB estimates are slightly different. For the EB analysis, the interpretation of the latent variable is not based on probability but based on one-sided confidence intervals. That is, if an individual hospital is classified to be in the top 5% of all hospitals, then this is a true top-performing hospital with 95% confidence.

4.8 Longitudinal Analysis

4.8.1 Composite Scores

I consider a longitudinal latent variable model to determine the amount of improvement, or lack thereof, a hospital has displayed over time for eight time periods. A composite rate is calculated for each of the four measure sets using only accountability measures as defined in section 3.1 for each quarter starting in the third quarter of the year 2011 to the second quarter of 2013. A hospital may have data in up to eight quarters for any given measure set and no restrictions on sample size is required.

Similar to the definitions stated for the static analysis, y_{ijkt} is a binomial variable that represents the number of patients that receives the k^{th} treatment in the j^{th} therapeutic area at the i^{th} hospital at the t^{th} time point. Similarly, let n_{ijkt} be the number of eligible patients to receive the k^{th} therapy in the j^{th} clinical area at hospital i for the t^{th} time period. The number of clinical areas is denoted by J_i that ranges 1 to 4 clinical areas depending on the services provided by the i^{th} hospital. Let the number of therapies in the j^{th} clinical area denoted by K_{ij} be 7 for AMI, 1 for HF, 3 for PN, and 6 for SCIP in the i^{th} hospital for the t^{th} time period. Let $y_{it} = \sum_j^{J_i} \sum_k^{K_{ij}} y_{ijkt}$ be the total number of times appropriate therapies were provided in all the clinical areas in the i^{th} hospital in the t^{th} time period. Similarly, let $n_{it} = \sum_j^{J_i} \sum_k^{K_{ij}} n_{ijkt}$ be the total number of opportunities to provide appropriate therapies in all the clinical areas at time period t in the i^{th} hospital. The overall observed composite rate of providing appropriate therapies in all clinical areas at time period t in the i^{th} hospital is defined as $p_{it}^{obs} = y_{it}/n_{it}$.

4.8.2 **Longitudinal Regression Model**

I extend the yearly latent variable model to incorporate changes longitudinally. Within the framework of the new model, the latent variable θ_{0i} represents the quality of care in the i^{th} hospital where larger values of θ_{0i} are associated with a higher quality of care. I introduce a new latent variable θ_{1i} associated with the time effect to represent the quality of performance improvement over time where larger values of θ_{1i} are associated with greater performance improvement efforts over time. Let α_{0j} be the overall mean rate on the log-odds scale for the j^{th} clinical area and α_{1j}be the average change from the overall rate at each time point (i.e., α_{0j} is the average intercept and α_{1j} is the average slope for the j^{th} area on the log-odds scale); σ_{0j} be a positive measure set-specific discrimination. Let p_{ijt} be the probability that the proper therapy was given within the j^{th} clinical area at the i^{th} hospital in the t^{th} time period.

The proposed longitudinal model is a random slope/random intercept model and is defined as follows:

$$logit(p_{ijt}) = h_{ij}(\alpha, \sigma, \beta, \theta, t), \tag{4.4.1}$$

where

$$h_{ij}(\alpha, \sigma, \beta, \theta, t) = (\alpha_{0j} + \sigma_{0j}\theta_{0i}) + (\alpha_{1j} + \theta_{1i}) * t$$

and $\begin{pmatrix}\sigma_{0j}\\ \sigma_{1j}\end{pmatrix} > \begin{pmatrix}0\\0\end{pmatrix}$ and

$\begin{pmatrix}\theta_j\\ \theta_{jt}\end{pmatrix} \sim MVN(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix})$.

Hence

$$\Pr(y_{ijt} = 1) = p_{ijt} = \frac{\exp(h_{ij}(\alpha,\sigma,\beta,\theta,t))}{1+\exp(h_{ij}(\alpha,\sigma,\beta,\theta,t))}. \quad (4.4.2)$$

4.8.3 **Longitudinal Classification**

From the model 4.4.1, a measure of each hospital's quality improvement effort for each measure set or clinical area is computed using the coefficients associated with the time variable, i.e. $\alpha_{1j} + \theta_{1i}$. In this situation the score $\widehat{\theta_{1i}}$ is of interest, which is derived from the posterior distribution of θ_{1i}. Assuming that national rates are positively increasing, i.e. $\alpha_{1j} > 0$, and if $\alpha_{1j} + \theta_{1i}$ is less than 0, then the hospital is classified as having a decrease in their quality improvement initiates over time for the measure set j which implies that θ_{1i} must be a negative quantity. This is equivalent of computing $P(\alpha_{1j} + \theta_{1i} < 0) > \vartheta$ for each j^{th} measure set for the i^{th} hospital where ϑ is a predefined threshold related to the degree of certainty. Performing some simple math, $P(\theta_{1i} < -\alpha_{1j}) > \vartheta$ is obtained. For example, if the threshold of certainty, ϑ, is decided to be 95%, a 90% credible interval of the posterior distribution of the score estimate, S_{1i}, is calculated based on θ_{1i} and compared to the upper bound of the credible interval to $-\alpha_{1j}$. If the upper bound is less than this quantity, then there is at least 95% of the distribution below this ratio and conclude that the hospital is declining in their quality initiatives for this measure set. Conversely, hospitals with ongoing quality improvement efforts will be identified by calculating $P(S_{1i} > -\alpha_{1j}) > \vartheta$. Determining the lower 90% credible interval of the posterior

distribution of θ_{i1} and comparing that to $-\alpha_{1j}$ as before for each measure set will satisfy this probability taking ϑ to be 0.95.

4.8.4 **Longitudinal Results**

There were 2,977 hospitals included in the longitudinal study that had at least one quarter of data in the AMI, HF, PN, or SCIP measure sets with a total of 85,220 total quarterly data points. A majority (85.0%) of the hospitals had data in all four measure sets, while 302 (10.1%) hospitals had data in three measure sets with 77 (2.6%) of hospital had data in two measure sets and 66 (2.3%) of hospitals had longitudinal data in only one measure set.

Descriptive statistics are displayed in Table XIII. All measure sets had minimum quarterly rates of 0.0, although the majority of these low rates had small denominators (<5) and all had maximum rates of 1.00 and all measure sets had average observe rates about 95%. All four measure sets show an average increase over time as shown in Figure 20. The AMI had the largest mean increase from the 3Q11 to 2Q13 with an average increase of 1.12% and PN had the lowest mean increase of 0.7% over the eight-quarter time interval.

TABLE XIII

DESCRIPTIVE STATISTICS BY MEASURE SET OVER TIME

	2011Q3	2011Q4	2012Q1	2012Q2	2012Q3	2012Q4	2013Q1	2013Q2
AMI								
-n	2,457	2,473	2,510	2,469	2,424	2,430	2,448	2,434
-Mean	0.9665	0.9722	0.9703	0.9737	0.9759	0.9763	0.9765	0.9777
-Max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-Median	0.9947	0.9954	0.9957	0.9966	0.9974	0.9970	0.9973	1.0000
-Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HF								
-n	2,702	2,720	2,718	2,706	2,656	2,682	2,698	2,687
-Mean	0.9586	0.9609	0.9570	0.9589	0.9621	0.9655	0.9630	0.9657
-Max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-Median	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PN								
-n	2,803	2,813	2,819	2,820	2,764	2,773	2,784	2,788
-Mean	0.9618	0.9600	0.9626	0.9651	0.9641	0.9649	0.9637	0.9633
-Max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-Median	1.0000	0.9857	0.9833	1.0000	1.0000	1.0000	0.9828	1.0000
-Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SCIP								
-n	2,697	2,697	2,724	2,731	2,686	2,694	2,700	2,713
-Mean	0.9753	0.9761	0.9779	0.9788	0.9812	0.9814	0.9842	0.9846
-Max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-Median	0.9830	0.9850	0.9861	0.9872	0.9881	0.9887	0.9908	0.9914
-Min	0.5000	0.4643	0.3636	0.5000	0.7143	0.0000	0.6000	0.5000

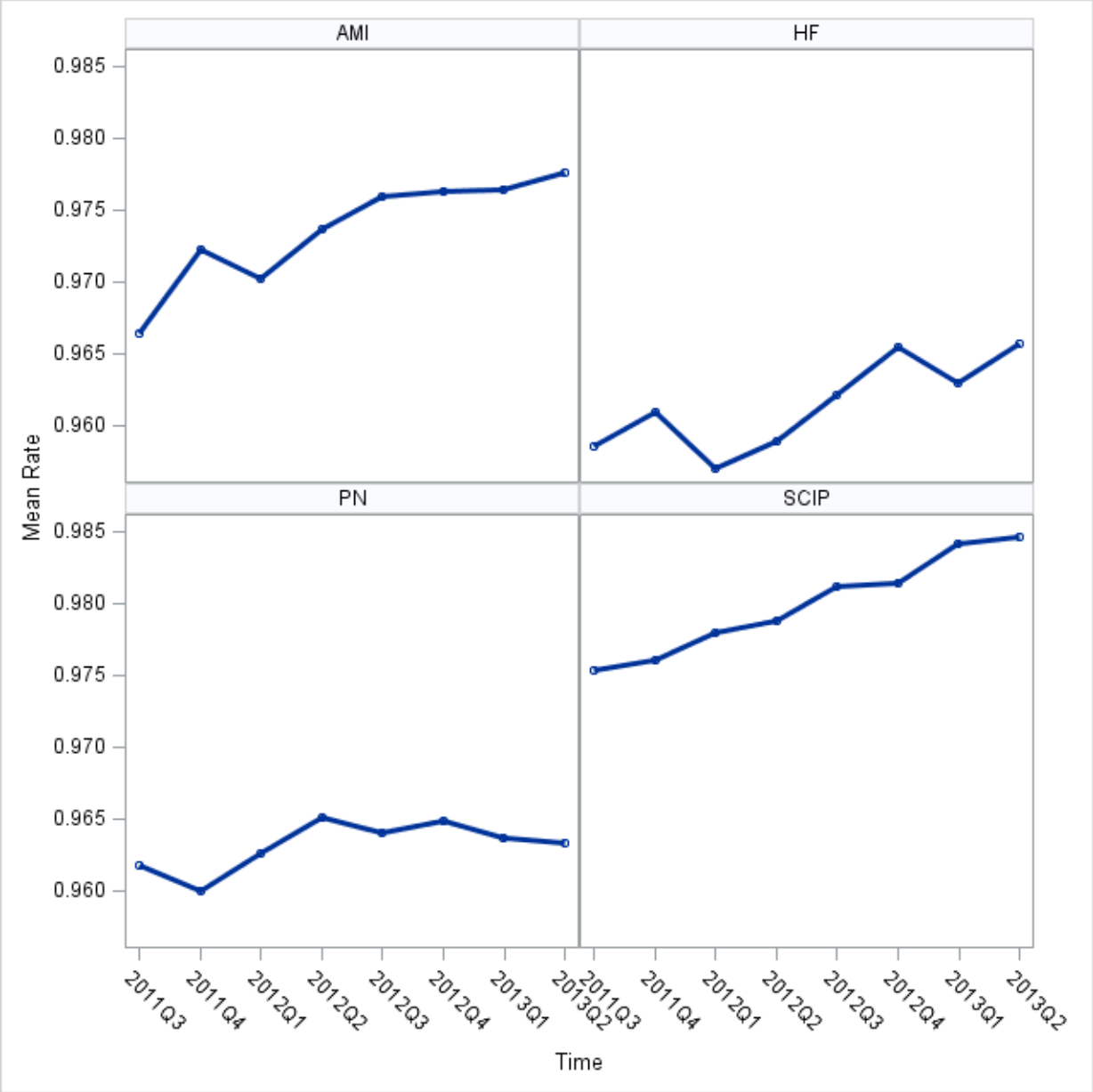

Figure 20. Mean observed rates by measure set over time.

Parameters of model 4.4.1 are estimated by using an FB approach in OpenBUGS by using two chains with different initial values. Convergence after 2,500 burn-in samples of the variance parameter, σ_{0j}, is assessed graphically in Figure 21. The random slope parameter, α_{1J}, converges slower than the variance parameter and converges after 10,000 burn-in samples and is shown in Figure 22. Additional assessment is evaluated using the Gelman-Rubin statistic and convergence is shown for each estimated variance parameter in Figure 23.

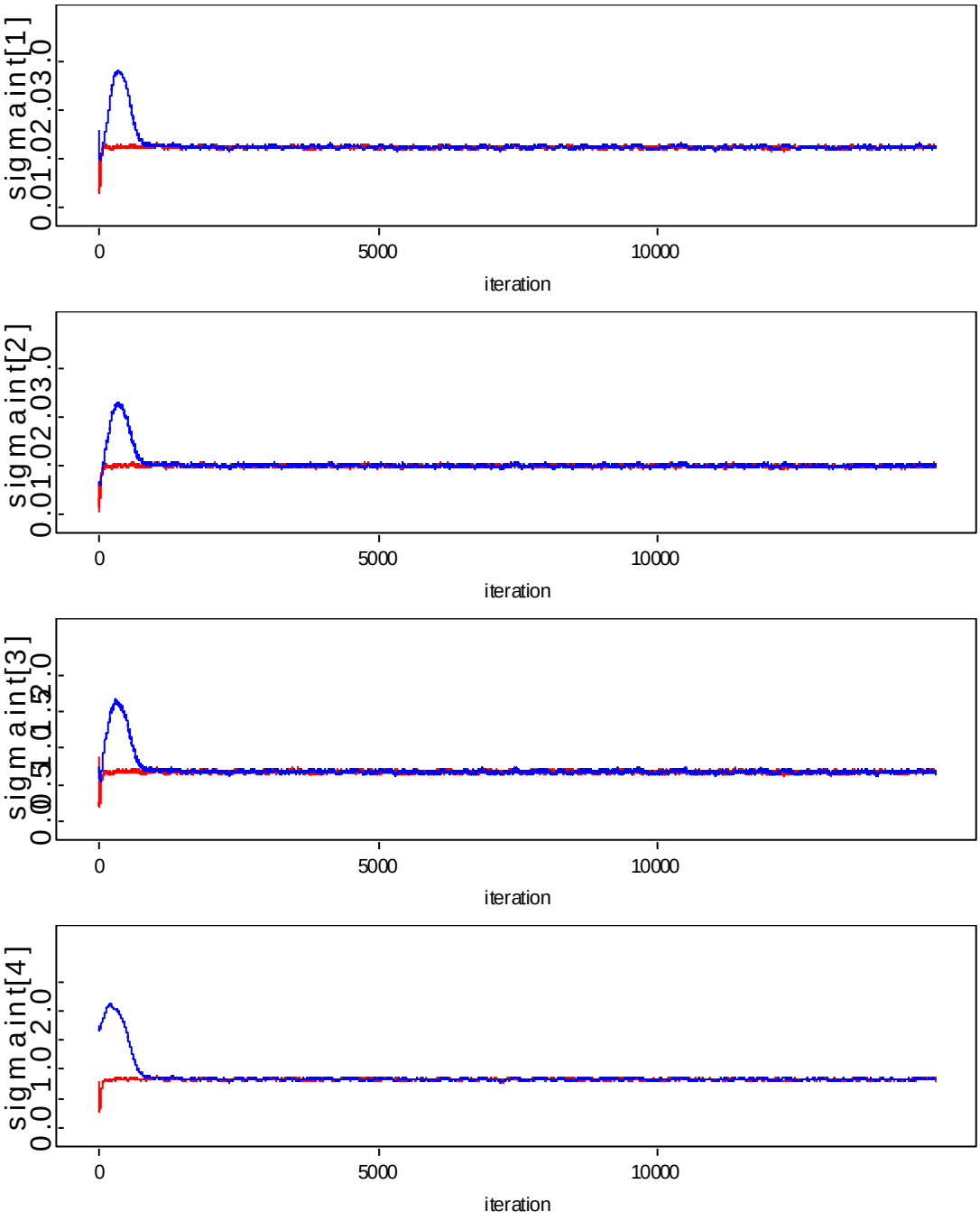

Figure 21. Diagnostic plots of convergence for random intercept variance, σ_{0j}.
1=AMI, 2=HF, 3=PN, 4=SCIP

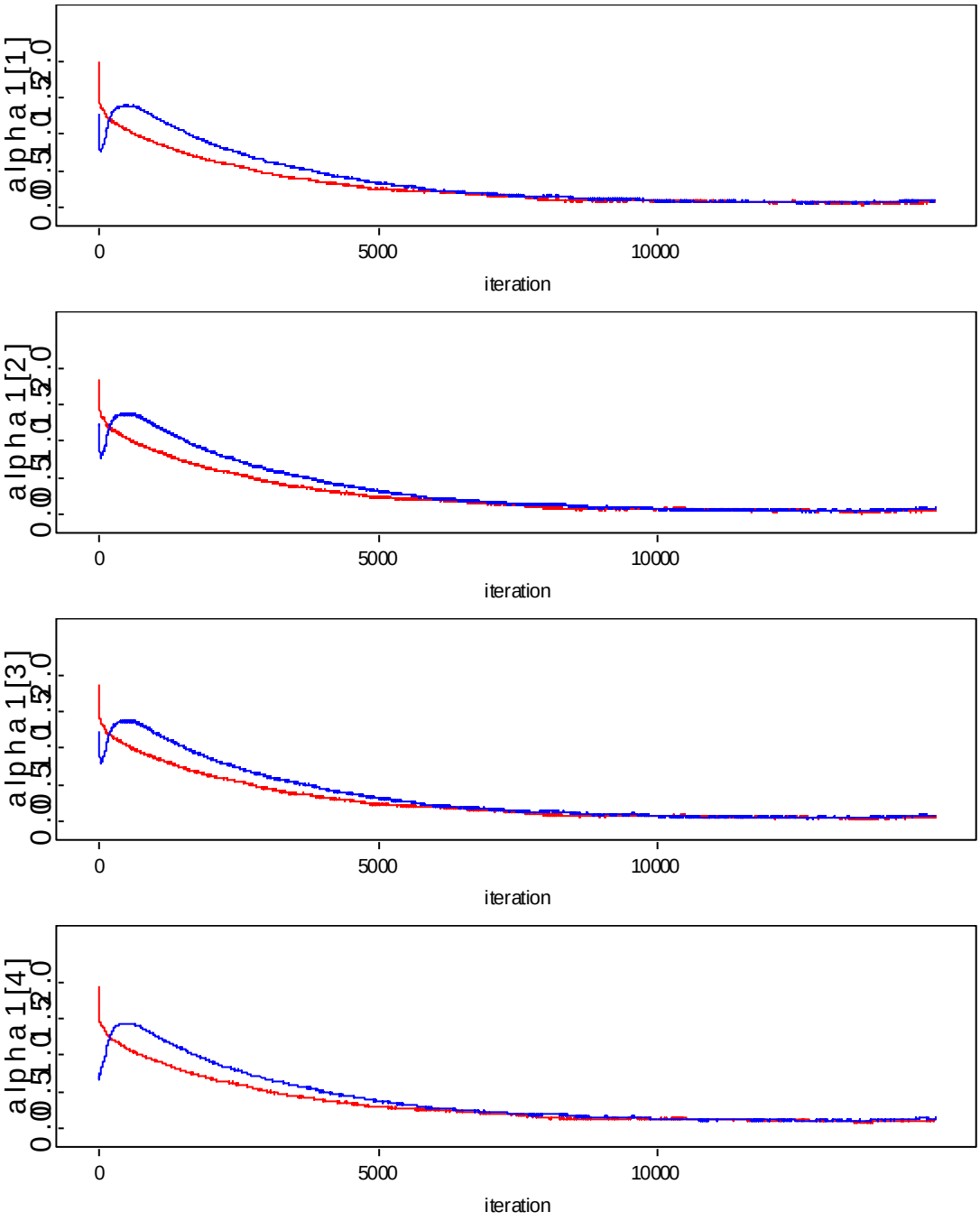

Figure 22. Diagnostic plots of convergence for random slope $\alpha_{1j.}$
1=AMI, 2=HF, 3=PN, 4=SCIP

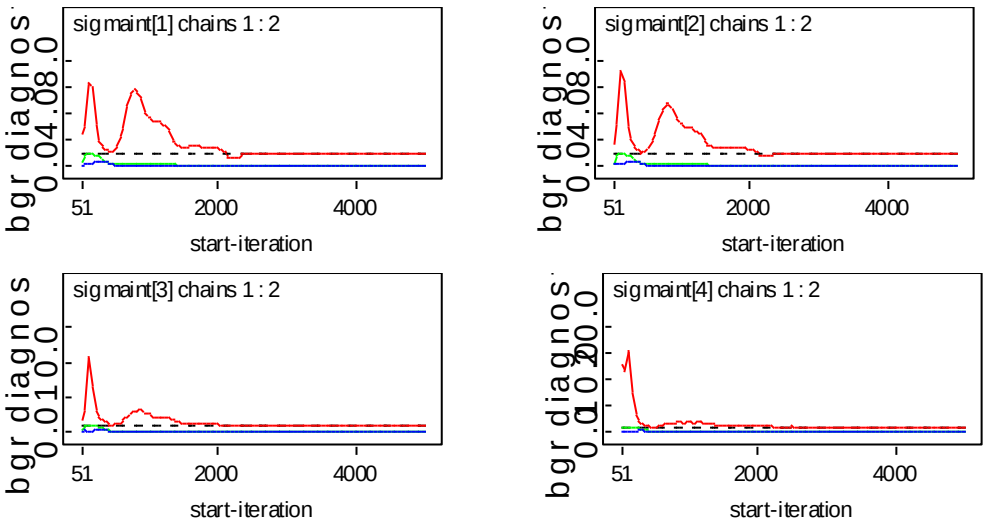

Figure 23. Gelman-Rubin statistics diagnostic plots of σ_{0i}.
1=AMI, 2=HF, 3=PN, 4=SCIP

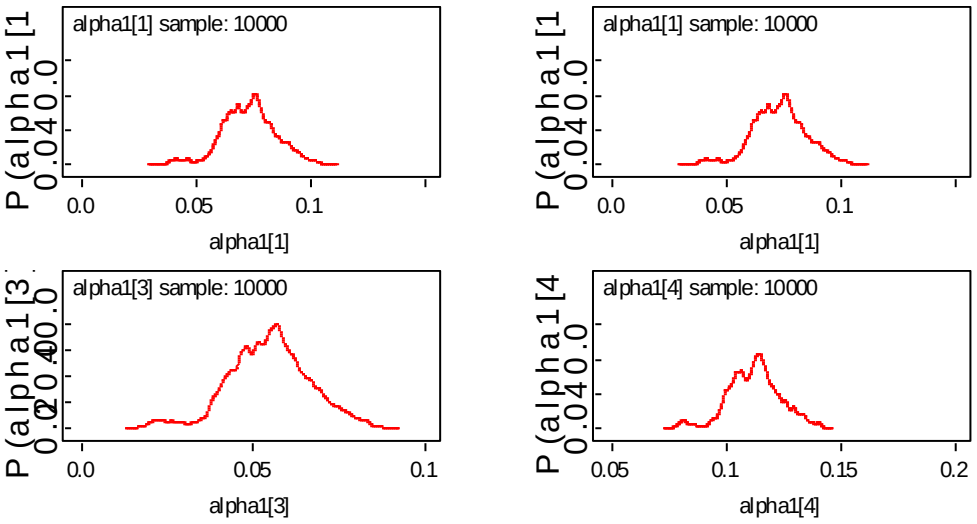

Figure 24. Posterior distribution of α_{1J}.
1=AMI, 2=HF, 3=PN, 4=SCIP

TABLE XIV

PARAMETER ESTIMATION USING FULL BAYES MODEL FOR LONGITUDINAL DATA

	Parameter	Full Bayes Estimate (SD)	Full Bayes 95% Credible Interval
Fixed Intercept	α_{0AMI}	4.316 (0.0262)	(4.264, 4.369)
	α_{0HF}	3.489 (0.0287)	(3.489, 3.545)
	α_{0PN}	3.368 (0.0199)	(3.329, 3.407)
	α_{0SCIP}	3.722 (0.0163)	(3.689, 3.754)
Fixed Slope	α_{1AMI}	0.0730 (0.0118)	(0.0458, 0.0966)
	α_{1HF}	0.0552 (0.0122)	(0.0278, 0.0794)
	α_{1PN}	0.0546 (0.0120)	(0.0264, 0.0783)
	α_{1SCIP}	0.1123 (0.0116)	(0.0842, 0.1351)
Intercept Variance	σ_{AMI}	1.2280 (0.0254)	(1.1850, 1.2730)
	σ_{HF}	0.9895 (0.0210)	(0.9488, 1.0320)
	σ_{PN}	0.6723 (0.0149)	(0.6432, 0.7011)
	σ_{SCIP}	0.8269 (0.0142)	(0.7987, 0.8550)

Parameter estimates are computed from using 5,000 posterior draws from two chains after the 10,000 chain burn-in was achieved and is shown in Table XIV. The AMI had the highest amount of discrimination over time and SCIP had the lowest amount of discrimination over time. Positive values of each time factor indicates that, on average, hospitals are showing improvement over time. Posterior distribution of the random slope parameter α_{1J} is shown in Figure 24 and shows each parameter estimate is approximately normal around the median parameter estimate.

Classifying hospitals based on the top 5th percentile of the posterior distribution of θ_{i1}, I find 588 (19.8%) hospitals showing superior QI efforts over time in all measure sets submitted for analysis. The minimum $\hat{p}_{it}$ of these top performers in the first time period is 67.87% while the minimum overall observed proportion in the first time period is 59.85% coming from the same hospital. For the hospital with the minimum rate for this group, the predicted and observed rates in the last time period are 99.30% and 97.46%, respectfully. The observed average overall improvement of composite rates over the three-year period for these hospitals is 9.94% ranging from 1.78% to 38.79%. Similarly, the predicted overall average improvement is 8.58% ranging from 1.01% to 31.43%.

There were 220 (7.4%) hospitals identified as having declining quality over the two-year period in at least one of the four measure sets. There were 74 hospitals with declining rates in all measure sets submitted, while 75 hospitals had declining rates in three out of the four measure sets.

5. HIERARCHICAL MIXED-EFFECTS POISSON REGRESSION MODELS

Poisson regression models are a natural choice when modeling rates, in particular rates of infections and mortality, to name just two. The goal in this section is to estimate an unobserved measure of hospital safety in addition to modeling performance measures used in the binomial method section of this study by utilizing mixed-effects Poisson hierarchical model. One method of determining an estimator of hospital quality is through a latent variable model for the observed individual safety measures. Construction of hospital profiles using a 2-parameter Normal-Ogive model or multivariate probit model has been proposed by Landrum et al. [2000]. This approach of profiling hospitals is the same method used in IRT models performed in psychological and educational testing. Gibbons et al. [2010] propose FB and EB methods on adverse events surveillance data utilizing a mixed-effects Poisson regression model. For each of these methods, these models assume that the underlying measure, hospital safety, is reflected through the individual measures although each measure may have a separate weight in the final score depending on the ability of the measures to discriminate between hospitals. Independence between the measures is assumed conditional on the latent variable.

5.1 The Poisson Regression Model

The research in this section explores methods to utilize the notion of latent variable models to determine hospital quality in two ways. The first is a univariate method. This method is similar to the binomial method addressed in the previous section in that it is using a latent

variable model to determine hospital quality. Instead of using hospital performance measures, the model determines hospital quality based on HAI safety measures that have not been published in the literature at the time of this writing. The measures used are CAUTI, SSI:colon, SSI:hysterectomy, CLABSI, MRSA, and *C. difficile*.

The second method employs a bivariate model to determine hospital quality. In the bivariate case, the model estimates hospital quality based on both hospital performance measures and HAI safety measures. From this model, two estimates based on both sets of data are obtained and an overall measure of hospital quality is constructed.

5.2 Univariate Poisson Regression Model for Hospital Safety Data

In this section, a univariate Poisson regression model was used to identify hospital quality based on HAI safety data. Each of the six measures is summarized and analyzed in the framework of a latent variable mixed-effects model. Graphical techniques will be used to determine convergence of the model, including the Gelman-Rubin statistics modified by Brooks and Gelman.

Let y_{ij} be the numerator for the j^{th} measure in the i^{th} hospital and n_{ij} be the denominator for the j^{th} measure in the i^{th} hospital where j=1,…6. Let θ_i be the latent variable representing the underlying measure of hospital safety for the i^{th} hospital. Within this framework, the Poisson model will be of the form as follows:

$$y_{ij} \sim f(\mu_{ij,} n_{ij})$$

$$g(\mu_{ij}) = \alpha_j + \sigma_j \theta_i \tag{5.1}$$

$$where\ \sigma_j > 0\ and\ \ \theta_i \sim N(0,1)$$

In this model 5.1, α_j represents the baseline rate for the j^{th} measure and σ_j represents the discrimination factor for the j^{th} measure. The function *f()* represents the assumed distribution with corresponding link function *g()*. For the Poisson model the following is proposed:

$$y_{ij} \sim Poisson(\mu_{ij,} n_{ij})$$

$$f(y_{ij}; \mu_{ij}) = \frac{\exp(-\mu_{ij})(\mu_{ij}{}^{y_{ij}})}{y_{ij}}$$

$$log(\mu_{ij}) = \alpha_j + \sigma_j \theta_i + \log(n_{ij}) \tag{5.2}$$

The HAI measures should have larger values of σ_j for those therapeutic areas that are less homogeneous among the hospitals. Contrary from performance measure data where the direction of improvement was an increase in the rates and larger values of the latent variable were indicative of high quality of care, the direction of improvement for HAI data is a decrease in rates. Therefore, where large negative values of the latent score, θ_i that ranges from -∞ to ∞, are indicative of high quality of care, values of zero are indicative of average quality of care, and large positive values are indicative of poor quality of care. Although assuming normality on the latent variable may be too restrictive, it solves the problem of the within-hospital correlation

between the therapeutic areas. As previously stated, the current nature of available data for HAIs, the within patient variability cannot be estimated.

The safety data were analyzed within the framework of an FB adaptation of model 5.1 using the OpenBUGS software. 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. Posterior means and posterior variances of θ_i are estimates of the quality of care and the variance. The parameter estimates are based on a chain of 5,000 iterations allowing for a 5,000 burn-in chain.

Posterior predictive checking was incorporated to determine the model fit by utilizing two separate chains with different starting values in addition to incorporating Gelman-Rubin convergence statistics and graphs. In this method, multiple chains are generated starting at over-dispersed initial values, and convergence is assessed by comparing within- and between-chain variability over the second half of those chains. The number of chains generated is denoted by M and the length of each chain by 2T. The measure of posterior variability is the width of the 100 (1 -α)% credible interval for the parameter of interest (in OpenBUGS, α=.2). From the final T iterations, the empirical credible interval for each chain is calculated. At each iteration, the average width of the intervals is computed across the M chains and denoted by W. Finally, the width B of the empirical credible interval is calculated based on all MT samples pooled together. The ratio R = B / W of pooled to average interval widths should be greater than 1 if the starting values are suitably over-dispersed; it will also tend to 1 as convergence is approached, and so

one might assume convergence for practical purposes if, for example, R <1.05. Rather than calculating a single value of R, the behavior of R can be examined over iteration-time by performing the above-procedure repeatedly for an increasingly large fraction of the total iteration range, ending with all of the final T iterations contributing to the calculation as described above.

Additionally, an EB adaptation of model 5.2 was used to analyze the safety data. Parameter estimates were obtained and compared with estimates obtained from the FB analysis. Additionally, estimates of θ_i will be obtained using the EB method and compared with estimates of θ_i from the FB model using Lin's CCC to determine the agreement rate between both methods of estimating hospital quality.

Model 5.1 was also implemented using the hospital performance measure data. Estimates of θ_i will be obtained from the FB approach and compared to the FB estimates obtained from model 4.1 and compared utilizing Lin's CCC.

5.3 **Classification of Hospitals**

As mentioned previously in section 4.1.3, the classification of hospitals is based on the estimate of quality of care score, S_i, derived from the posterior distribution of θ_i, as opposed to ranking hospitals. In this case where a decrease in rates is indicative of better performance which implies that lower values of S_i denote higher quality of care, the lower end of the distribution is of particular interest.

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. I take into account this variability by using a predefined threshold by determining the probability that the estimate falls below the 5th percentile as shown as follows:

$$P(S_i < \eta_{05}) > \gamma,$$

where η_{05} is the 5th percentile of θ_i and γ is a predefined threshold. This guarantees that a hospital is classified as being a high-performing hospital when the true score is below η_{05} 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, in order to identify exactly 5% of the hospitals as being high performers, the upper percentile of the score is used as a cutoff. Determining η_{05} such that 5% of the hospitals have a probability exceeding γ is identical to finding the threshold in which 5% of the γ100%-credible intervals of hospital performance lie below it. The symmetry of the credible intervals is used to determine the cutoff point. Note that the probability that θ_i lies below the upper bound of the ((2γ -1)x100)% credible intervals is γ. For example, if γ =.9, then the probability that θ_i is below the upper bound of its ((2 * 0.9 -1)x100)%=80% credible intervals is 0.9 since the probability of being in the credible interval 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 5th percentile of the upper bound of the ((2γ -1)x100)% credible intervals.

Classification of low-performing hospitals uses a similar approach as classifying high-performing hospitals except $P(S_i > \eta_{99}) > \gamma$ is determined where η_{99} is the 99th percentile of S_i and γ is a predefined threshold. Thus, the 1st percentile of the upper bounds of the ((2γ -1)x100)% credible intervals is calculated. Using the previous example, probability that S_i is above the lower bound of its ((2 0.9 -1)x100)%=80% credible intervals is 0.9 since the probability of being in the credible interval is 0.8 and the probability of being above the upper bound of the credible interval is 0.1. Therefore, classifying hospitals as low performers is determined by identifying the 99th percentile of the lower bound of the ((2γ -1)x100)% credible intervals.

5.4 **Results of Poisson Regression**

There were a total of 3,729 hospitals that had at least one HAI measure representing 19,295 data points for analysis. The median number of HAI measures among the hospitals was six and the first quartile was five measures. The lower 10% of hospitals had two or fewer HAI measures.

Descriptive statistics for each of the safety measures across all hospitals reporting HAIs are shown in Table XV. The highest rate of infection is seen in colon surgeries with 0.0257, with one hospital having a rate of 1.0, whereas the MRSA infection rate was the lowest at 0.000042. Half of the hospitals reporting MRSA and SSI:Abdominal had zero infections in the one year time period. The mean number of observed infections was highest in the *C. difficile* group with 13.9, and SSI:Abdominal had the lowest mean number of observed infections with 0.8 infections

per hospital. There were two separate hospitals that had an observed number of 303 infections, one hospital in CAUTI and one hospital in *C. difficile*.

TABLE XV

DESCRIPTIVE STATISTICS OF HAI MEASURE RATES

AND NUMBER OF INFECTIONS

Measure	n	Mean	Standard Deviation	Max	Median	Min
Observed Rates						
CAUTI	3,060	0.001466	0.001772	0.02564	0.001027	0
SSI:Colon	3,102	0.025727	0.042314	1.0	0.016103	0
SSI:Abdominal	2,989	0.007711	0.020646	0.5	0	0
CLABSI	3,066	0.000908	0.003843	0.2	0.000431	0
C. difficile	3,563	0.000517	0.000619	0.02273	0.000454	0
MRSA	3,515	0.000042	0.000098	0.00239	0	0
Number of Infections						
CAUTI	3,060	8.2307	19.1219	303	2	0
SSI:Colon	3,102	2.4752	3.9957	53	1	0
SSI:Abdominal	2,989	0.8073	1.6941	20	0	0
CLABSI	3,066	3.4775	7.3609	94	1	0
C. difficile	3,563	13.9722	23.2014	303	5	0
MRSA	3,515	1.398	3.0715	60	0	0

Figure 25 shows the convergence of α_i for each of the safety measures for the two separate chains. Convergence for all the measures occurred relatively quickly in OpenBUGS within 1,000 chains. Similarly, Figure 26 shows the convergence of each value of σ_i.

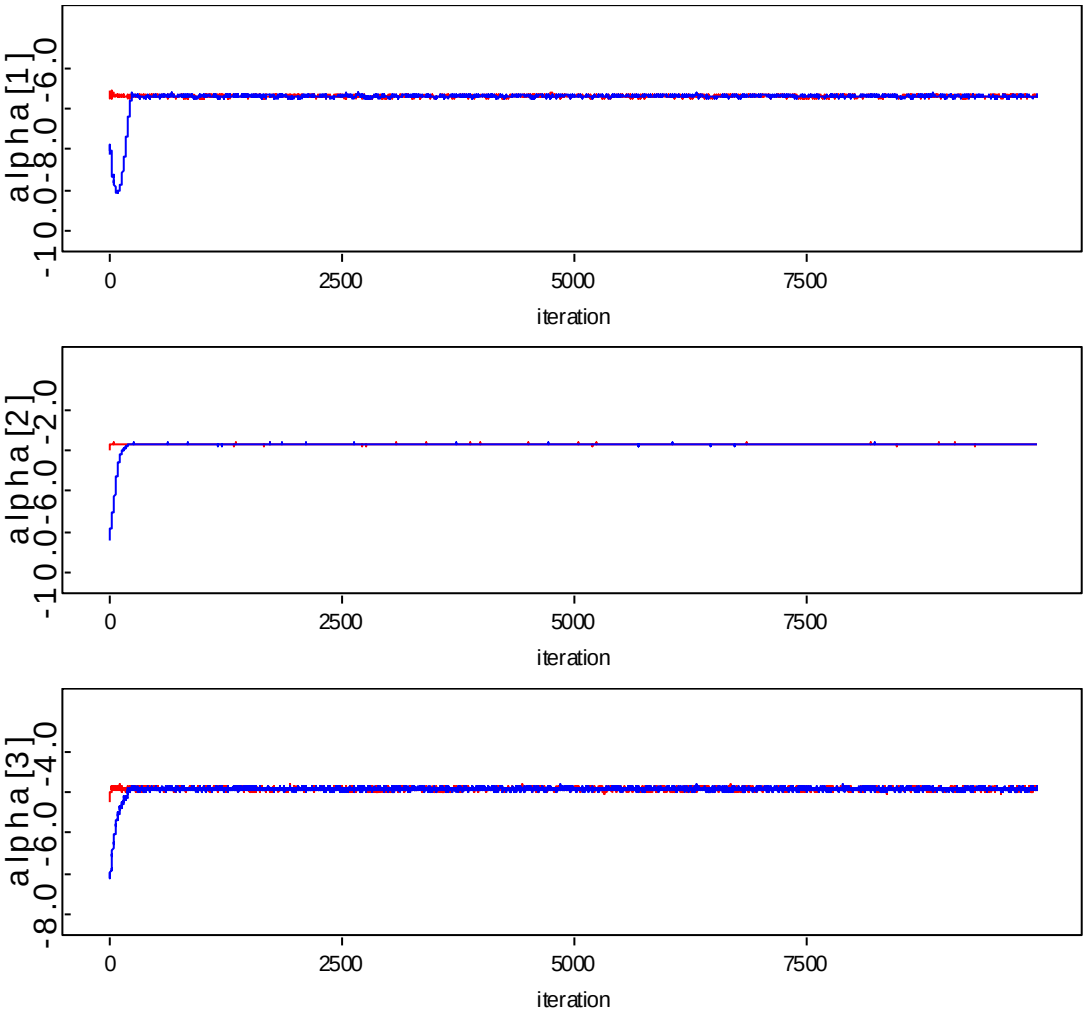

Figure 25. Diagnostic plots of α_i with Normal prior of $\boldsymbol{\theta_i}$.
1=CAUTI, 2=SSI:Colon, 3=SSI:Abdominal,

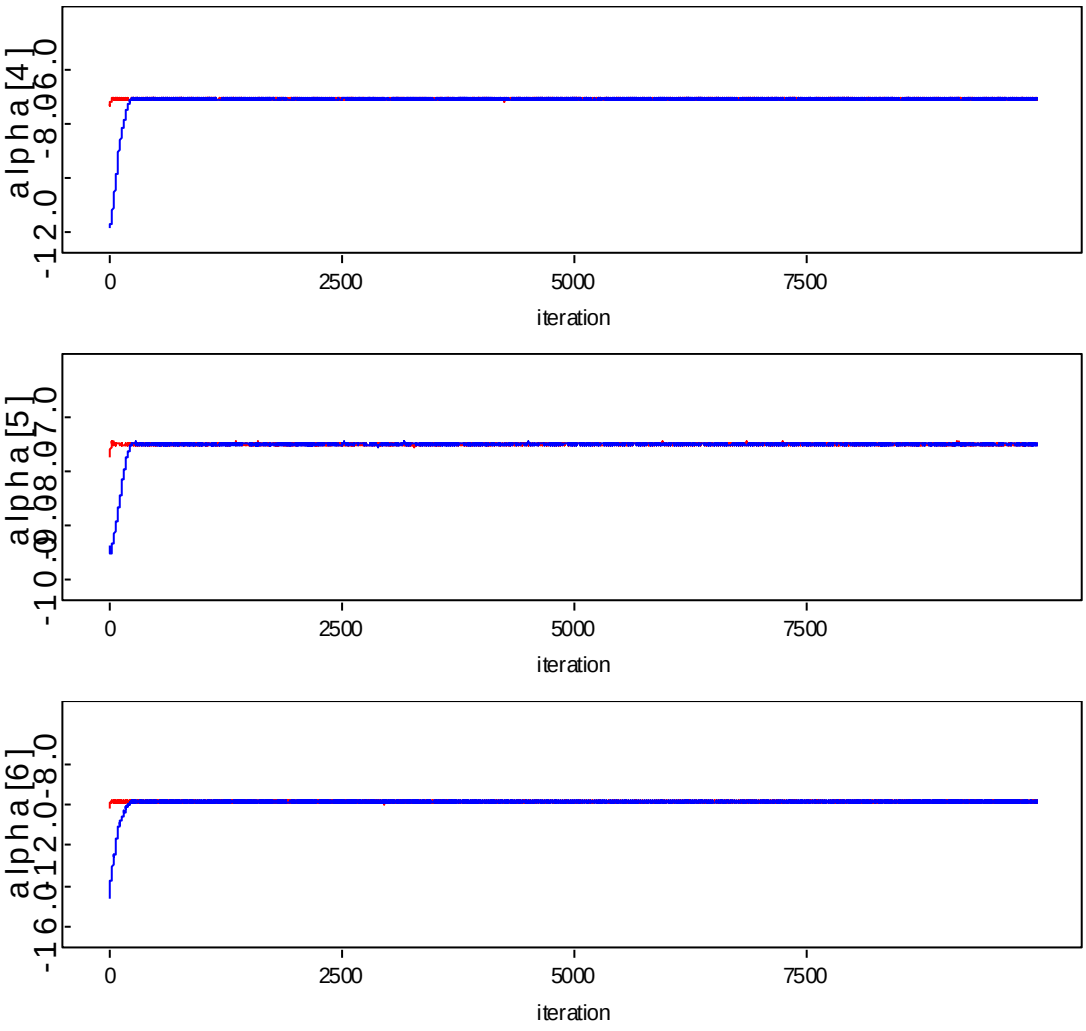

Figure 26. Diagnostic plots of α_i with Normal prior of $\boldsymbol{\theta_i}$.
4=CLABSI, 5=*C. difficile*, 6=MRSA

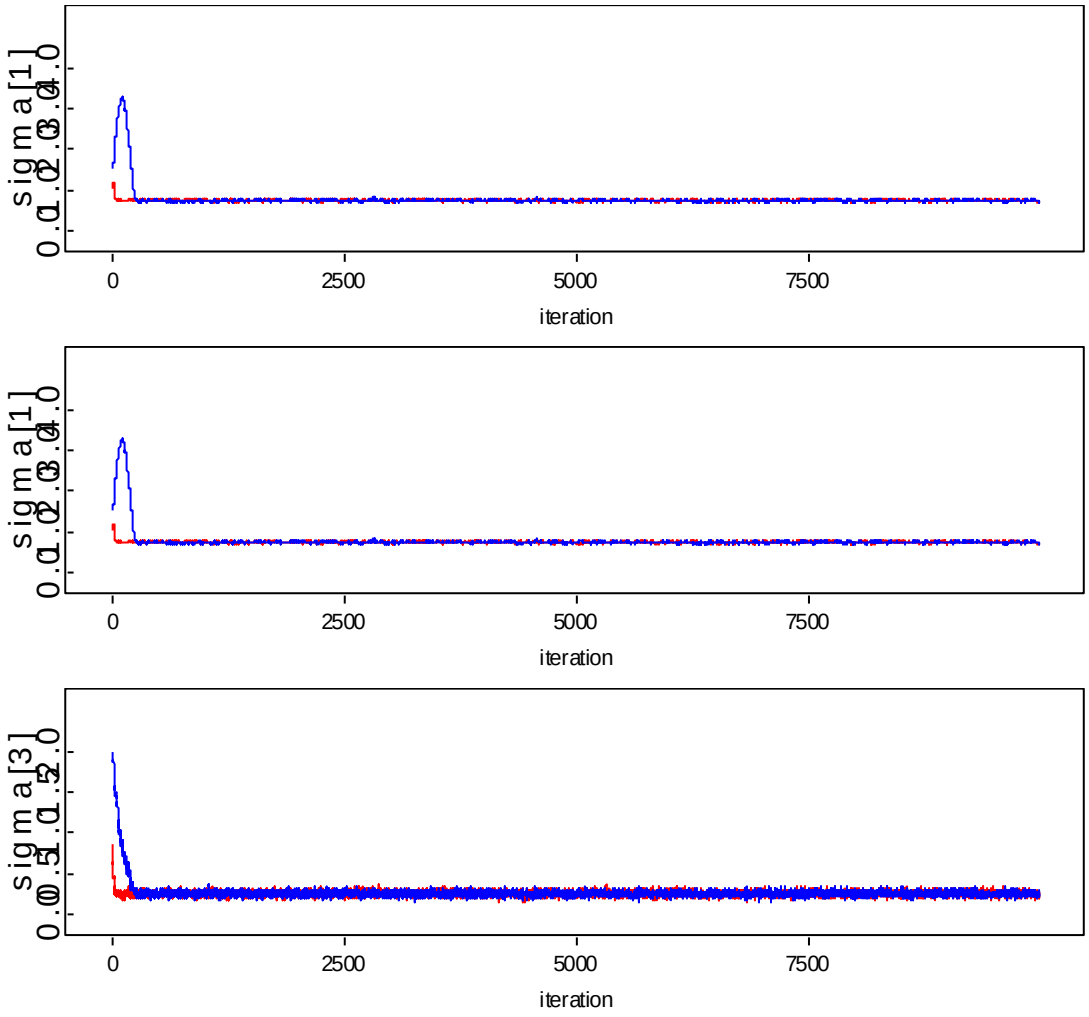

Figure 27. Diagnostic plots of σ_i with Normal prior of $\boldsymbol{\theta_i}$.
1=CAUTI, 2=SSI:Colon, 3=SSI:Abdominal,

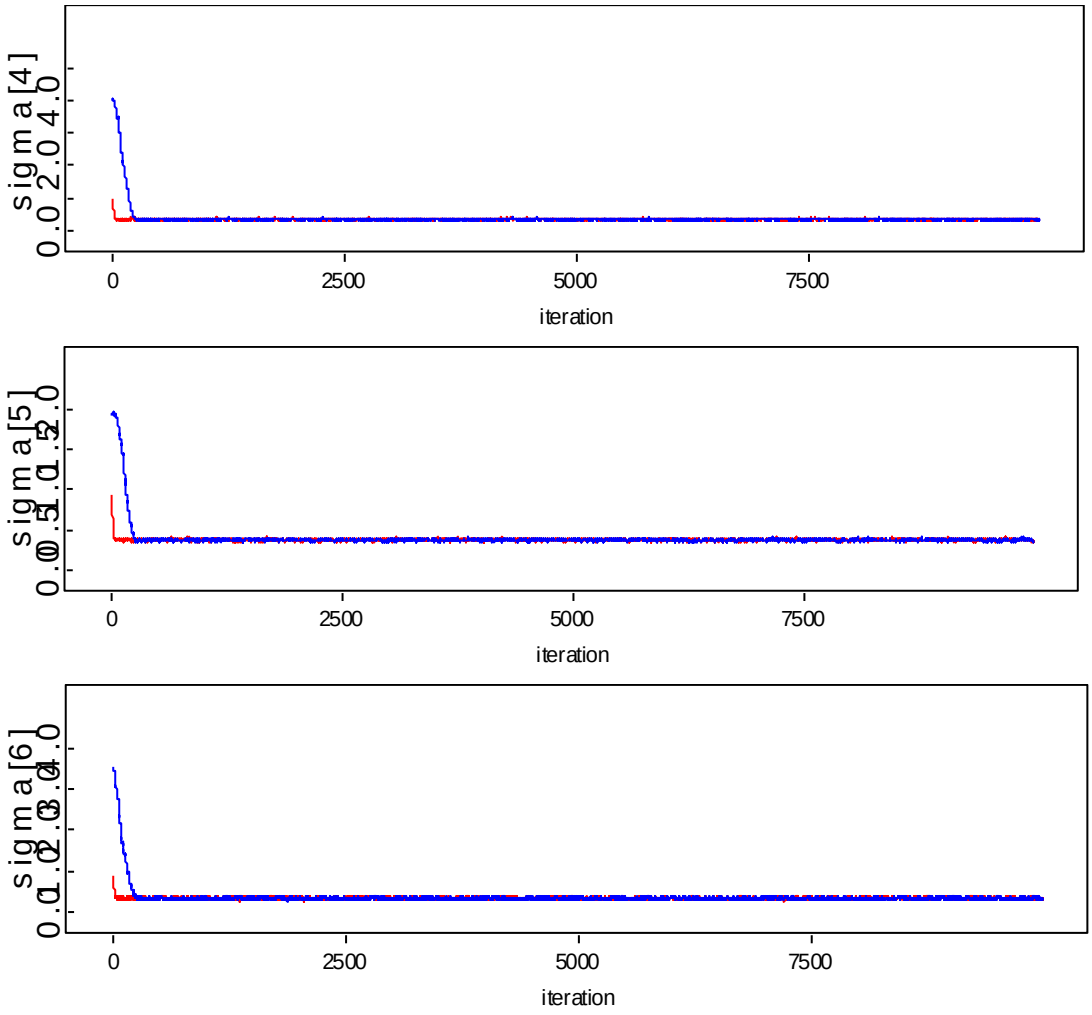

Figure 28. Diagnostic plots of σ_i with Normal prior of $\boldsymbol{\theta_i}$.
4=CLABSI, 5=*C. difficile*, 6=MRSA

Similar to the convergence of α_i, σ_i converged relatively quickly as well within 1,000 iterations. To further assure the convergence of σ_i, Figure 29 shows diagnostic plots of the Gelman-Rubin convergence statistics for each estimate of the variance for each measure. The red line represents the value of R, the pooled value of B is represented in blue and the average, W, is

shown in green. As shown with each disperse initial value, each chain is represented and not only does R converge to 1, but both B and W converges in stability. Additionally, the Deviance Information Criteria (DIC) for this model was 70,738.0.

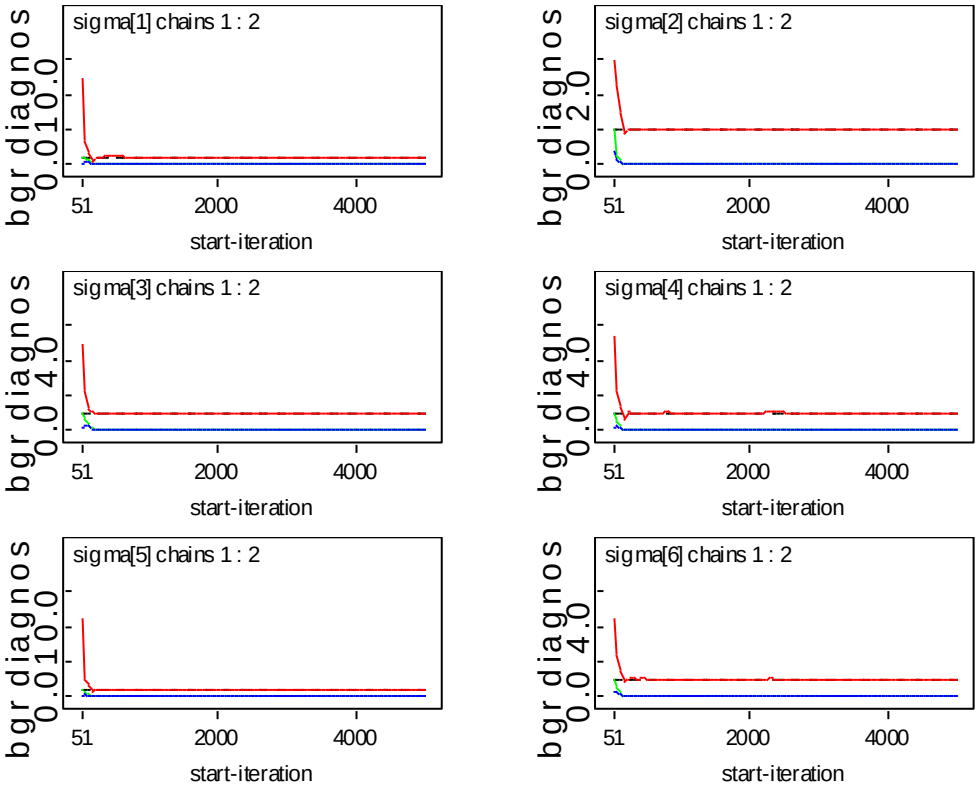

Figure 29. Gelman-Rubin statistics diagnostic plots of σ_i with Normal prior of $\boldsymbol{\theta_i}$.
1=CAUTI, 2=SSI:Colon, 3=SSI:Abdominal,
4=CLABSI, 5=*C. difficile*, 6=MRSA

Parameter estimates obtained from the model are represented in Table XVI and the posterior distribution plots of each parameter, α_i and σ_i, are shown in Figures 30 and 31,

respectively. Based on the FB implementation, each parameter has an approximately normal distribution and estimates are based on the final 5,000 iterations of the MCMC. Measure SSI:Colon had the largest estimate of α_i with mean value -3.71, and MRSA had the lowest estimate with mean value of -9.87. The CLABSI and *C. difficile* measures had similar estimates of -7.11 and -7.51, respectively. The CAUTI measure had the largest discrimination, with parameter estimate of σ_i being 0.74. Both SSI measures had similar estimates of σ_i with mean posterior values of 0.23 for colon surgeries and 0.25 for abdominal surgeries. The remaining HAI measures had similar infection rates ranging from 0.33 to 0.37. Estimates obtained from the EB implementation of model 5.1 were similar to those obtained by the FB model.

TABLE XVI

PARAMETER ESTIMATES OBTAINED FROM MODEL 5.1

Parameter	Full Bayes Estimate (95% Credible Interval)	Empirical Bayes Estimate (95% Confidence Interval)
$\alpha_{(CAUTI)}$	-6.6964 (-6.7350, -6.6566)	-6.6960 (-6.7361, -6.6559)
$\alpha_{(SSI:Colon)}$	-3.7083 (-3.7402, -3.6779)	-3.7081 (-3.7393, -3.6769)
$\alpha_{(SSI:Abdominal)}$	-4.9264 (-4.9817, -4.8740)	-4.9250 (-4.9785, -4.8716)
$\alpha_{(CLABSI)}$	-7.1051 (-7.1389, -7.0705)	-7.1052 (-7.1395, -7.0709)
$\alpha_{(C.\ difficile)}$	-7.5108 (-7.5309, -7.4897)	-7.5103 (-7.5311, -7.4896)
$\alpha_{(MRSA)}$	-9.8680 (-9.9097, -9.8273)	-9.8668 (-9.9079, -9.8258)
$\sigma_{(CAUTI)}$	0.7408 (0.7064, 0.7740)	0.7393 (0.7048, 0.7739)
$\sigma_{(SSI:Colon)}$	0.2337 (0.2009, 0.2674)	0.2334 (0.1999, 0.2669)
$\sigma_{(SSI:Abdominal)}$	0.2548 (0.1999, 0.3110)	0.2533 (0.1984, 0.3082)
$\sigma_{(CLABSI)}$	0.3503 (0.3196, 0.3807)	0.3498 (0.319, 0.3807)
$\sigma_{(C.\ difficile)}$	0.3747 (0.3555, 0.3945)	0.3737 (0.3539, 0.3934)
$\sigma_{(MRSA)}$	0.3357 (0.2978, 0.3754)	0.3343 (0.2951, 0.3735)

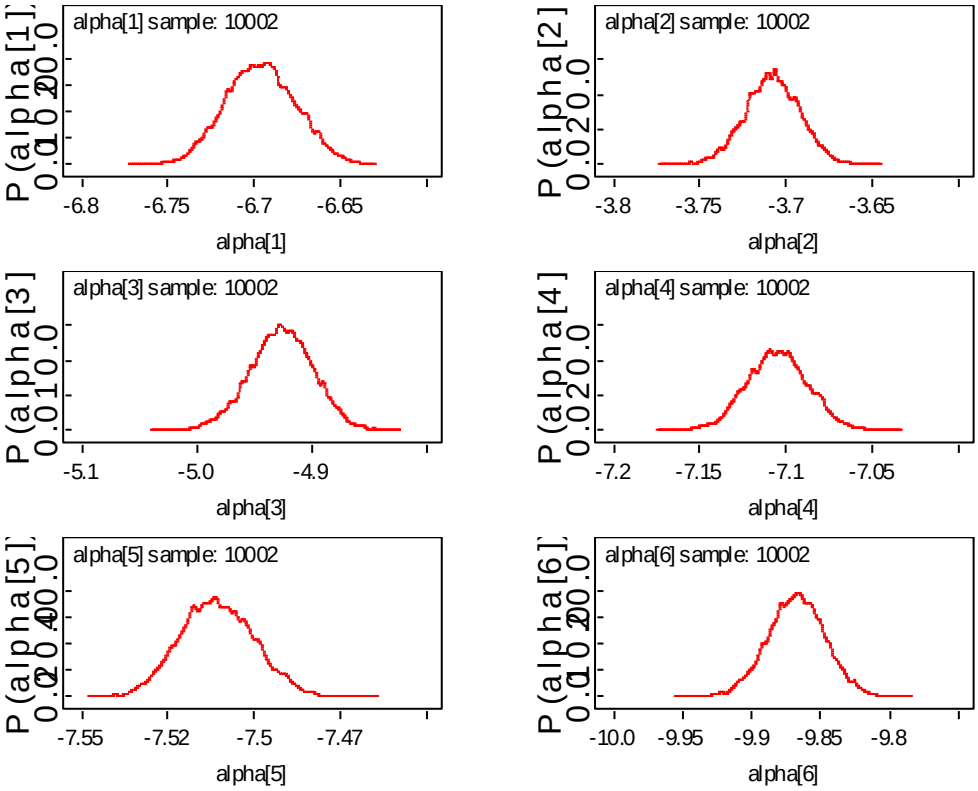

Figure 30. Posterior distribution of α_i.
1=CAUTI, 2=SSI:Colon, 3=SSI:Abdominal,
4=CLABSI, 5=*C. difficile*, 6=MRSA

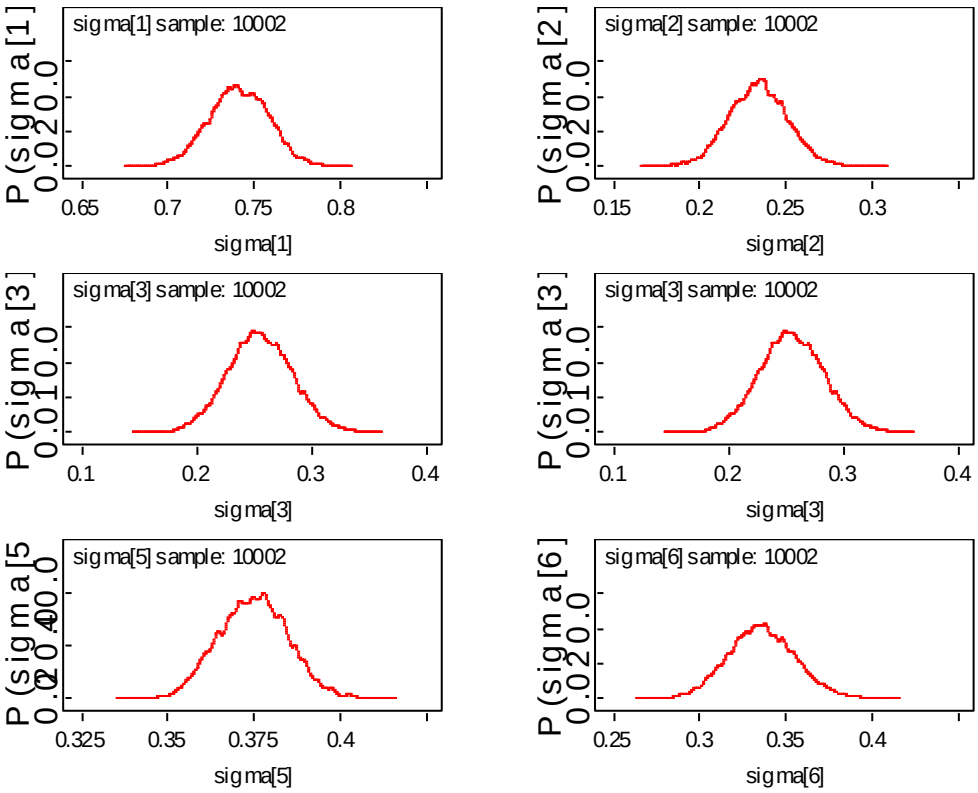

Figure 31. Posterior distribution plots of σ_i.
1=CAUTI, 2=SSI:Colon, 3=SSI:Abdominal,
4=CLABSI, 5=*C. difficile*, 6=MRSA

Classification is based on determining the bottom 5% of the posterior distribution of θ_i. Figure 32 shows the normal and kernal density plots for the estimator of θ_i and indicates they are both approximately the same. Figure 33 shows the observed infections for each of the HAI measures versus the estimate θ_i. Lower values of $\widehat{\theta_i}$ are associated with lower number of observed infections and higher values of $\widehat{\theta_i}$ are based on larger number of observed infections.

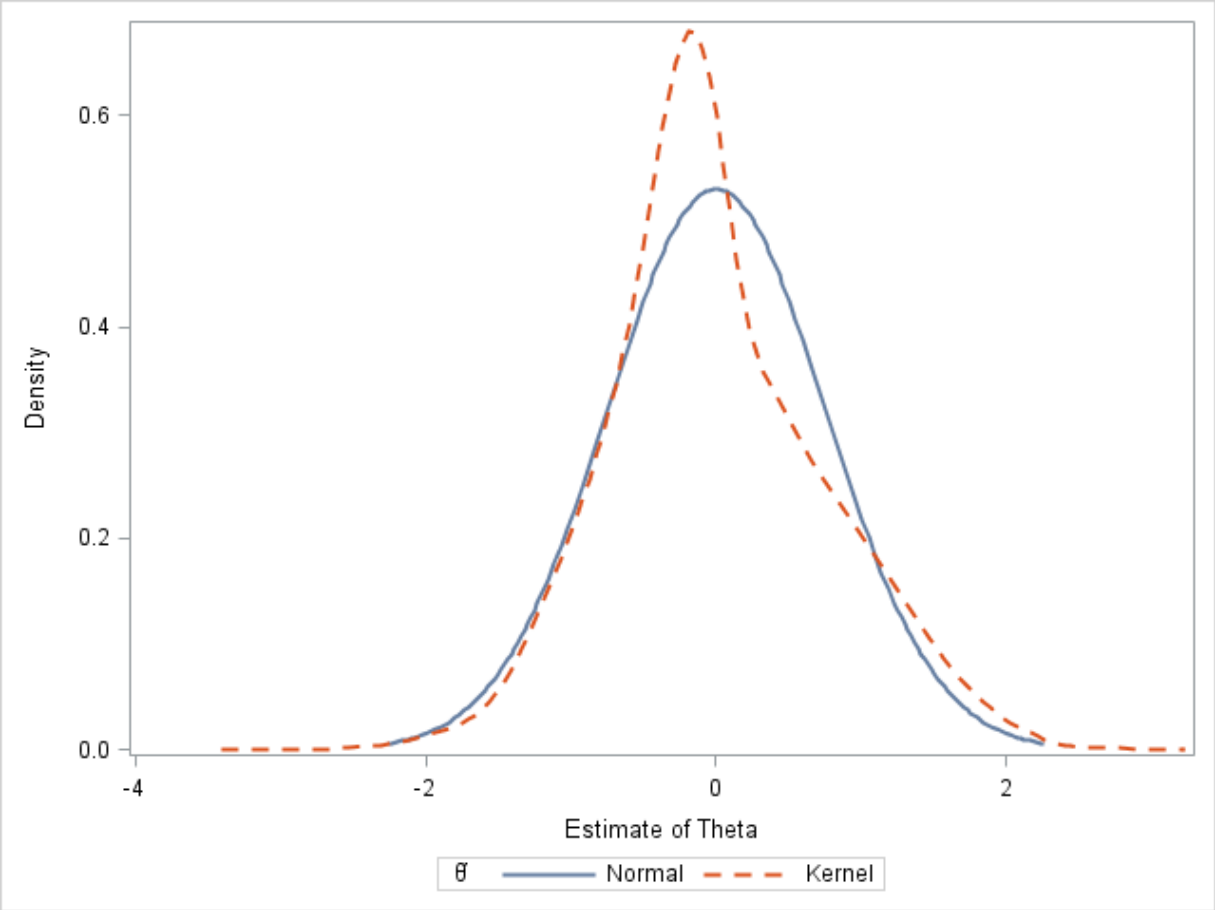

Figure 32. Normal and kernal density plot of $\widehat{\boldsymbol{\theta}_{\boldsymbol{\iota}}}$.

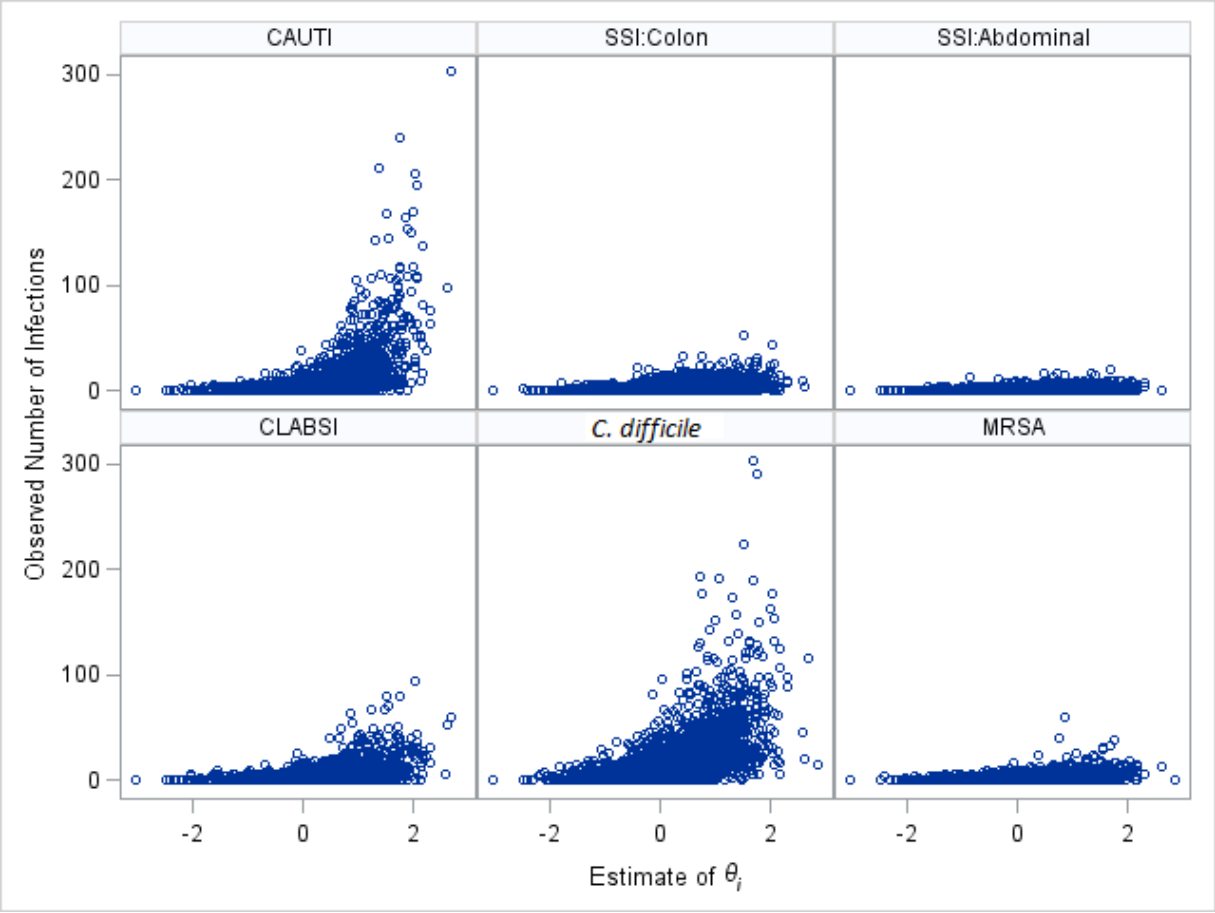

Figure 33. Observed number of infections versus $\boldsymbol{\theta_i}$.

Based on the classification scheme defined above, there are 186 hospitals identified as high-performing hospitals. Of these 186 hospitals, 167 hospitals had data in all 6 HAI measures. There were 19 hospitals that had less than six measures. Ten of the 19 hospitals had data in five measures, one hospital had data four measures, one hospital had data in three measures, and seven hospitals that were identified as a top performer had data in two measures. The rates for the individual measures for the high-performing hospitals are displayed in Table XVII. All rates

in the high-performing hospitals are lower than the national rates, with MRSA having the lowest overall mean observed rate. The number of infections for high-performing hospitals is displayed in Figure 34 with the majority of hospitals having zero infections over the one-year time period and *C. difficile* being the only measure containing hospitals having a larger number of infections above 10 over the given time period.

TABLE XVII

HAI RATES OF HIGH-PERFORMING HOSPITALS

Measure	N	Mean	Max	Median	Min
CAUTI	181	0.00026402	0.00182	0	0
SSI:Colon	179	0.009744941	0.11905	0	0
SSI:Abdominal	167	0.004373731	0.08333	0	0
CLABSI	181	0.000412064	0.00524	0	0
C. difficile	183	0.000178966	0.00076	0.000164	0
MRSA	182	0.000025614	0.00021	0	0

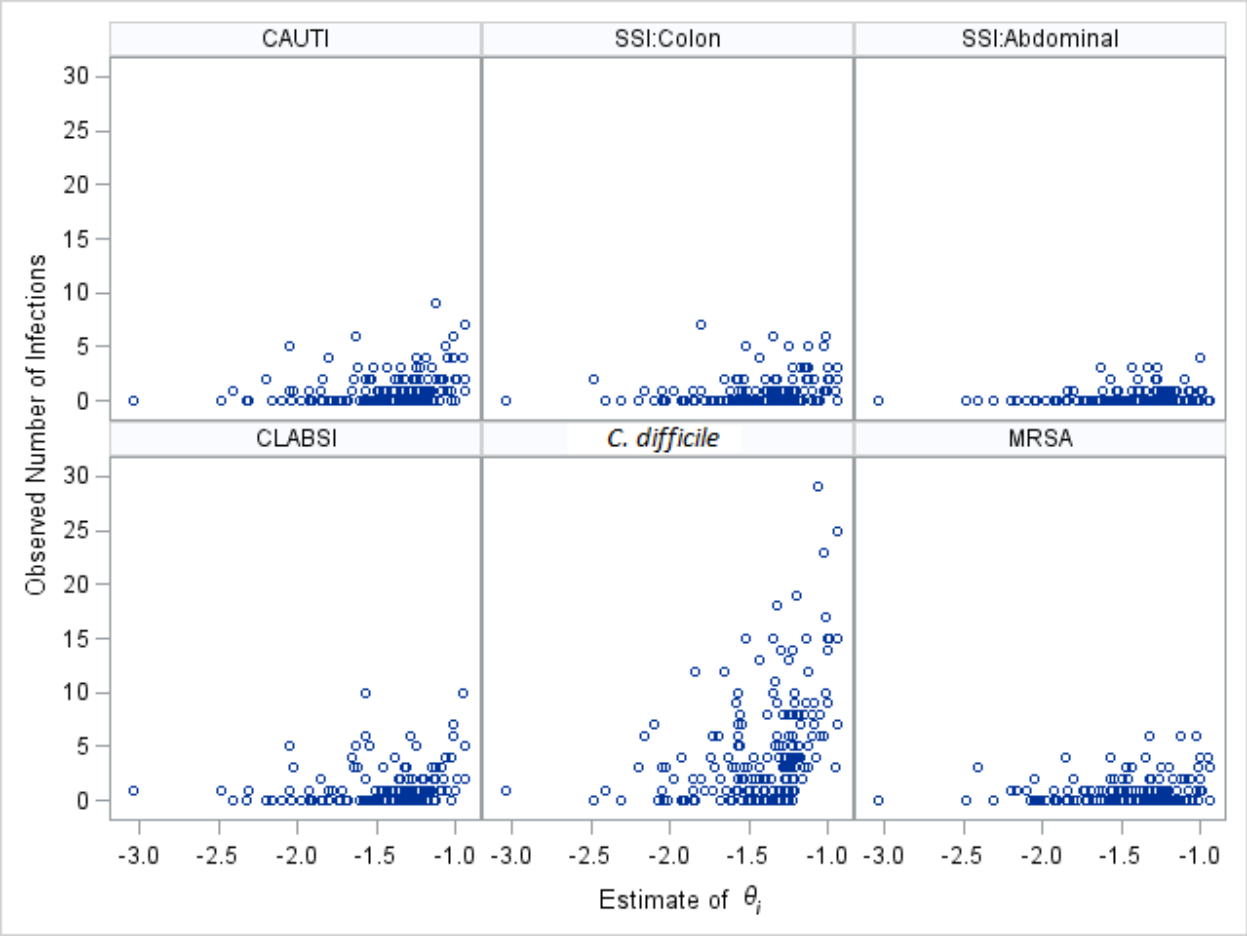

Figure 34. Observed number of infections of high-performing hospitals.

There were 37 hospitals identified as low-performing hospitals based on the method outlined in this study. There were two hospitals that submitted data for two measures, two hospitals that submitted data for three HAI measures, two hospitals that submitted data for five measures, and the remaining 31 hospitals sent data in all six of the HAI measures. Table XVIII shows the distribution of observed rates for low-performing hospitals. All rates in all the measures were higher than the national rates with both SSIs having the highest rates. Figure 35

shows the observed number of infections by the estimate of θ_i for hospitals identified as low-performing hospitals. Catheter associated urinary tract infections and *C. difficile* contained the hospitals with the highest number of infections in the low-performing group among all the measures.

TABLE XVIII

HAI RATES OF LOW-PERFORMING HOSPITALS

Measure	N	Mean	Max	Median	Min
CAUTI	35	0.006372	0.00991	0.006293	0.003038838
SSI:Colon	34	0.045344	0.12	0.043407	0
SSI:Abdominal	32	0.014438	0.09091	0.010262	0
CLABSI	36	0.002076	0.00428	0.001893	0
C. difficile	36	0.001111	0.00323	0.001063	0.000413234
MRSA	33	0.000108	0.00033	0.000103	0

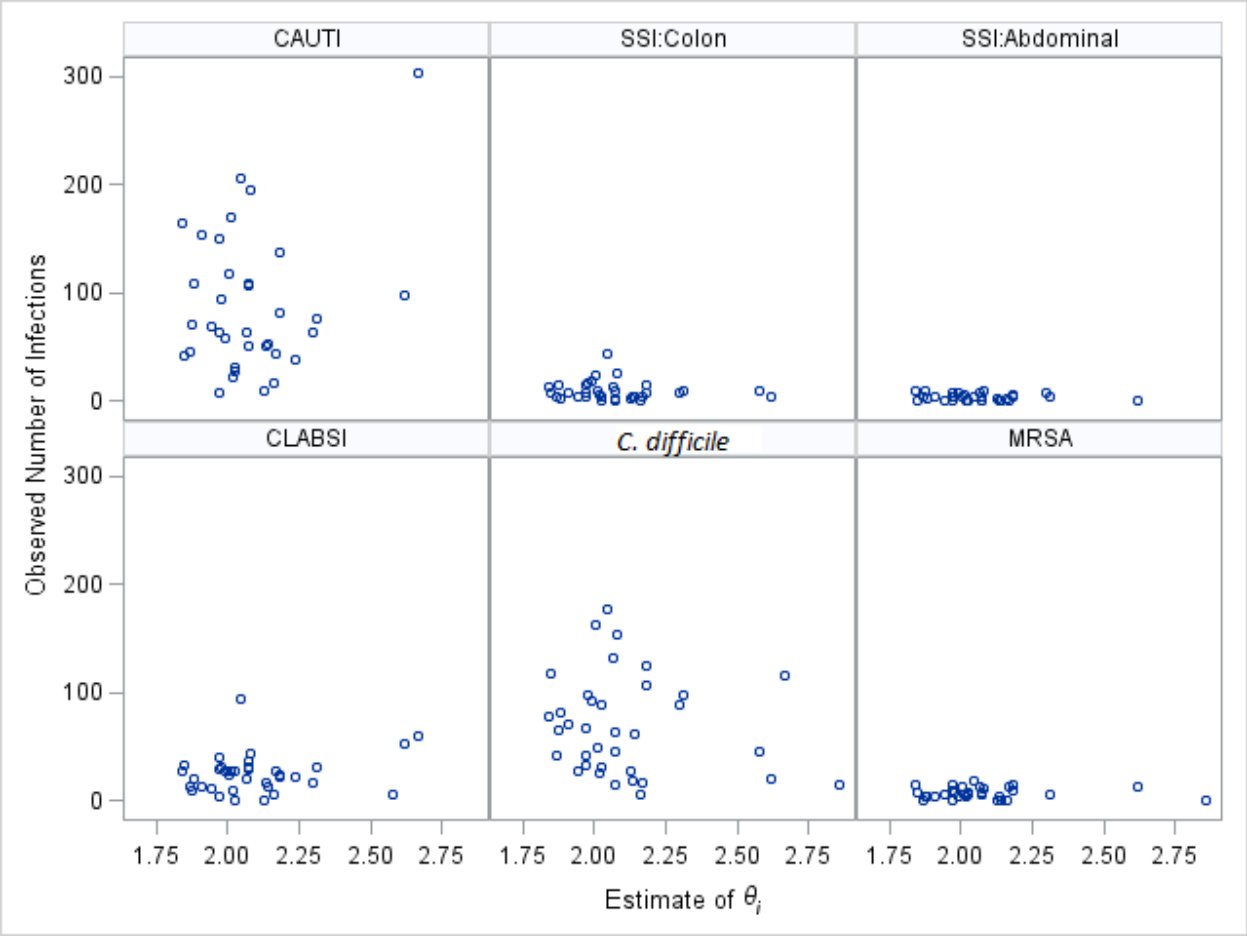

Figure 35. Observed number of infections of low-performing hospitals.

Figure 36 plots the estimates from both EB and FB analyses and the correlation between both methods is almost 1. The CCC obtained has a value of 0.9996 indicating there is almost a 100% agreement between both methods.

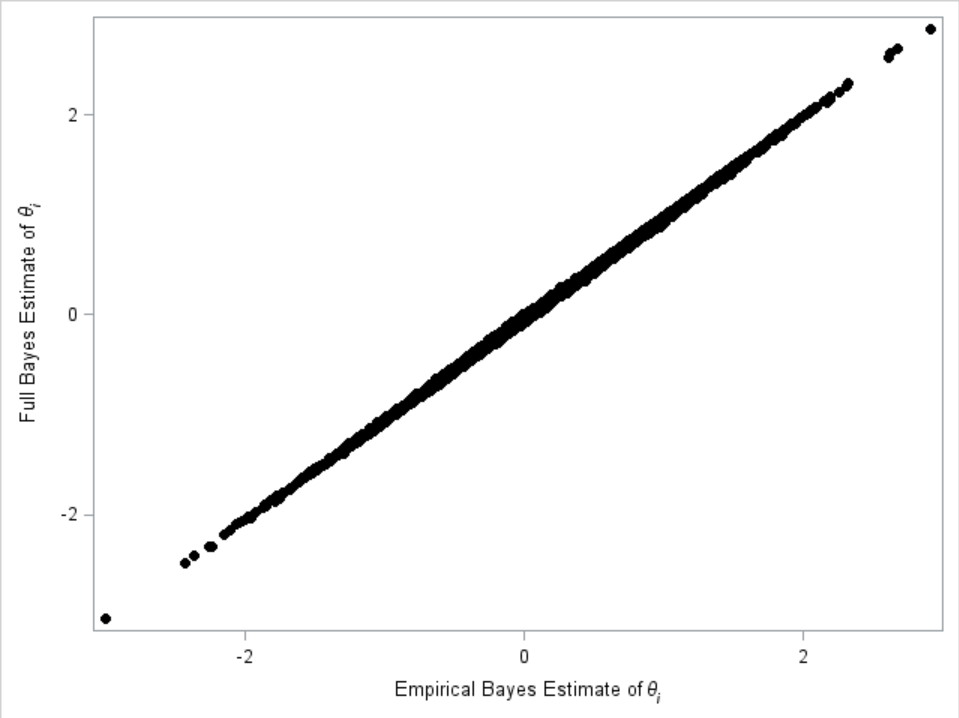

Figure 36. Plot of Empirical Bayes estimates versus Full Bayes estimates of $\boldsymbol{\theta_i}$.

With respect to identifying high and low performers, the Poisson model properly identified those hospitals with low number of observed infections in the high-performing group and the model identified hospitals with high numbers of infections as low-performing hospitals. The overall fit of the Poisson model had similar DIC when compared to models utilizing the binomial models. In conclusion, the Poisson model works well in identifying top performers utilizing only aggregated summary data. If patient level data were available, then more appropriate techniques would be able to be studied to verify the validity of the models.

5.5 **Poisson Model Applied to Performance Measure Data**

Model 5.1 has been applied to hospital performance measure data using the FB method to estimate the model parameters. All hospital performance measure data was used as opposed to analyzing those hospitals that had complete data in all four measure sets. That is, all 2,957 hospitals used in section 4.1.4 are included in the univariate Poisson model.

Convergence of the Poisson model occurred before the 2,000 iterations as seen in the diagnostic convergence plots of σ_i displayed in Figure 37. Thus, the final 5,000 iterations from both chains of the MCMC procedure were used in calculating the parameter estimates displayed in Table XIX. The DIC computed for the Poisson model was 89,070.0 compared to 79,959.4 using the binomial model. In this model, HF had the highest amount of discrimination and AMI had the lowest amount of discrimination.

TABLE XIX

PARAMETER ESTIMATION USING FULL BAYES MODEL FOR STATIC DATA

Parameter	Full Bayes Estimate (95% Credible Interval)
α_{AMI}	-0.0104 (0.0000, -0.0097)
α_{HF}	-0.0280 (0.0000, -0.0259)
α_{PN}	-0.0305 (0.0000, -0.0290)
α_{SCIP}	-0.0135 (0.0000, -0.0131)
σ_{AMI}	0.0022 (0.0000, 0.0033)
σ_{HF}	0.0054 (0.0001, 0.0082)
σ_{PN}	0.0039 (0.0001, 0.0060)
σ_{SCIP}	0.0026 (0.0000, 0.0036)

Estimates of θ_i are obtained from the model. Applying the same methodology for calculating high- and low-performing hospitals yields 147 high performers and 29 low performers. The high-performing group had an average observed rate, p_i, of 99.57% ranging from 98.99% to 100% whereas the low-performing group had an average rate of 91.23% ranging from 78.43% to 97.52%.

In comparing estimates of θ_i from both the binomial and the Poisson method, the correlation was 0.6479, which indicates a slight positive linear relationship. The density of both estimates are displayed in Figures 38 and 39.

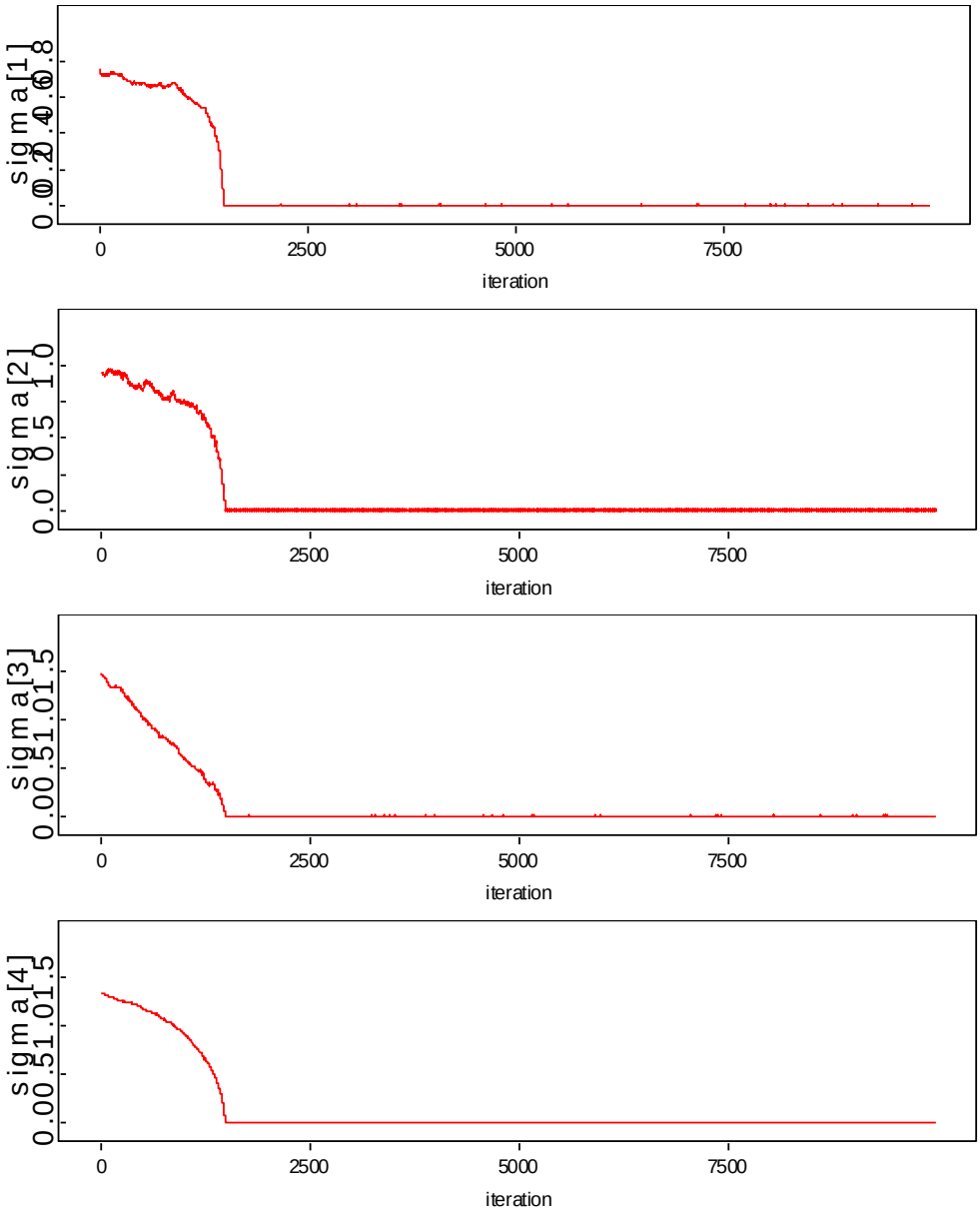

Figure 37. Diagnostic plots of σ_i.
1=AMI, 2=HF, 3=PN, 4=SCIP

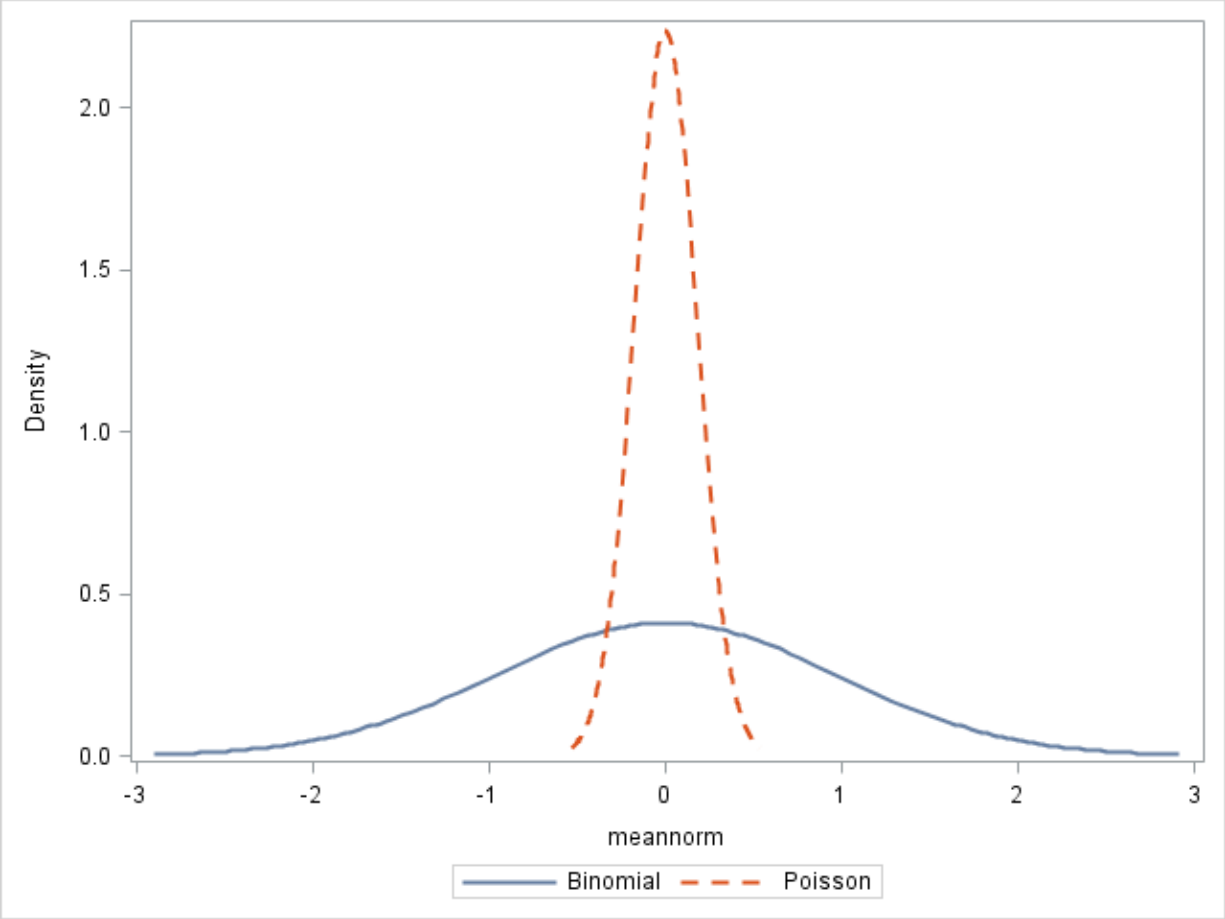

Figure 38. Density plot for binomial and Poisson estimates of $\theta_{i.}$

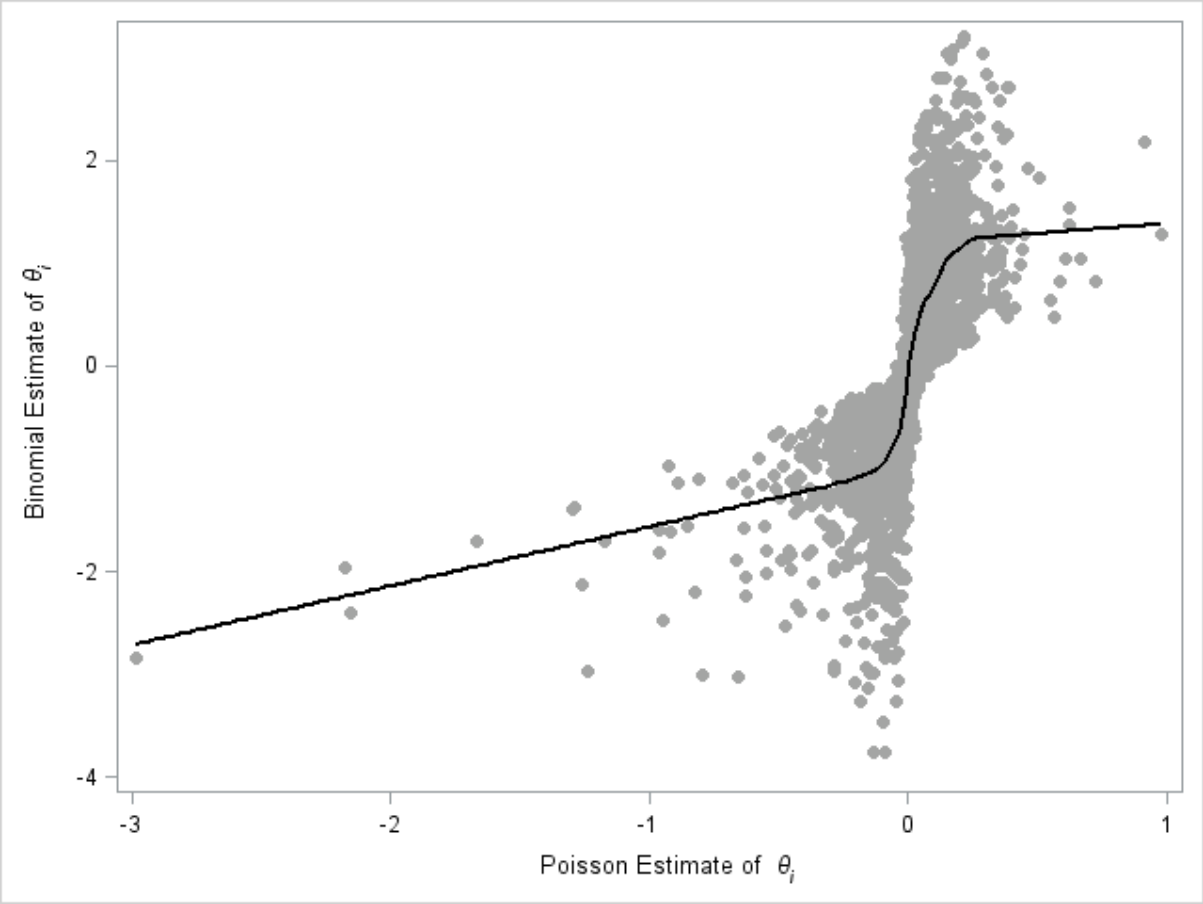

Figure 39. Plot of estimates of θ_i from Poisson and binomial models.

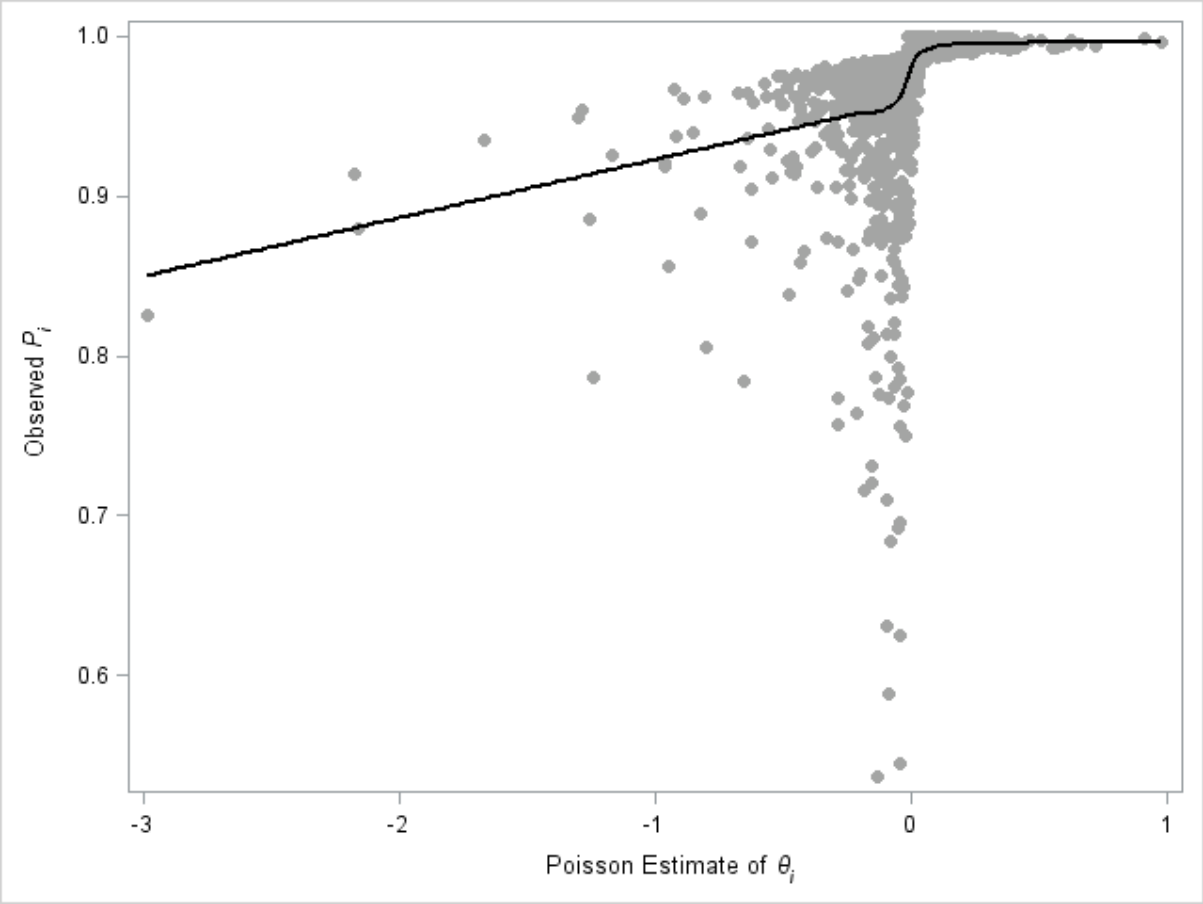

Figure 40. Plot of the estimate of $\boldsymbol{\theta_i}$ versus observed rate P_i**.**

The Poisson estimates had a narrower curve with a smaller variance compared to the binomial estimates. As seen in the plot of the kernel density, the estimates obtained from the Poisson model are slightly skewed and fall below 1.0 ranging from -2.98 to 0.97, whereas estimates obtained from the binomial model had a more symmetrical bell shaped, normal, curve. Figure 40 shows the plots of each estimates of θ_i from each model with the Loess line, which is a local regression smoothing plot and shows there is slight agreement in the middle and tails but

not total agreement, as some hospitals with lower rates were not identified as low-performing hospitals by the Poisson method as confirmed by the calculation of the CCC with a value of 0.229. One possible reason is that the restriction of the variance being the same as the mean is too restrictive for the binomial data that skewed towards 1.

6. BIVARIATE MODEL

In this section, hospital quality is estimated using a vector $\boldsymbol{\theta} = (\theta_1, \theta_{2,\ldots}, \theta_k)$ that has a generalized multivariate normal distribution based on *k* different aspects of quality of care within the framework of a Poisson and binomial model. In this study, $\boldsymbol{\theta}$ is will 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} be the numerator for the j^{th} measure in the i^{th} hospital and n_{ij} be the denominator for the j^{th} measure in the i^{th} hospital where j=1,..,6. Let θ_s be the latent variable representing the underlying quality measure of hospital safety. Similarly, for hospital performance measures, I assume *j* ranges from 1 to 4 in this study and let θ_{pm} be the latent variable representing the underlying quality measure of hospital performance measures. Under this framework, the assumption that the overall vector of hospital quality is represented by $\theta_i = (\theta_{pm}, \theta_s)$, are exchangeable to the joint distribution, and $p(\theta_{pm}, \theta_s)$ is invariant to permutations of the indices for the i^{th} hospital. This means 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, then this yields the following hospital specific:

$$p(\boldsymbol{\theta}_i|\varphi) = \prod_{j}^{k} p(\theta_{ij}|\varphi)$$

$$p(\boldsymbol{\theta}_j) = \int \left[\prod_{i}^{k} p(\theta_{ij}|\varphi) \right] p(\varphi) d\varphi$$

And the joint posterior distribution is

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

Within this framework, a two component bivariate regression model is used to determine the two estimates of hospital quality based on performance measures and hospital safety derived from HAIs. In general, the model will be of the form previously shown as follows:

$$\begin{aligned} y_{ij} &\sim f(\mu_{ij,} n_{ij}) \\ g(\mu_{ij}) &= \alpha_j + \sigma_j \theta_i \\ where\ \sigma_j &> 0\ and\ \ \theta_i \sim MVN(0, \sum). \end{aligned} \tag{5.3}$$

In this model 5.3, α_i represents the baseline rate for the j^{th} measure and σ_j represents the discrimination factor for the j^{th} measure. The function *f(*)* represents the assumed distribution with corresponding link function *g().* Based on previous sections, *f(*)* will be the binomial distribution for the performance measure data and *f(*)* will be the Poisson model for the safety data and will be as follows:

For hospital safety data

$$y_{ij} \sim Poisson(\mu_{ij,} n_{ij})$$

$$f(y_{ij}; \mu_{ij}) = \frac{\exp(-\mu_{ij})(\mu_{ij}{}^{y_{ij}})}{y_{ij}}$$

$$log(\mu_{ij}) = \alpha_j + \sigma_j \theta_i + \log(n_{ij})$$

where $\log(n_{ij})$ is the offset term.

And for hospital performance measure data:

$$y_{ij} \sim Bin(n_{ij}, p_{ij})$$

$$f(y_{ij}; n_{ij}, p_{ij}) = \binom{n_{ij}}{y_{ij}} p_{ij}^{y_{ij}} (1 - p_{ij})^{n_{ij} - y_{ij}}$$

$$log(p_{ij}) = \alpha_j + \sigma_j \theta_i$$

$$\theta_i = \begin{pmatrix} \theta_{pm} \\ \theta_s \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right)$$

where Σ is a Wishart distribution with hyperparameters analogous to precision.

A full Bayesian model was employed in OpenBUGS to obtain hospital estimates of using the same prior distributions on α_i and β_i as in the previous models in sections 5.1 and 5.2, and Σ will have a Wishart prior distribution with hyperparamenters associated with precision.

6.1 **Estimating Hospital Quality from the Bivariate Model**

For each hospital, estimates of $\boldsymbol{\theta} = (\theta_{pm}, \theta_s)$, are obtained where θ_s represents the estimate of hospital safety quality based on the HAIs and θ_{pm}represents the estimate of hospital quality based on the hospitals' ability to provide the appropriate treatment for performance measures. In general, $\boldsymbol{\theta}$ can be a vector with *k* items depending on the number of valid estimates of hospital quality that can 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. The method in the bivariate case is as follows. First is to determine the credible region for each hospital by finding the points $\{\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}: \left(\widehat{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_{\boldsymbol{i}}\right)^T \Sigma^{-1} \left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}} - \bar{\boldsymbol{\theta}}_{\boldsymbol{i}}\right) \leq \chi^2_{1-\gamma}(k)\}$ where $\bar{\theta}_i$ is the centroid of the ellipse and k=2 (Berger, 1980). For each hospitals define $\boldsymbol{\theta}^{max} = (\max(\theta_{pm}), \max(\theta_S))$ and $\boldsymbol{\theta}^{min} = (\min(\theta_{pm}), \min(\theta_S))$. The lower bound of the credible region is defined as the point $\widehat{\boldsymbol{\theta}}^{\boldsymbol{LB}} = argmin\left(\left|\left|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{min}\right|\right|\right)$, and the upper bound of the credible region is defined as $\widehat{\boldsymbol{\theta}}^{\boldsymbol{UB}} = argmin(||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{max}||)$ (see Figure 41). From the lower bounds, a linear combination for each upper and lower bound is defined as $LB = \boldsymbol{w}^{\boldsymbol{T}}\widehat{\boldsymbol{\theta}}^{\boldsymbol{LB}}$ and the upper bound is defined as $UB = \boldsymbol{w}^{\boldsymbol{T}}\widehat{\boldsymbol{\theta}}^{\boldsymbol{UB}}$ where $\boldsymbol{w}^{\boldsymbol{T}}$ 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. Thus, classification follows as previously mentioned.

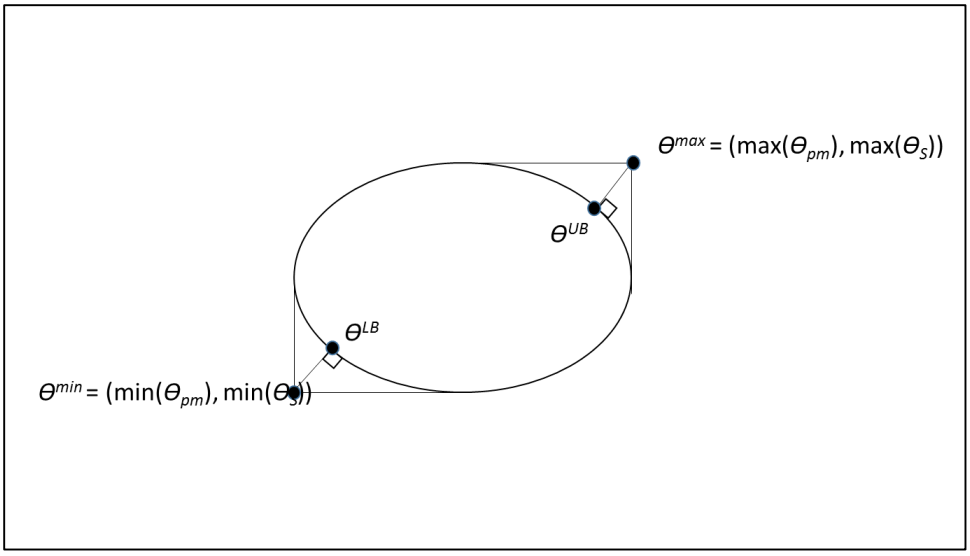

Figure 41. Identification of upper and lower bounds of $\boldsymbol{\theta_i}$.

6.2 **Results of the Bivariate Model**

This analysis consists of those hospitals that sent in data for both safety measures and hospital performance measures without any other constraints on the number of measures or sample sizes. There were 2,432 hospitals that had data in at least one measure set for the performance measure data and had data in at least one of the patient safety infection measures. The distribution of the data for each data grouping is presented in Table XX. For the performance measures, all of the measures sets had mean overall rates above 95% and 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 XX

SUMMARY STATISTICS OF THE RATES FOR BIVARIATE DATA

Measure Performance Measures	n	Mean (Std)	Max Rate	90th Percentile	Median	10th Percentile	Min
AMI	2,293	0.976 (0.0594)	1	1	0.99265	0.94118	0
HF	2,385	0.965 (0.0652)	1	1	0.99057	0.9	0
PN	2,374	0.968 (0.0416)	1	1	0.97727	0.93293	0.41176
SCIP	2,337	0.984 (0.0207)	1	0.99745	0.989	0.96862	0.66667
Infection Measures							
CAUTI	2,327	0.0016 (0.0016)	0.01613	0.00377	0.00121	0	0
SSI:Colon	2,330	0.0257 (0.0316)	0.33333	0.06122	0.01839	0	0
SSI: Abdominal	2,256	0.0080 (0.0183)	0.33333	0.02273	0	0	0
CLABSI	2,332	0.0010 (0.0043)	0.2	0.00222	0.00059	0	0
C. difficile	2,408	0.0006 (0.0004)	0.00765	0.00108	0.00054	0	0
MRSA	2,399	0.0001 (0.0001)	0.00095	0.00013	0.00002	0	0

Convergence results of model 5.3 are displayed in Figures 42–45. 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 and Figure 46 shows that both chains converge utilizing the Gelman-Rubin statistic. For each estimate, the ratio for each of the two chains converges to 1, indicating agreement and convergence.

Density plots for each parameter estimate are presented in Figures 47 and 48. Plots of α show the estimates follow a normal distribution. Similarly, the estimates of σ_i also follow a normal distribution and those with lower amounts of discrimination near zero show a truncated normal distribution which follows the initial assumption.

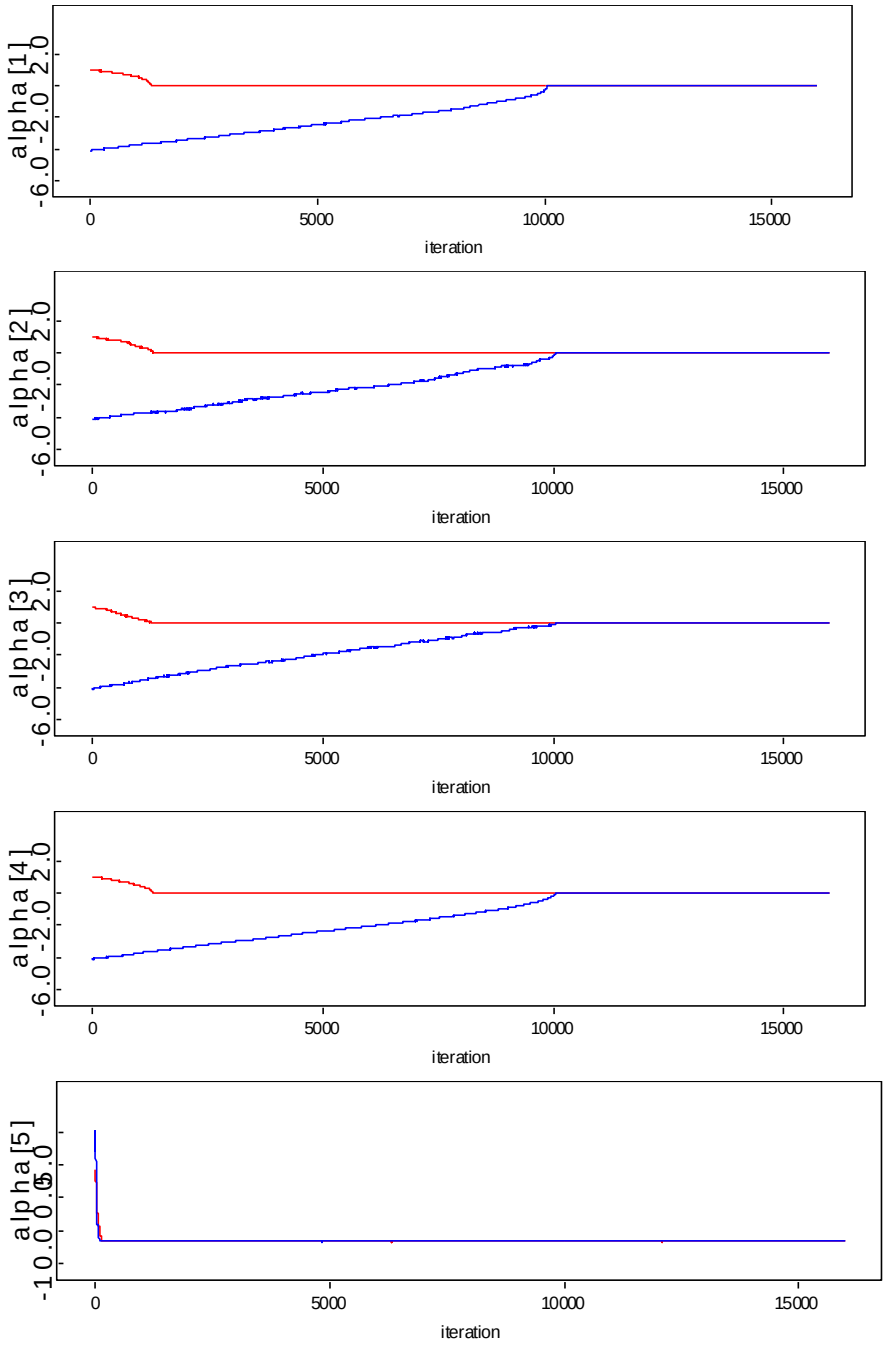

Figure 42. Convergence plots for α_i.
1=AMI, 2=HF, 3=PN, 4=SCIP, 5=CAUTI

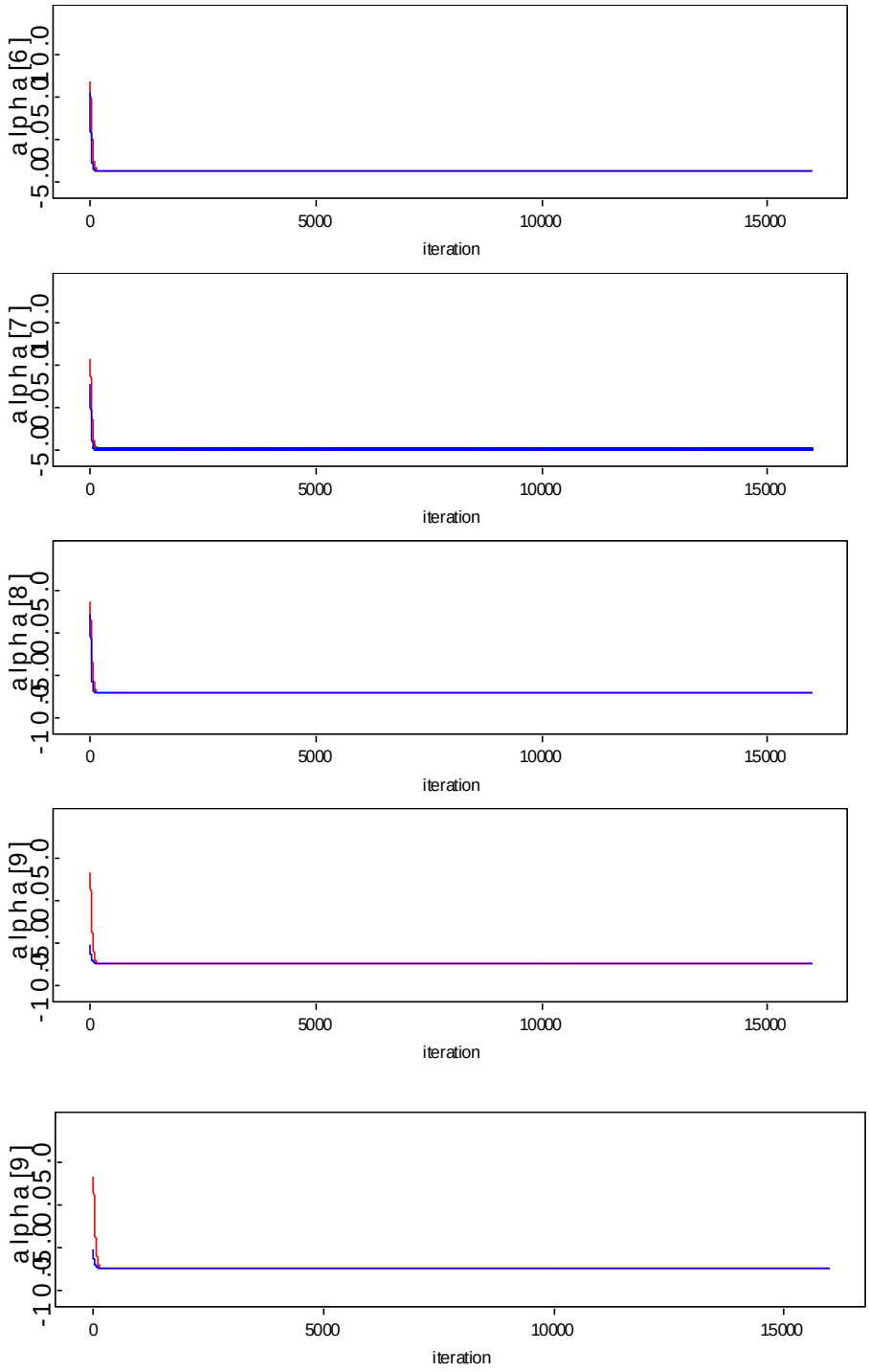

Figure 43. Convergence plots for $\alpha_{i.}$
6=SSI:Colon, 7=SSI:Abdominal, 8=CLABSI, 9=*C. difficile*, 10=MRSA

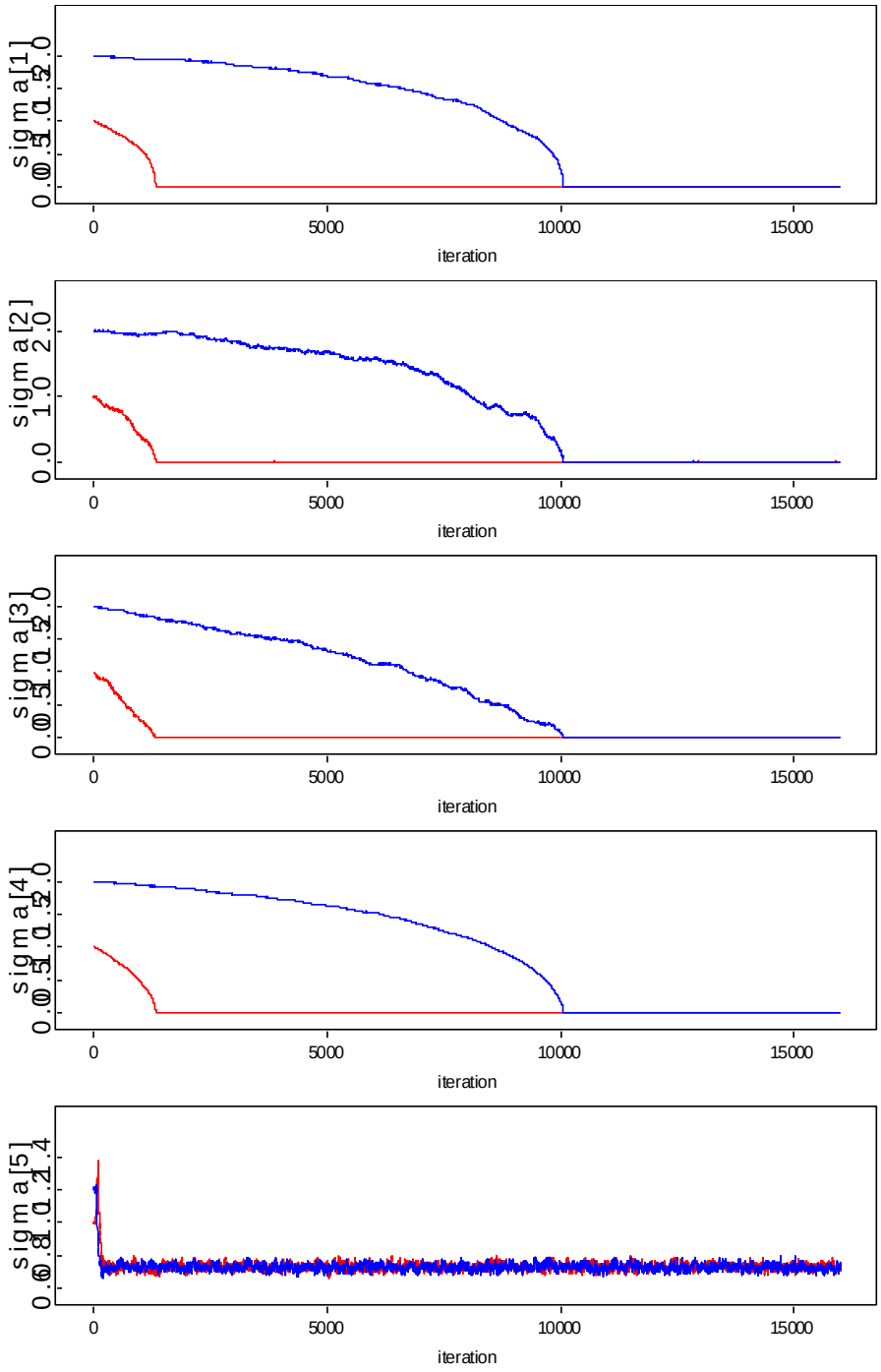

Figure 44. Convergence plots for σ_i.
1=AMI, 2=HF, 3=PN, 4=SCIP, 5=CAUTI

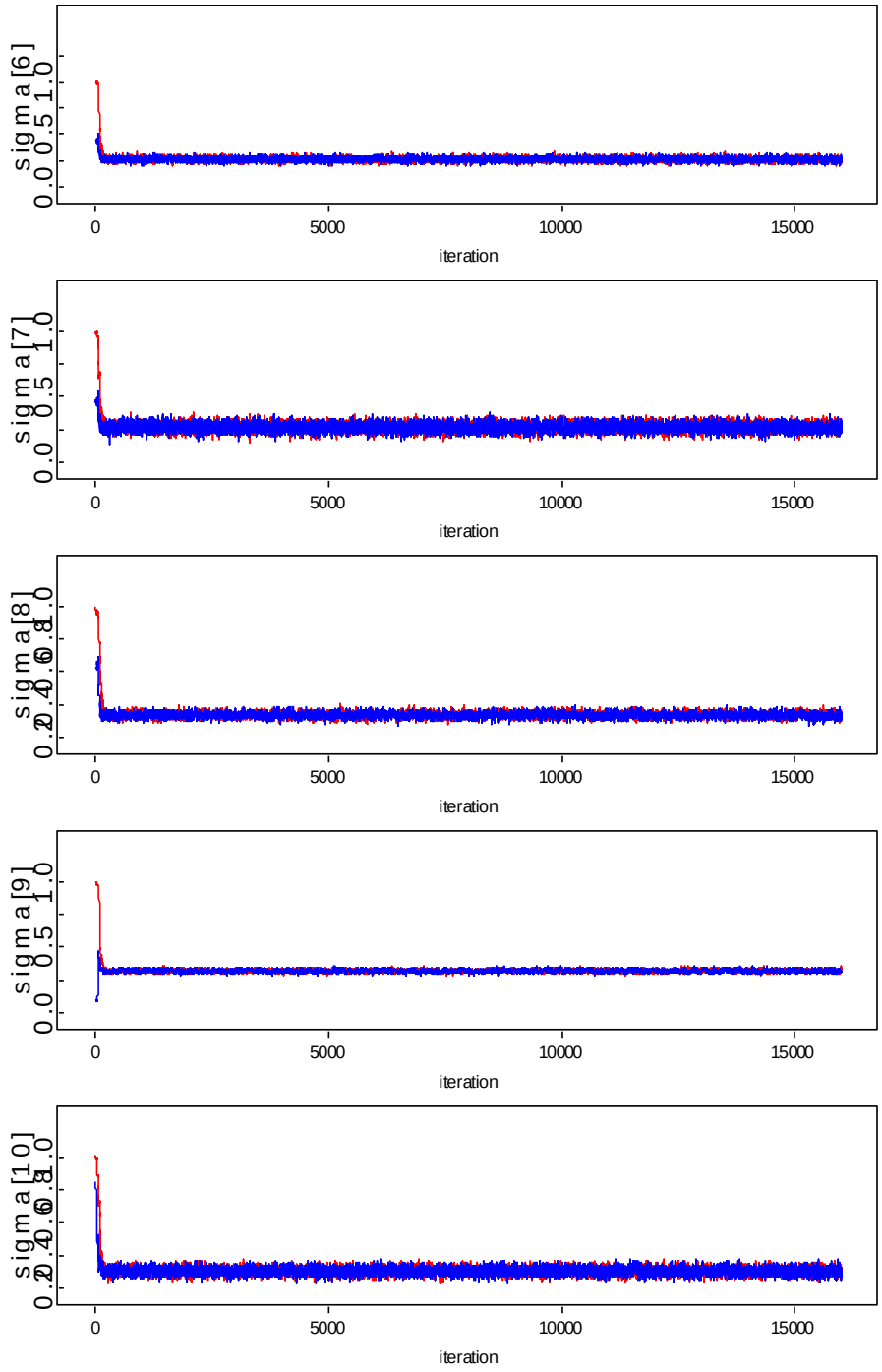

Figure 45. Convergence plots for σ_i.
6=SSI:Colon, 7=SSI:Abdominal, 8=CLABSI, 9=*C. difficile*, 10=MRSA

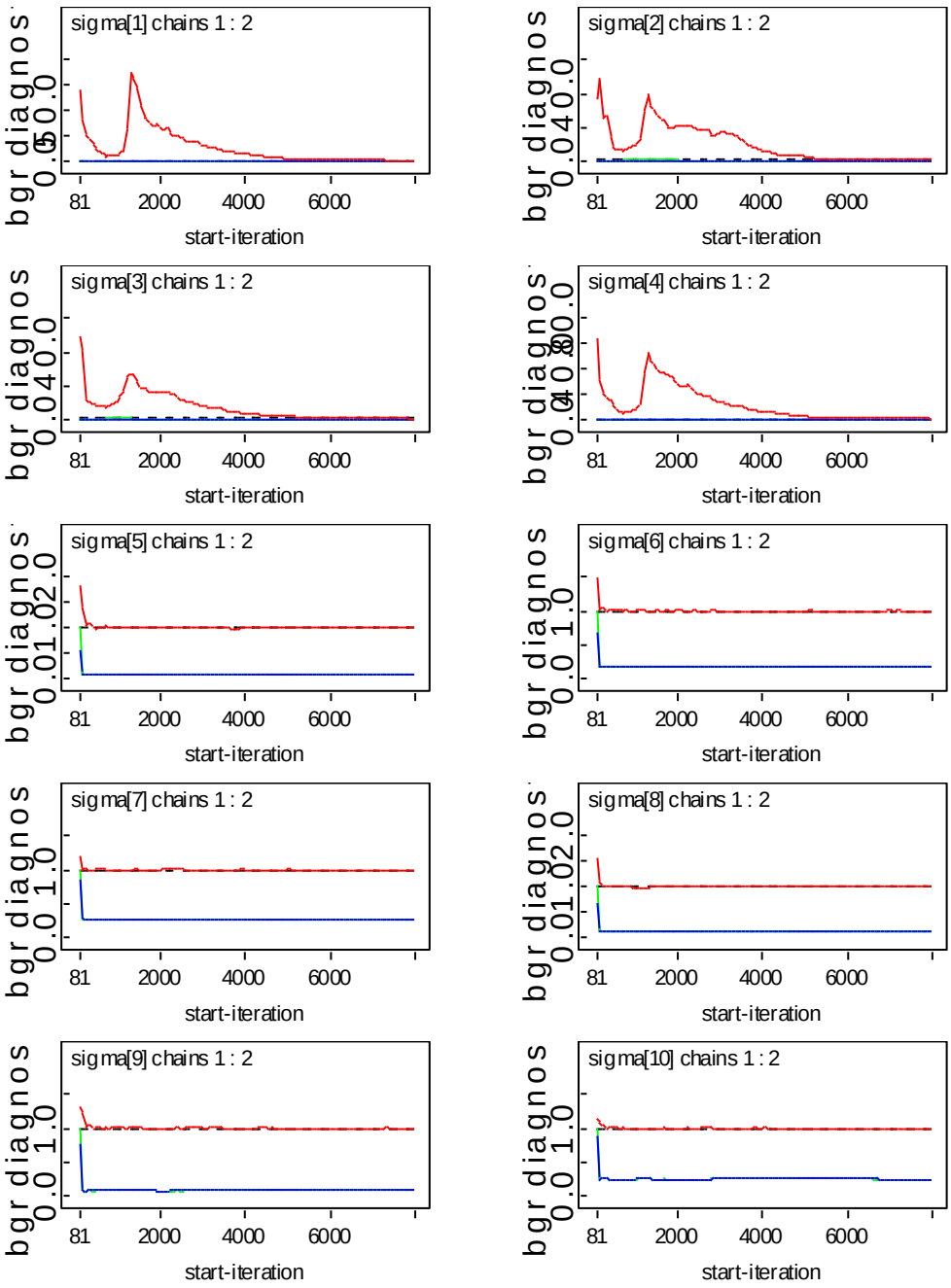

Figure 46. Gelman-Rubin statistics diagnostic plots of σ_i.
1=AMI, 2=HF, 3=PN, 4=SCIP, 5=CAUTI, 6=SSI:Colon,
7=SSI:Abdominal, 8=CLABSI, 9=*C. difficile*, 10=MRSA

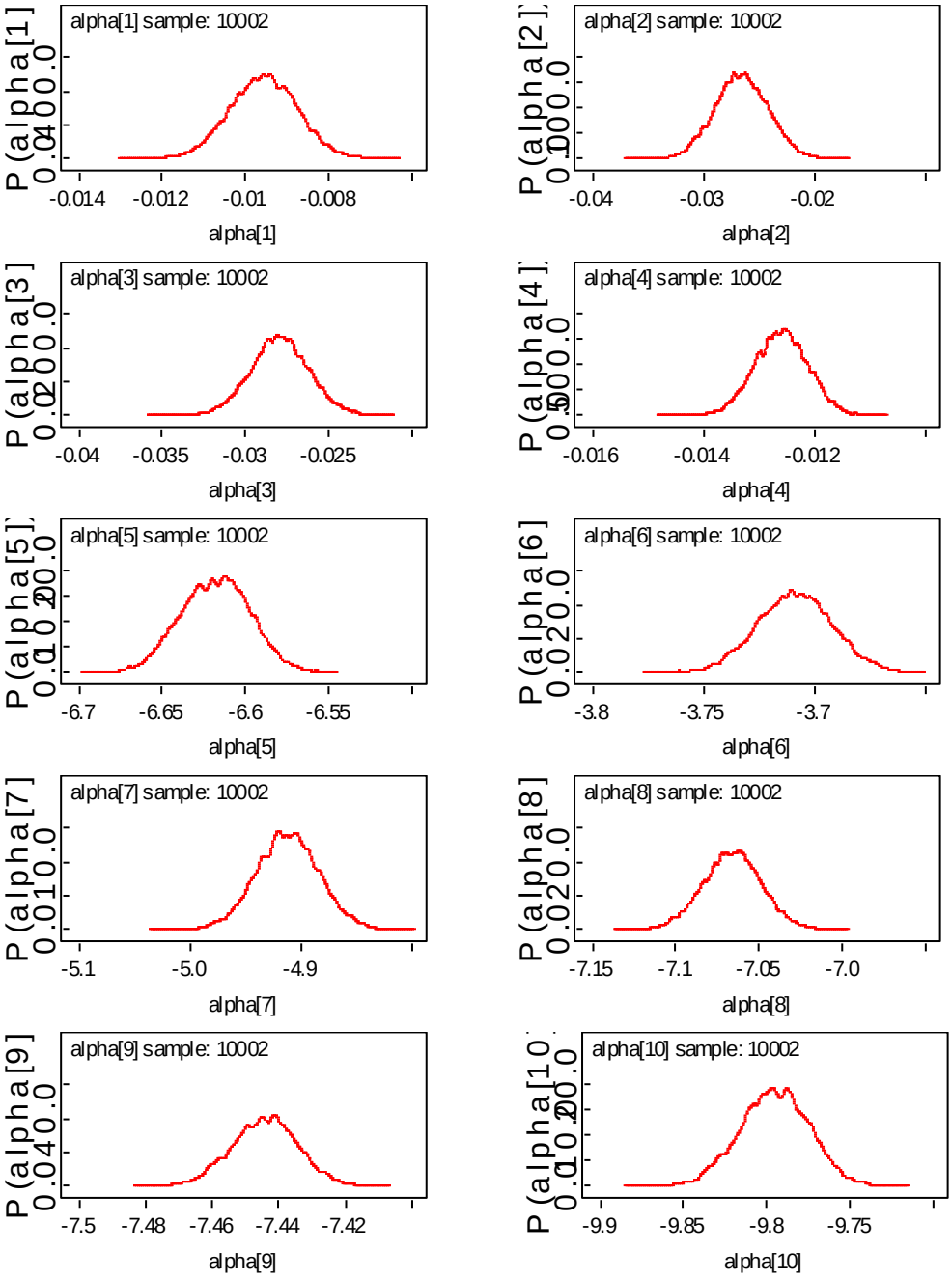

Figure 47. Density plots for α_i.
1=AMI, 2=HF, 3=PN, 4=SCIP,5=CAUTI, 6=SSI:Colon,
7=SSI:Abdominal, 8=CLABSI, 9=*C. difficile*, 10=MRSA

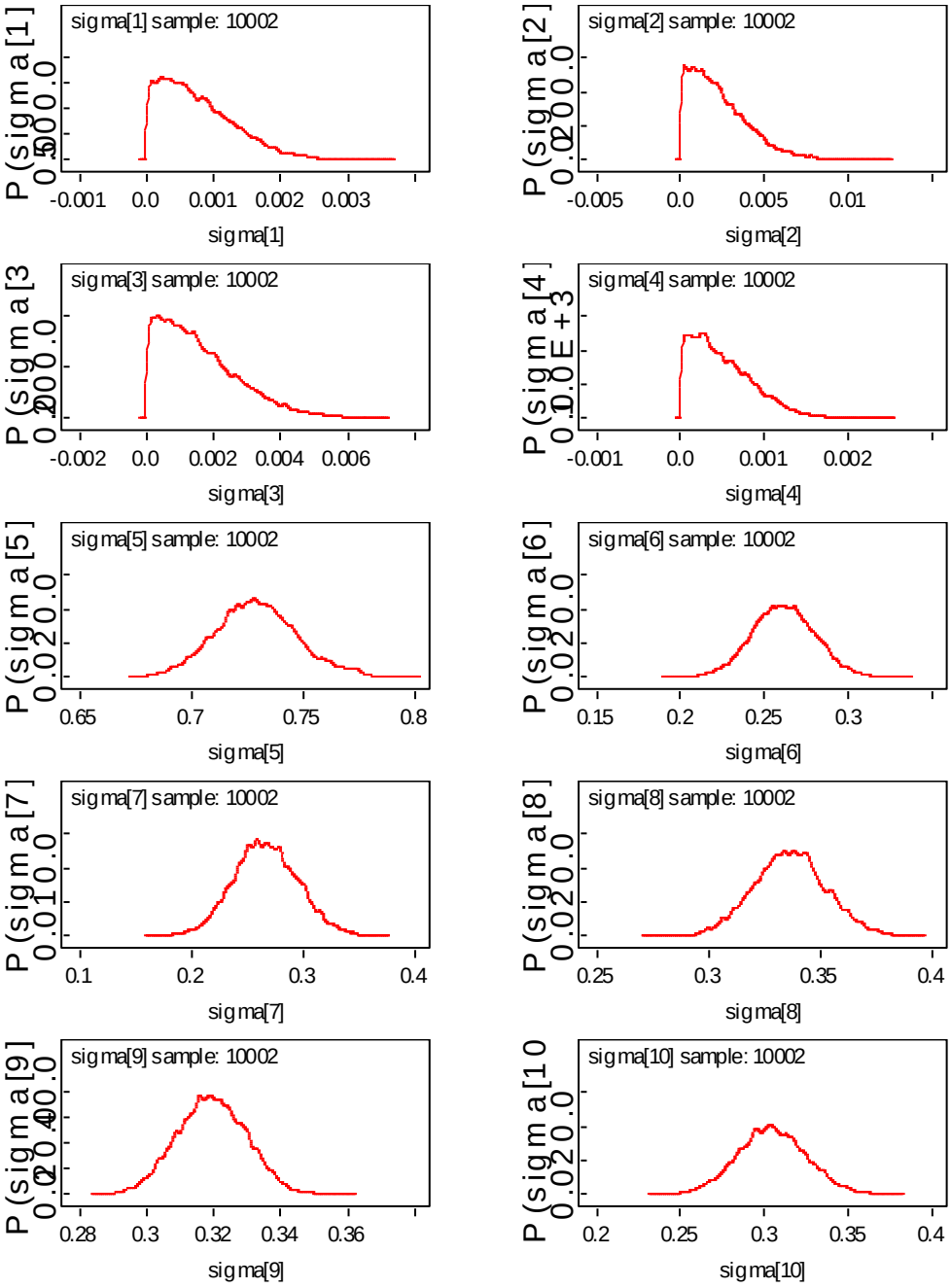

Figure 48. Density plots for σ_i.
1=AMI, 2=HF, 3=PN, 4=SCIP,5=CAUTI, 6=SSI:Colon,
7=SSI:Abdominal, 8=CLABSI, 9=*C. difficile*, 10=MRSA

TABLE XXI

PARAMETER ESTIMATES AND CREDIBLE INTERVALS

Measure	α Full Bayes Estimate (95% Credible Interval)	σ Full Bayes Estimate (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)

Parameter estimates with their 95% credible intervals are shown for each measure in Table XXI. Similar patterns with respect to α_i are seen in the bivariate case as in the univariate case. For the performance measures, AMI had the largest estimate of α along with the largest credible interval, while PN had the lowest α estimate. For 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. The density of the posterior estimates for θ_{pm} and θ_s is presented in Figure 49, which shows both posterior distributions are approximately normal with mean 0 and standard deviation of 1. Figure 50 plots the mean posterior estimates for each hospital for θ_{pm} and θ_s. The correlation between the two estimates is -0.05397, which is significantly different from 0 (p=.0078). Figure 51 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.

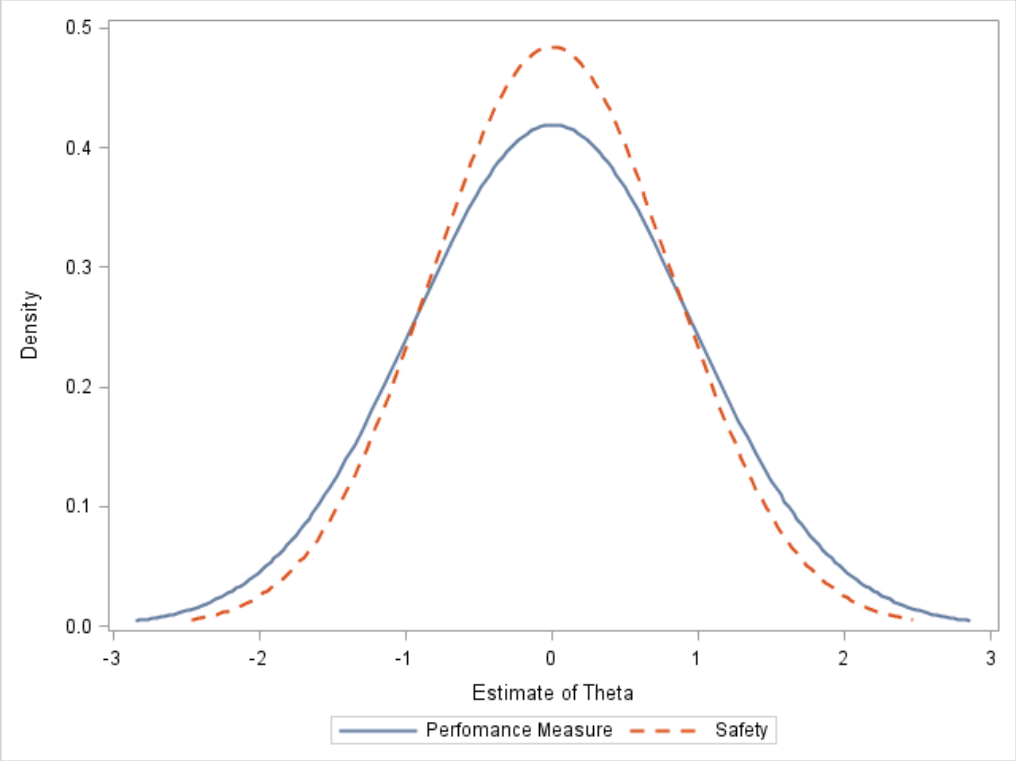

Figure 49. Posterior density graphs of θ_{pm} and $\theta_{s.}$

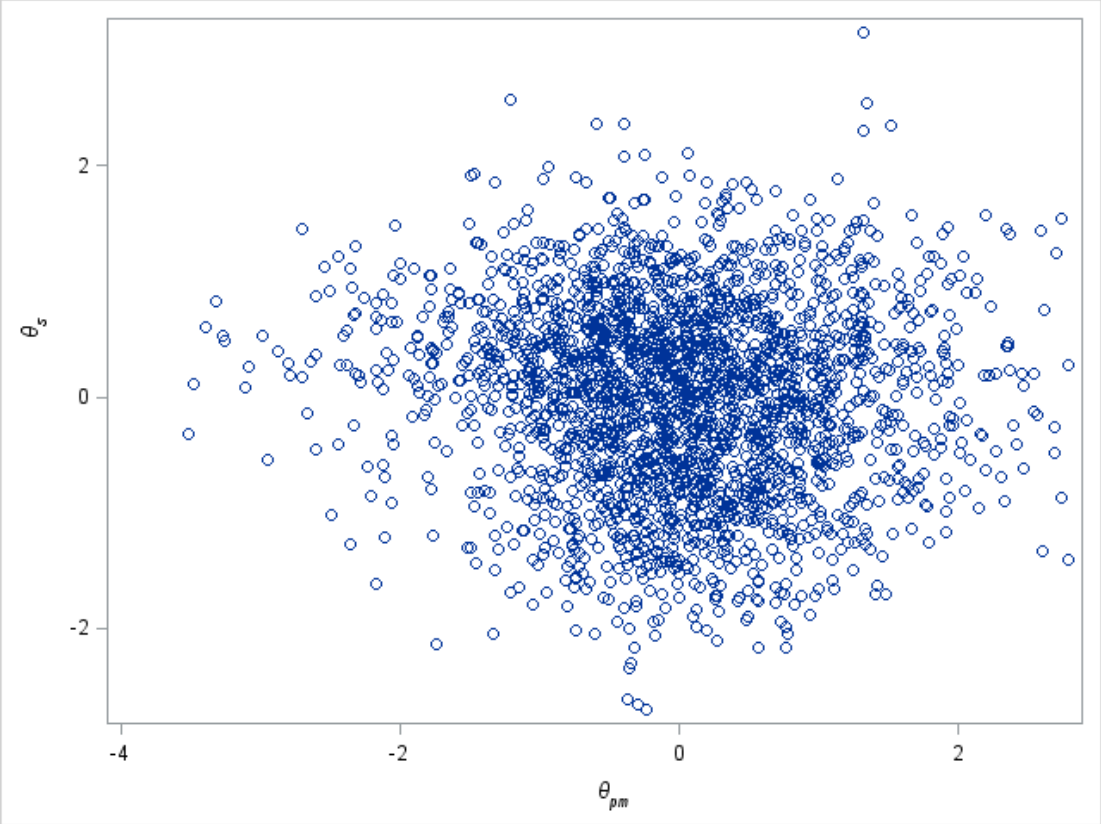

Figure 50. Plots of mean posterior estimates of θ_{pm} and $\theta_{s.}$

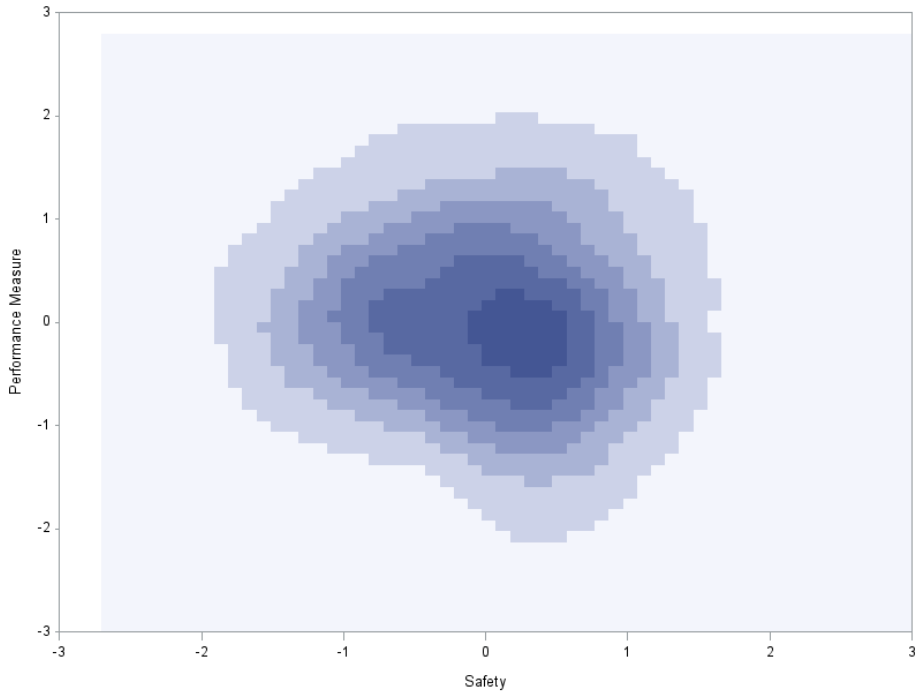

Figure 51. Contour plot of mean posterior estimates of θ_{pm} and $\theta_{s.}$

Plots of θ_{pm} versus p_{ij}, the observed measure composite rates, are displayed in Figure 52, which shows increasing values of θ_{pm} for increasing observed rates for each measure set. The Loess line included in those plots also indicates that higher performing hospitals are indicative of higher values of θ_{pm}. Similarly, Figure 53 shows plots of θ_s versus the observed number of infections of the safety infection measures. Increased values of θ_s are associated with lower numbers of infections for each of the measures in this study, which indicates that higher-quality hospitals have lower infections and infection rates.

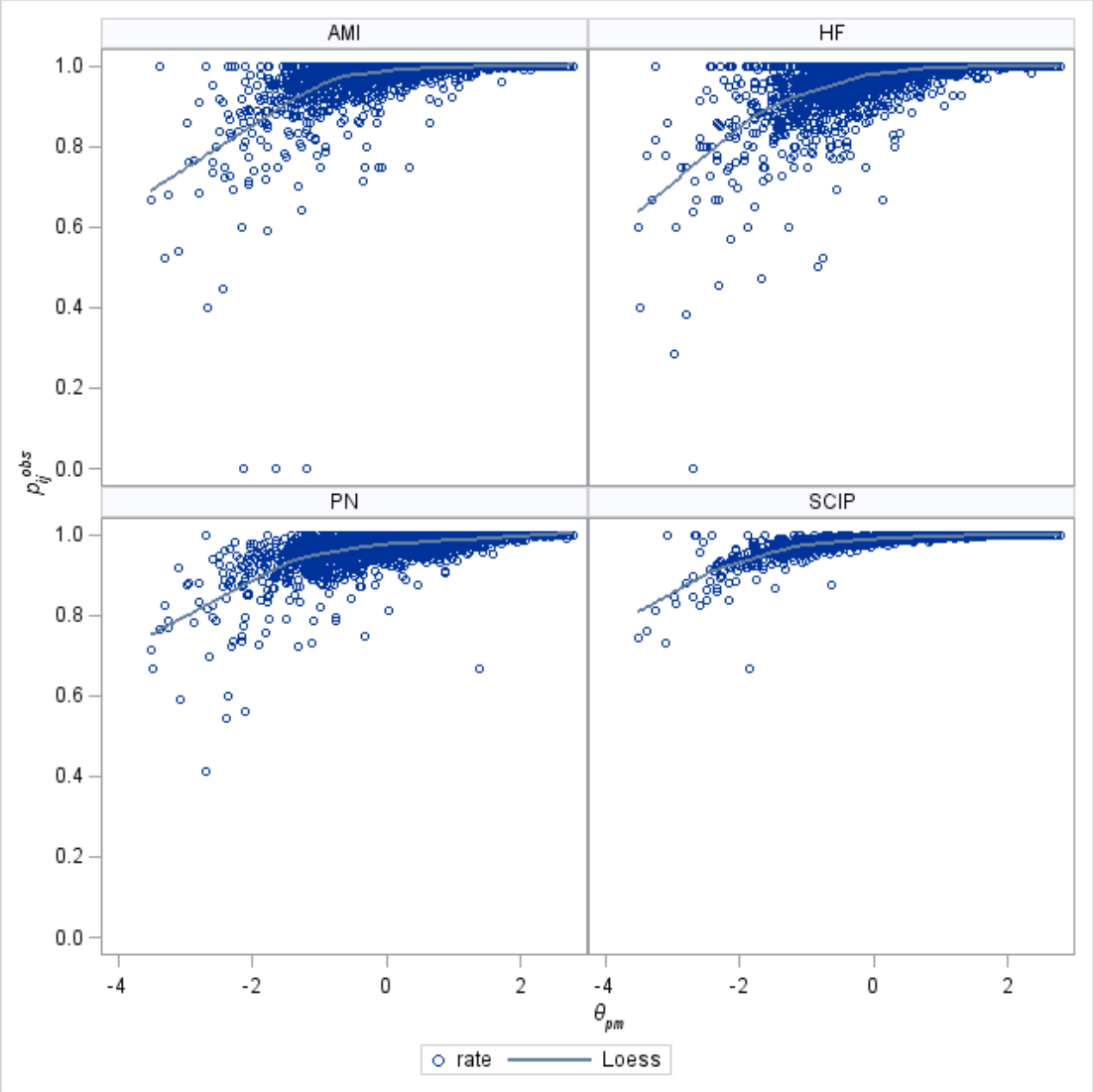

Figure 52. Plots of estimates of θ_{pm} versus observed performance measure rates.

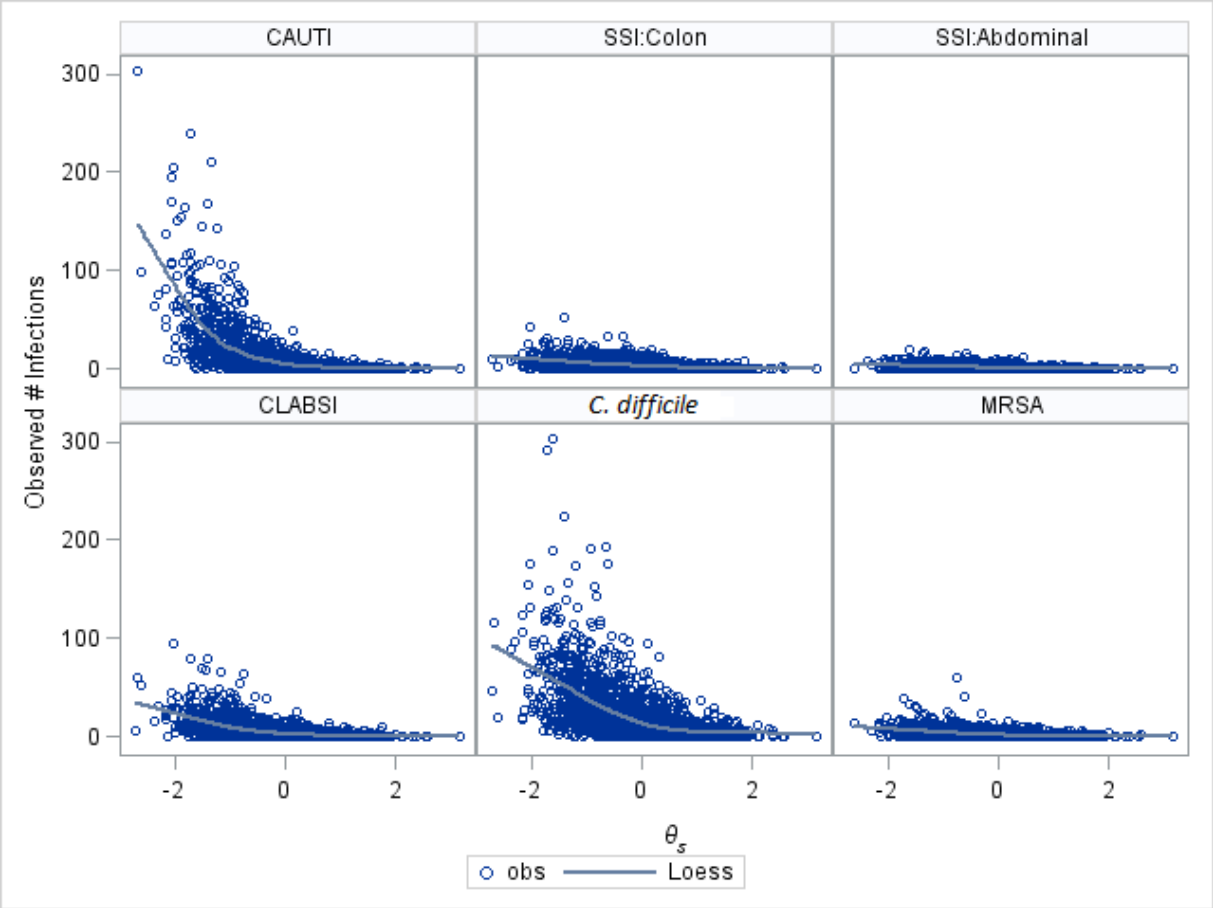

Figure 53. Plots of estimates of θ_s versus observed number of infections.

There are 121 hospitals 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 that ranged 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 that ranges from 0.806 to 2.24 with 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, 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 -.23 to -3.51, and the mean estimate for θ_s is -.72 with a maximum value of -.32 and a minimum value of -2.70. For the performance composite measures, the lower 1% of the hospitals identified had an 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.

7. DISCUSSION

7.1 **Binomial Regression Model**

As the world of hospital performance measurement evolves, innovative and robust statistical techniques are important, more than ever, in order to properly evaluate hospitals. The methods outlined in this study can be extended to other areas of interest such as education and behavioral sciences. Using static yearly data, the method outlined accurately models the overall composite hospital rate adjusting for the variation between the measure sets. The advantage of using composite measures relates to the cumulative effect of combining all opportunities within a measure set to an overall rate based on the accountability measures within the set. This enables better assessment of hospital care within a measure set for those hospitals with low patient volume in addition to giving a person a single number by which to assess those hospitals that are high performers. The exception is with HF because the composite measure for HF contains only one accountability measure. One issue involved with using composite rates is the loss of variation of individual measures within a measure set or therapeutic area.

I explored three models to classify hospitals using yearly static data, an FB model with a standard normal distribution assumption on the latent variable, an FB model with a rectangular distribution assumption on the latent variable, and an EB model with a standard normal assumption on the latent variable. The two models (FB and EB) with the standard normal prior distribution yielded similar results. Classification of hospital quality using FB is based on probabilities of the posterior distribution of the latent variable. Classifying hospitals based on the

EB model is based on using confidence intervals, which is comparable to the FB method utilizing a large number of hospitals and data elements. Since FB methods are easy to compute with programs such as OpenBUGS, classifying hospitals based on probabilities is easy to interpret and preferred over EB methods. The model using the rectangular prior produces inferences that lend to misclassifying hospitals due to the larger variation, i.e., a larger DIC, and more work in assessing more appropriate parameters of the rectangular distribution is needed.

Using the methods outlined in this study, I identify a fixed number of hospitals as being top-performing hospitals, i.e., the top 5% of all hospitals, as opposed to other organizations such as the Joint Commission that identify top performers based on a minimum threshold of raw observed rates. One limitation of using raw composite rates is that the variability between the measure sets is not taken into account, thus introducing higher error rates identifying top performers. This method takes this variability into account and gives a higher degree of confidence of correctly identifying top-performing hospitals. This methodology is extended to identify low-performing hospitals as well in order to provide motivation for improvement. As shown, hospitals with higher rates in the performance measures yielded higher latent scores and hospitals with lower performance measure rates are associated lower estimates of the latent score.

Within the framework of the longitudinal analysis, hospitals with declining measure rates over a three-year period based on quarterly composite rates are identified. For each of the measure sets, an average increase in composite measure rates is observed over the three-year period of evaluation, with HF and PN having the most opportunity for improvement. One

limitation in this study is that not all hospitals had enough longitudinal data in all measure sets as some hospitals were represented by just one measure set. Another limitation is utilizing quarterly data where the number of opportunities was not enough to be included in the study. This methodology is important as a surveillance system to identify those hospitals that are not maintaining an acceptable level of quality. Additionally, the current methods show hospitals with superior quality improvement efforts over time. With this information, better hospital performers will be recognized so other hospitals in need of improvement can learn from their experiences to increase their rates to an acceptable level.

7.2 **Poisson Regression Model**

Using a Poisson regression model for the hospital performance measure data is not advisable when the binomial rates are close to 100%. The Poisson distribution is better used for data where the occurrence of the outcome is a rare event, i.e., close to zero. Therefore, identifying top-performing hospitals based on the Poisson model is not an effective choice of models with performance measure data.

For the patient safety data, the Poisson model fits the data appropriately because infections rates are very low. Hospital safety infection data cannot be aggregated to a composite rate and therefore each measure is represented individually. Because of very small rates of infections among the hospitals used in this study, the assumptions of the Poisson model are a good choice to fit these data.

In this study, I have demonstrated that using a Poisson model is effective in identifying top-performing hospitals in the areas of patient safety. Those top-performing hospitals share characteristics of having a lower number of reported infections and therefore a lower rate of infections.

7.3 **Bivariate Model**

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 FB method to model the performance measure data with a binomial model and to model 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. By extending the concept of credible intervals to credible regions, I was able to identify top-performing hospitals that have both high performance 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, although there are some hospitals with high performance measure composite rates that were also identified as low performers.

7.4 Conclusion

Model-based scoring methods offer more advantages than methods based on point estimates, as the former method borrows strength from other measure sets. The proposed methodology, in the binomial, Poisson, or bivariate regression models, accounts for hospital variation in order to make proper comparisons and categorizations of hospitals. Extending this method to a longitudinal setting offers similar results in identifying top performers in quality improvement efforts.

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 FB approach is preferred. This final 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.

7.5 Beyond This Study

The work of identifying best hospital practices should not stop here. There are many different types of data that are collected by the hospitals and I am only scratching the surface of what is available. As healthcare organizations move into electronic medical records and new electronic measures (e-measures), more data will become available. Future work in modeling the data with extensions of the Poisson model such as the zero-inflation Poisson model, the over-

dispersed Poisson model, and the negative binomial model will be further explored. Future exploration of different prior distributions on the latent variables and weighting schemes in construction of the composite measures is also required. Another area of research is to compare classifications based on composite measures to classification of hospitals using models at the measure level in order to determine if composite measures yield the same inferences.

CITED LITERATURE

Alagumalai, Sivakumar., D. D. Curtis, and N. Hungi. *Applied Rasch Measurement: A Book of Exemplars. Papers in Honour of John P Keeves*. The Netherlands: Springer, 2005.

Albert, J. H., and S. Chib. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88 (1993): 669–679.

Austin, P. C. "A Comparison of Bayesian Methods for Profiling Hospital Performance." *Medical Decision Making* 22 (2002): 35–45.

Austin, P. C., D. C. Naylor, and J. V. Tu, "A Comparison of a Bayesian vs. a Frequentist Method for Profiling Hospital Performance." *Journal of Evaluation in Clinical Practice* 7 (2001): 35–45.

Awad, S. S. "Adherence to Surgical Care Improvement Project Measures and Post-Operative Surgical Site Infections." *Surgical Infection* 13, no. 4 (2012): 234–237.

Bedrick, E. J., R. Christensen, and W. Johnson. "A New Perspective of Priors for Generalized Linear Models." *Journal of the American Statistical Association* 91 (1996): 1450–460.

Bedrick, E. J., R. Christensen, and W. Johnson. "Bayesian Binomial Regression: Predicting Survival at a Trauma Center." *The American Statistician* 51 (1997): 211–218.

Berger, J. O. "A Robust Generalized Bayes Estimator and Confidence Region for a Multivariate Normal Mean." *Annals of Statistics* 8, no 4 (1980): 716–761.

Casey, D. E., Jr. "Performance Measurement 2.0: Time to Raise the Bar." *American Journal of Medical Quality* 25, no. 4 (2010): 246–248.

Chassin, M. R., J. M. Loeb, S. P. Schmaltz, and R. M. Wachter. "Accountability Measures—Using Measurement to Promote Quality Improvement." *New England Journal of Medicine* 363, no. 20 (2010): 683–688.

"CMS Quality Improvement Roadmap," accessed April 16, 2012, https://www.cms.gov/Medicare/Coverage/CouncilonTechInnov/downloads//qualityroadmap.pdf.

Cohen, J., M. Seburn, T. Antal, and M. Gushta. "Linking Rasch Scales across Grades in Clustered Samples." *American Institutes for Research* (2003), accessed November 23, 2014, http://assessment.air.org/psychometrics_docs/linking%20_rasch_scales.pdf.\\

Damien, P., J. Wakefield, and S. Walker. "Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variable." *Journal of the Royal Statistical Society* 61 (1990): 331–344.

Data from the National Hospital Discharge Survey. Accessed November 15, 2012, http://www.cdc.gov/nchs/data/nhds/4procedures/2010pro_numberpercentage.pdf.

Dellaportas, P., and J. Forster. "Markov Chain Monte Carlo Model Determination for Hierarchical and Graphical Log-Linear Models." *Biometrika* 86 (1999): 615–633.

Dellaportas, P., J. Forster, and I. Ntzoufras. "On Bayesian Model and Variable Selection Using MCMC." *Statistics and Computing* 12 (2002): 27–36.

Dellaportas, P., and A. Smith. "Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling." *Journal of the Royal Statistical Society* 42 (1993): 443–460.

Elixhauser, A., and C. Steiner. "Infections with Methicillin-Resistant *Staphylococcus aureus* (MRSA) in U.S. Hospitals, 2003–2005." *AHRQ Healthcare Cost and Utilization Project Statistical Brief* 35 (2007): 1–10.

"Facts on About Accountability Measures," accessed April 15, 2012, http://www.jointcommission.org/facts_about_accountability_measures/.

Gamerman, D., and H. Lopes. *Markov Chain Monte Carlo*. New York: Chapman and Hall, 2006.

Gelfand, A.E., S. E. Hills, A. Racine-Poon, and A. F. M. Smith. "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling." *Journal of the American Statistical Association* 85 (1990): 972–985.

Gelfand, A.E., and A. F. M. Smith. "Sampling-Based Approaches to Calculation Marginal Densities." *Journal of the American Statistical Association* 85 (1990): 398–409.

Gelfand, A. E., A. Smith, and T. M. Lee. "Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling." *Journal of the American Statistical Association* 87 (1992): 523–532.

Gelman, A. “Prior Distributions for Variance Parameters in Hierarchical Models.” *Bayesian Analysis* 1, no. 3 (2006): 515–533.

Gelman, A., and. D. B. Rubin. “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical Science* 7 (1992): 457–472.

Geman, S., and D. Geman. “Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1984): 721–741.

Gibbons, R. D., A. K. Amatya, C. H. Brown, K. Hur, S. M. Marcus, D. K. Bhaumik, and J. J. Mann. “Post-Approval Drug Safety Surveillance.” *Annual Review of Public Health* 31 (2010): 419–437.

Gilks, W., S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Suffolk, UK: Chapman and Hall, 1996.

Gilks, W., and P. Wild. “Adaptive Rejection Sampling for Gibbs Sampling.” *Journal of the Royal Statistical Society* 41 (1992): 337–348.

Givens, G. H., and. J. A. Hoeting. *Computational Statistics*. New York: Wiley, 2005.

Gould, C. V., C. A. Umscheid, R. K. Agarwal, G. Kuntz, and D. A. Pegues. “Guideline for Prevention of Catheter Associated Urinary Tract Infections 2009.” *Infect Control Hospital Epidemiology* 31 (2010): 319–26.

Green, P. “Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination.” *Biometrika* 82 (1995): 711–732.

Hans, C., A. Dobra, and M. West. “Shotgun Stochastic Search for Large P Regression.” *Journal of the American Statistical Association* 102 (2007): 507–516.

Hastings, W. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications.” *Biometrika* 57 (1970): 97–109.

Higdon, D. “Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications.” *Journal of the American Statistical Association* 93 (1998): 585–595.

Hulin, C. L., F. Drasgow, and C. K. Parsons. *Item Response Theory: Application to Psychological Measurement*. Homewood, IL: Dow Jones-Irwin, 1983.

"Improving America's Hospitals—The Joint Commission's Annual Report on Quality and Safety 2011," accessed September 13, 2011, http://www.jointcommission.org/accreditation/top_performers.aspx

Klevens, R. M., J. Edwards, C. Richards, T. Horan, R. Gayens, D. Pollock, and D. Cardo. "Estimating Health Care Associated Infections and Deaths in U.S. Hospitals, 2002." *Public Health Reports* 122 (2007): 160–166.

Klevens, R. M., J. R. Edwards, et al. "Estimating Health Care-Associated Infections and Deaths in U.S. Hospitals, 2002." *Public Health Reports* 122 (2007): 160–166.

Kuhn, H. "Statement on Pay for Performance Initiatives. Testimony before: Subcommittee on Health of the Committee on Ways and Means," accessed October 5, 2011, http://www.hhs.gov/asl/testify/t050315a.html.

Laird, N. M., and T. A. Louis. "Empirical Bayes Ranking Methods." *Journal of Educational Statistics* 14 (1989): 29–46.

Landrum, M. B., S. E. Bronskill, and S-L. T. Normand. "Analytic Methods for Constructing Cross-Sectional Profiles of Health Care Providers." *Health Services & Outcomes Research Methodology* 1, no. 1 (2000): 23–47.

Lin, L. A. "A Concordance Correlation Coefficient to Evaluate Reproducibility." *Biometrics* 45, no. 1 (1989): 255–268.

Lin, L. A. "Assay Validation Using the Concordance Correlation Coefficient." *Biometrics* 48, no. 2 (1992): 599–604.

Lockwood, J. R., T. A. Louis, and D. F. McCaffrey. "Uncertainty in Rank Estimation: Implications for Value-Added Modeling Accountability Systems." *Journal of Educational and Behavioral Statistics* 27 (2002): 255–270.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. "Winbugs—A Bayesian Modelling Framework: Concepts, Structure, and Extensibility." *Statistics and Computing* 10 (2000): 325–337.

Magill, S. S., W. Hellinger, J. Cohen, R. Kay, C. Bailey, B. Boland, D. Carey, J. de Guzman, K. Dominguez, J. Edwards, L. Goraczewski, T. Horan, M. Miller, M. Phelps, R. Saltford, J. Seibert, B. Smith, P. Starling, B. Viergutz, K. Walsh, M. Rathore, N. Guzman, and S. Fridkin. "Prevalence of Healthcare-Associated Infections in Acute Care Facilities." *Infect Control Hospital Epidemiology* 33, no. 3 (2012): 283–91.

"Making Healthcare Safer: Reducing Bloodstream Infections." CDC Vital Signs, accessed March 2011 at: http://www.cdc.gov/VitalSigns/HAI/index.html.

Malpiedi, P. J., K. D. Peterson, M. M. Soe, J. R. Edwards, R. D. Scott II, M. E. Wise, J. Baggs, S. H. Yi, M. A. Dudeck, K. E. Arnold, L., M. Weiner, C. A. Rebmann, A. Srinivasan, S. K. Fridkin, and L. C. McDonald. "2011 National and State Healthcare-Associated Infection Standardized Infection Ratio Report," accessed February 11, 2013, http://www.cdc.gov/hai/national-annual-sir/index.html.

"Medicare Pay for Performance (P4P) initiatives," CMS Office of Public Affairs, accessed October 5, 2011, http://www.cms.hhs.gov/apps/media/press/factsheet.asp?Counter1343.

Merller, J. "Perfect Simulation of Conditionally Specified Models." *Journal of the Royal Statistical Society B* 61 (1999): 251–264.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. "Equations of State Calculations by Fast Computing Machine." *Journal of Chemical Physics* 21 (1953): 1087–1092.

"National Action Plan to Prevent Healthcare-Associated Infections; Roadmap to Elimination—Executive Summary," CMS, accessed October 15, 2012, http://www.hhs.gov/ash/initiatives/hai/exec_summary.html.

Neal, R. "Slice Sampling." *The Annals of Statistics* 31 (2003): 705–767.

Normand, S-L. T., M. E. Glickman, and C. A. Gatsonis. "Statistical Methods for Profiling Providers of Medical Care: Issues and Applications." *Journal of the American Statistical Association* 92 (1997): 803–814.

Ntzoufras, I. *Bayesian Modeling using WinBUGS*. New York: Wiley, 2009.

O'Brien, S. M., F. H. Edwards, C. K. Haan, J. B. Rich, E. D. Peterson, S-L. T. Normand, E. R. DeLong, C. M. Shewan, R. S. Dokholyan, and D. Shahian. "Quality Measurement in Adult Cardiac Surgery: Part 2—Statistical Considerations in Composite Measure Scoring and Provider Rating." *Annals Thoracic Surgery* 83 (2007): 13–26.

O'Grady, N. P., M. Alexander, L. A. Burns, E. P. Dellinger, J. Garland, S. O. Heard, P. A. Lipsett, H. Masur, L. A. Mermel, M. L. Pearson, I. I. Raad, A. G. Randolph, M. E. Rupp, and S. Saint. "Guidelines for the Prevention of Intravascular Catheter-Related Infections." *Clinical Infectious Diseases* 52 (a) (2011): 1087–99.

Propp, J., and D. Wilson. “Exact Sampling With Coupled Markov Chains and Applications to Statistical Mechanics.” *Random Structures and Algorithms* 9 (1996): 223–252.

Robert, C., and G. Casella. *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 2004.

Sisson, S. “Trans-Dimensional Markov Chains: A Decade of Progress and Future Perspectives.” *Journal of the American Statistical Association* 100 (2005): 1077–1089.

Shen, W., and T. A. Louis. “Triple-Goal Estimates in Two-Stage Hierarchical Models.” *Journal of the Royal Statistical Society* 60 (1998): 455–471.

Tanner, M., and W. Wong. “The Calculation of the Posterior Distributions by Data Augmentation.” *Journal of the American Statistical Association* 82 (1987): 528–549.

Teixeira-Pinto, A., and S-L. T. Normand. “Statistical Methodology for Classifying Units on the Basis of Multiple-Related Measures.” *Statistics in Medicine* 27 (2008): 1329–1350.

van der Linden, W. J., and R. K. Hambleton. *Handbook of Modern Item Response Theory*. New York: Springer-Verlag, 1997.

Williams, S. C., R. G. Koss, D. J. Morton, and J. M. Loeb. “Performance of Top-Ranked Heart Care Hospitals on Evidence-Based Process Measures.” *Circulation* 114 (2006): 558–564.

Williams, S. C., S. P. Schmaltz, D. J. Morton, R. G. Koss, and J. M. Loeb. “Quality of Care in Hospitals as Reflected by Standardized Measures, 2002–2004.” *New England Journal of Medicine* 353, no. 3 (2005): 28–37.

Yi, M., J. R. Edwards, T. C. Horan, S. I. Berrios-Torres, and S. K. Fridkin. “Improving Risk-Adjusted Measures of Surgical Site Information for the National Healthcare Safety Network.” *Infection Control Hospital Epidemiology* 2, no. 10 (2011): 970–986.

APPENDICES

APPENDIX A

Sample of composite measure data and OpenBUGS code to fit the latent variable model. The complete data has 1,625 hospitals, each having four composite measures representing AMI, HF, PN, and SCIP. The variable *sumnum* refers to the number of time the appropriate therapy was provided within each therapeutic area; *sumden* refers to the total number of opportunities available to each therapeutic area; *set* refers the therapeutic area and *hconum* represents the hospital.

Sample of Static Year (3Q2012–2Q2013) Data:

sumnum[]	sumden[]	set[]	hconum[]
52	55	1	1
31	32	2	1
303	307	3	1
673	688	4	1
1456	1467	1	2
288	288	2	2
407	450	3	2
2135	2163	4	2
79	83	1	3
42	43	2	3
145	158	3	3
444	467	4	3

APPENDIX A (continued)

OpenBUGS Code – Static Year Analysis

```
model {
    for (i in 1:6500) {
       sumnum[i]~ dbin(p[hconum[i],set[i]],sumden[i])
       logit(p[hconum[i],set[i]]) <- a[set[i]] +b[set[i]]*theta[hconum[i]]    }
  for (j in 1:1625){
      theta[j] ~dnorm(0,1)
   }
   for (k in 1:4){
        a[k] ~dnorm(0,0.0001)
        b[k] ~dnorm(0,0.0001)I(0,)
   }
        betabm ~dnorm(0,0.0001)
        betaT  ~dnorm(0,0.0001)
        betaBL ~dnorm(0,0.0001)
        betaRural ~dnorm(0,0.0001)
}
```

APPENDIX A (continued)

SAS code - Static Year Analysis for EB analysis

```
proc nlmixed data=analysis empirical;
   parms alpha1=4 alpha2=4 alpha3=3 alpha4=3.5  beta1=1.2 beta2=1 beta3=1.2 beta4=2 ;

   if set=1 then
      eta = alpha1*measet1 + beta1*theta ;
   if set=2 then
      eta = alpha2*measet2 + beta2*theta ;
   if set=3 then
      eta = alpha3*measet3 + beta3*theta;
  if set=4 then
      eta = alpha4*measet4 + beta4*theta;

  pij= exp(eta)/(1+exp(eta));
  bounds beta1,beta2,beta3,beta4>0;

   model sumnum ~ binomial(sumden,pij);

  random theta~normal(0,1) subject=hcoid out=thetaout;
  predict pij out=predout;
run;
```

APPENDIX B

Sample of composite measure data and OpenBUGS code to fit the latent variable model. The complete data have 1,588 hospitals, each having four composite measures representing AMI, HF, PN, and SCIP. The variable *sumnum* refers to the number of time the appropriate therapy was provided within each therapeutic area; *sumden* refers to the total number of opportunities available to each therapeutic area; *set* refers the therapeutic area; *t* refers to the time period and *hconum* represents the hospital.

Sample Longitudinal Data:

sumnum[]	sumden[]	set[]	t[]	hconum[]
85	89	3	1	1
98	102	3	2	1
88	89	3	3	1
95	95	3	4	1
91	94	3	5	1
111	112	3	6	1
88	90	3	7	1
91	94	3	8	1
78	80	3	9	1
72	73	3	10	1
69	69	3	11	1
84	85	3	12	1

APPENDIX B (continued)

OpenBUGS Code – Longitudinal Analysis

```
model {
   for (i in 1:51672) {
      sumnum[i]~ dbin(p[hconum[i],set[i],t[i]],sumden[i])

    logit(p[hconum[i],set[i],t[i]]) <- a[set[i]]+a2[set[i]]*t[i] +b[set[i]]*theta1[hconum[i]]
+b2[set[i]]*theta2[hconum[i]]*t[i]
  }
 for (j in 1:1588){
    theta1[j] ~dnorm(0,1)
    theta2[j] ~dnorm(0,1)
  }
  for (k in 1:4){
      a[k] ~dnorm(0,0.0001)
      b[k] ~dnorm(0,0.0001)I(0,)
      a2[k]~dnorm(0,0.0001)
      b2[k]~dnorm(0,0.0001)I(0,)

  }
      betabm ~dnorm(0,0.0001)
      betaT  ~dnorm(0,0.0001)
      betaBL ~dnorm(0,0.0001)
      betaRural ~dnorm(0,0.0001)

}
```

APPENDIX B (continued)

SAS Code – Longitudinal Model

```
proc nlmixed data=long2 tech=nrridge empirical;
  parms alpha10=3.39 alpha20=2.5 alpha30=2.23 alpha40=2.77
     alpha11=.1 alpha21=.09 alpha31=.105 alpha41=.115
       beta10=.64 beta11=.091  gamma1=-.35 gamma2=.16 gamma3=.07 gamma4=.051
       beta20=.65 beta21=.11 beta30=.53 beta31=.08 beta40=.70 beta41=.05;

    if set=1 then
      eta=alpha10*measet1+alpha11*measet1*t+beta10*theta1+
            beta11*theta2*t;

    if set=2 then
      eta=alpha20*measet2+alpha21*measet2*t+beta20*theta1+
            beta21*theta2*t;

    if set=3 then
      eta=alpha30*measet3+alpha31*measet3*t+beta30*theta1+ beta31*theta2*t;

    if set=4 then
      eta=alpha40*measet4+alpha41*measet4*t+beta40*theta1+
            beta41*theta2*t;

      pijt= exp(eta)/(1+exp(eta));

      bounds beta10,beta11,beta20,beta21,beta30,beta31,beta40,beta41>0;

     model sumnum ~ binomial(sumden,pijt);
      random theta1 theta2~normal([0,0],[1,0,1]) subject=hcoid out=thetaout;
      predict pijt out=predout;

run;
```

APPENDIX C

OpenBUGS code for the bivariate model.

```
model {
      for (i in 1:9389){
        obs[i]~ dbin(p[hospnum[i],meas[i]],den[i])
        logit(p[hospnum[i],meas[i]]) <- alpha[meas[i]] +
                                          sigma[meas[i]]*theta[hospnum[i],areacode[i]]
          }

      for (i in 9390:23441)  {
            obs[i] ~ dpois(mu[hospnum[i],meas[i]])
           log(mu[hospnum[i],meas[i]]) <- alpha[meas[i]] +
                                    sigma[meas[i]]*theta[hospnum[i],areacode[i]] + log(den[i])
           }

       for (j in 1:2432) {
             theta[j,1:2] ~ dmnorm(beta[],prec.sigma[,])
          }

        for (k in 1:10)  {
            alpha[k] ~ dnorm(0,0.0001)
            sigma[k] ~ dnorm(0,0.0001)I(0,)
           }

beta[1] <-0
 beta[2] <-0


 prec.sigma[1:2,1:2] ~ dwish(Omega[,], 2)
 Sigma[1:2,1:2] <- inverse(prec.sigma[,])

 Omega[1,1] <- 0.0001
 Omega[2,2] <- 0.0001
 Omega[1,2] <- 0
 Omega[2,1] <- 0

}
```

VITA

NAME David J. Morton

EDUCATION: BS, Mathematics, Purdue University, West Lafayette, Indiana, 1989

MS, Mathematics, Purdue University, West Lafayette, Indiana, 1991

MS, Statistics, University of Chicago, Chicago, Illinois, 1998

TEACHING: Department of Mathematics and Department of Business, Elmhurst College, Adjunct Faculty, 2000–present

Department of Statistics, University of Chicago, course assistant, 1996–1998

Department of Mathematic, Purdue University, Teaching Assistant, 1990–1991

PUBLICATIONS: Albers, G. W., J. M. Yim, K. M. Belew, N. Bittar, C. R. Hattemer, B. G. Phillips, S. Kemp, E. A. Hall, D. J. Morton, and P. H. Vlasses. ”Status of Antithrombotic Therapy for Patients with Atrial Fibrillation in University Hospitals.” *Archives of Internal Medicine* 156 (1996): 2311–2316.

Ashesh, J., P. H. Vlasses, D. J. Morton, and J. L. Bauman. “Economic Impact of Dioxin Toxicity.” *Pharmacoeconomics* 12 (1997): 175–181.

Demirtas, H., A. Amatya, O. Pugach, J. Cursio, F. Shi, D. J. Morton, and B. Doganay. “Accuracy Versus Convenience: A Simulation-Based Comparison of Two Continuous Imputation Models for Incomplete Ordinal Longitudinal Clinical Trials Data.” *Statistics and Its Interface* 2, no. 4 (2009): 449–456.

Turpin, R., L. Darcy, R. G. Koss, C. McMahill, K. Meyne, D. J. Morton, J. Rodriguez, S. P. Schmaltz, P. Schyve, and P. Smith. “A Model to Assess the Usefulness of Performance Indicators.” *International Journal for Quality in Health Care* 8 (1996): 321–329.

Williams, S. C., R. G. Koss, D. J. Morton, S. P. Schmaltz, and J. M. Loeb. "Case Volume and Hospital Compliance with Evidence-Based Processes of Cardiovascular Care." *International Journal for Quality in Health Care* 20, no. 2 (2008): 79–87.

Williams, S. C., D. J. Morton, K. Jay, R. G. Koss, S. Schroeder, and J. M. Loeb. "Smoking Cessation Counseling in U.S. Hospitals: A Comparison of High and Low Performers." *JCOM* 12 (2005): 7.

Williams, S. C., S. P. Schmaltz, D. J. Morton, R. G. Koss, and J. M. Loeb. "Quality of Care in U.S. Hospitals as Reflected by Standardized Measures, 2002–2004." *N. Engl. J. Med.* 353 (2005): 255–264.