Project Number:BN2010

A Bayesian Analysis of the Parameter of the Noncentral Hypergeometric Distribution: An Application to Several Small Hospitals for Liver Scan Recipients by Race

A Major Qualifying Project

Submitted to the Faculty

of

Worcester Polytechnic Institute

in Partial Fulfillment of the Requirements

for the Degree of Bachelor of Science.

by

Yiwei Liu

Adam Tragellis

April 28, 2011

Approved:

Professor Balgobin Nandram, Project Advisor

Abstract

The concept of small area estimation is growing in the statistics field due to the demand for reliable small area estimates which can be used to influence government decision-making. Hospitals are an example of small areas used in such statistical studies. Previous studies have shown that white and black breast-cancer patients had received some significantly different patterns of care. However this evaluation could be inaccurate due to some of the data being from small areas. The goal of our project was to estimate the odds ratios of the two groups of breast-cancer patients that received a treatment known as a liver scan. We built two statistical models in a Bayesian setting. Two methods of parameter estimation were used and compared. The main goal was to analyze and compare the reliability of the two methods of estimation.

Acknowledgements

We would like to thank Professor Nandram for his consistent patience and guidance throughout this project. We would also like to thank Professor Bhadra and Ms. Adriana Hera for their helpful suggestions.

Contents

List of Tables

List of Figures

1 Introduction

The concept of small area estimation is growing in the statistics field due to the need for reliable small area statistics. Rao (2003) stated that recently the demand for reliable small area estimates has increased due to their growing use in formulating government policies, allocating of government funds, etc.[10]. The estimation involves developing estimates of parameters of the distribution of a given sub-population or "small area". It is an expensive and unrealistic procedure to gather data from an entire population for a statistical study which is why sub-populations are used. However, often times sample sizes may be too small to develop accurate estimates since small sample sizes can lead to unacceptably large standard errors for the parameter estimates. This issue leads to the need to "borrow strength" from related areas to find more accurate estimates for a given area or several areas. This leads to the need to incorporate and develop alternative estimation methods such as hierarchical Bayes estimation, which this project investigates [7].

An example of small areas that would be used in a statistical study to formulate government decisions are hospitals. Patients in a given hospital are considered a sub-population of all hospital patients. Small area estimation can use data based on specific treatments administered to patients in a hospital to estimate parameters of a distribution of interest. These estimates are used to make inferences about the population as whole. The population of interest in our study was breast-cancer patients categorized by race and whether or not they received one specific treatment known as a liver scan.

Our data set came from hospital data from a 1989 National Cancer Institute study of $7,781$ breast-cancer patients. Several types of treatments were examined to assess treatment modality (or a tendency to conform to a general pattern) and quality differences between black and white breast-cancer patients. Our experiment used the liver scan data from this

study, which involved 1,856 patients [6]. We developed a statistical model with this data in order to examine and compare two methods of estimating its parameters.

Potential concerns with the accuracy of small area estimates motivated us to use a Bayesian framework for our study. We implemented Bayesian methods with MATLAB to determine estimates for the parameters of the prior and posterior distributions of this Bayes hierarchical procedure. We did this with the goal of determining a valid method of obtaining parameter estimates from small areas to draw inferences from.

Chapter 1 gives a brief background of the breast cancer treatment study from which we obtained the data and a review of classical statistical procedures used to setup our experimental methods. Chapter 2 discusses the methods of hospital specific estimation in which we developed parameter estimates for each hospital separately. Chapter 3 discusses the methods used to develop estimates after pooling the data in order to reduce the variability in estimation. Chapter 4 gives a comparison and analysis of both estimation methods. Chapter 5 gives a summary of our methodologies and conclusions drawn from our results.

1.1 Dataset

The data we used for our experiment came from a 1989 study of breast-cancer patients treated in community hospitals written by Paula Diehr PhD et al. The cancer study assessed the relationship of race and patterns of care for 7,781 patients with breast cancer in 107 hospitals [6]. We narrowed our focus to 1,856 black and white patients from the 19 hospitals with the most black patients, and to one pattern of care known as a liver scan, as did Hollander and Wolfe (1999).

Liver scan is a specialized radiological procedure used to examine the liver to identify certain conditions or to assess the function of the liver. It is a type of nuclear radiologi-

cal procedure that uses a radioactive substance called a radionuclide. The radionuclide is absorbed by the liver tissue and then emits gamma radiation. The gamma radiation is detected by a scanner which processes the information into a picture of the liver. Physicians can measure the behavior of the radionuclide in the body during a liver scan and assess and diagnose various conditions such as tumors, abscesses, hematomas, organ enlargement, or cysts [1].

In the Diehr et al. (1989) study, the patterns of care were defined by an expert National Cancer Institute appointed committee. The committee defined it as less appropriate if a patient received a liver scan. This is because

> "Liver scans and CT scans are not routinely required for a patient with local or regional disease because the likelihood of finding an abnormality is low in the absence of abnormal liver chemistries or hepatomegaly (pp.951)."

The data is organized in section A of the appendix in Table 13 based on race and the event of receiving a liver scan [6].

1.2 Brief Review of Classical Statistical Procedures

The liver scan data is considered binomial data from two samples. A binomial distribution is the discrete probability distribution, $B(n, p)$, of the number of successes Y in a sequence of independent trials, n, with Y obtaining a success probability of $p = Y/n$. The number of occurrences of the treatment of interest, or successes, in a sequence of independent trials follows a binomial distribution. Binomial data can more easily be analyzed after being gathered into 2×2 contingency tables. Table 1 is an example of a 2×2 contingency table.

Table 1: 2x2 Contingency table for ith strata.

	Successes	Failures	Totals
Sample 1	Y_1	F_1	n_1
Sample 2	Y_2	F_2	n_2
Totals:	$Y = Y_1 + Y_2$	$F = F_1 + F_2$	$n = n_1 + n_2$

1.2.1 Fisher's Exact Test

Fisher's exact test is a classical method used to compare two sample proportions. It is used, when sample sizes are small, to analyze the relation between two or more categorical variables. Fisher's exact test can be conducted with analysis of a 2×2 contingency tables displaying categorical data, such as the outcomes of two treatments administered to two groups [2]. Table 2 is a 2×2 contingency table of data from hospital 1.

Table 2: 2x2 Contingency table for hospital 1.

Hospital 1	Liver scan	No liver scan	Totals
Black patients	4	9	13
White patients	12	34	46
Totals:	16	43	59

In this case, the two groups are the black patients and the white patients, and the two treatments are receiving a liver scan and not receiving a liver scan. A success is receiving a liver scan and each patient is a trial. For a given group, the proportion of interest is the number of patients who received a liver scan out of the total number of patients. Fisher's exact test can be used to test whether any difference observed in the proportions of black and white patients that received a liver scan is significant.

1.2.2 Mantel-Haenszel Test

The appropriate extension of Fisher's exact test used to analyze sets of 2×2 contingency tables is known as the Mantel-Haenszel test. For this particular analysis, the data from the

19 hospitals is used to construct 19 2×2 contingency tables. Each hospital's 2×2 table allows for one to analyze and compare the two independent proportions of black and white patients that received a liver scan within a hospital. We are interested in comparing each hospital's 2×2 table with one another. Table 3 shows how the data for hospital i is formed.

Table 3: 2x2 Contingency table for hospital i.

	Liver scan	No liver scan	Totals
Black patients	x_i	$n_{1i} - x_i$	n_{1i}
White patients	$n_i - x_i$	$n_{2i} - n_i - x_i$	n_{2i}
Totals:	n_i	$N_i - n_i$	N_i

In our model, $\hat{p}_1^{(i)} = \frac{x_i}{n_{1i}}$ is the probability that a black patient in hospital $i (i = 1, 2, \ldots, 19)$ will receive a liver scan and $\hat{p}_2^{(i)} = \frac{n_i - x_i}{n_{2i}}$ is the probability that a white patient in hospital $i (i = 1, 2, \ldots, 19)$ will receive a liver scan. The Mantel-Haenszel procedure tests the null hypothesis, that within each hospital i the probabilities of receiving a liver scan are equal. Thus our hypotheses are constructed as follows:

$H_o : p_1^{(1)} = p_2^{(1)}, p_1^{(2)} = p_2^{(2)}, \ldots, p_1^{(19)} = p_2^{(19)};$

$H_a : p_1^{(1)} \geq p_2^{(1)}, p_1^{(2)} \geq p_2^{(2)}, \ldots, p_1^{(19)} \geq p_2^{(19)}.$

Alternatively we can test the null hypothesis that the probabilities $p_1^{(i)}$ and $p_2^{(2)}$ have a common odds ratio of 1, against the alternative hypothesis, that the common odds ratio differs from 1 [12]. The odds of a black patient receiving a liver scan is $\frac{p_1^{(i)}}{1-p_1^{(i)}}$ and the odds of a white patient receiving a liver scan is $\frac{p_2^{(i)}}{1-p_2^{(i)}}$. If we let $\Gamma_i = \frac{p_1^{(i)}}{1-p_1^{(i)}} / \frac{p_2^{(i)}}{1-p_2^{(i)}}$, the odds ratio for the i^{th} table, our hypotheses can be rewritten as :

$H_o : \Gamma_i = 1; i = 1, \ldots, 19.$

$H_a : \Gamma_i \geq 1, \ i = 1, \ldots, 19.$

The procedure of hypothesis testing is well known but in general estimation is considered to be more difficult. Estimation is less favored because there are no analytical forms for estimation. Our contribution of this project is our discussion of implementing different statistical analysis to estimate Γ_i. In either methods, we need to fit the observed data with a distribution. Our choice of distribution is noncentral hypergeometric distribution which we will explain in details in the next section.

1.2.3 Noncentral Hypergeometric Distribution

A noncentral hypergeometric distribution can be used to make inference on 2×2 contingency tables. Usually, the number of subjects assigned to the two different treatment groups are denoted as n_1, and n_2, respectively, and the counts of success are denoted as Y_1 and Y_2 respectively. Here $Y_i \sim Binomial(n_i, p_i)$, independent binomial distributions with probability for success p_i for $i = 1, 2$. For this design, the odds ratio $\Gamma = \frac{p_1(1-p_2)}{p_2(1-p_1)}$ is often used to measure the relationship between treatment outcomes and treatment groups. Let n be the sum of observed values of Y_1 and Y_2. Then the conditional distribution of Y_1 given $Y_1 + Y_2 = n$ follows a noncentral hypergeometric distribution [9].

$$P(Y_1 = x | Y_1 + Y_2 = n) = \frac{\binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x}{\sum_{x \in \mathcal{L}} \binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x}, x \in \mathcal{L}$$

where $\mathcal{L} : \{x : max(0, n - n_2) \leq x \leq min(n_1, n)\}$.

Based on the experimental data we used for our project, Y_{1i} and Y_{2i} represent the counts of black and white patients who received a liver scan for breast cancer respectively in the i^{th} hospital. It follows that:

$$P(Y_{1i} = x_i | Y_{1i} + Y_{2i} = n_i) = \frac{\binom{n_{1i}}{x_i}\binom{n_{2i}}{n_i - x_i}\Gamma_i^{x_i}}{\sum_{x_i \in \mathcal{L}} \binom{n_1 i}{x_i}\binom{n_2 i}{n_i - x_i}\Gamma_i^{x_i}}, x_i \in \mathcal{L}_i$$

where $\mathcal{L}_i : \{x_i : max(0, n_i - n_{2i}) \leq x_i \leq min(n_{1i}, n_i)\}$ and Γ_i is i^{th} odds ratio $\frac{p_{1i}(1-p_{2i})}{p_{2i}(1-p_{1i})}$.

Inferences drawn about the difference between treatment groups with respect to treatment outcomes are based on the odds ratios Γ_i. If the credible intervals of the odds ratios contain 1, this implies that the two groups are not significantly different. For example, if the credible interval of odds ratio for one hospital is $[0.800, 1.200]$, then the odds of the black patients receiving a liver scan and the odds of the white patients receiving a liver scan are not significantly different. If 1 is not contained in the credible intervals, this indicates a significant difference of the odds of the two groups receiving a scan. This leads to the assumption that if the odds ratios, Γ_i, are significantly greater than 1 then the black patients have greater odds of receiving a liver scan in the i^{th} hospital. As a result, we are interested in determining what is the best way to estimate Γ_i given the data from the local hospitals. We are also interested in inferences can be drawn from Γ_i from combined hospital data by borrowing strength across hospitals.

1.3 Proposed Methodology

1.3.1 Bayesian Analysis

Our procedure differs from the classic Mantel-Haenszel test which uses marginal densities with set parameters. The parameters in our model were not prespecified but estimated with Bayesian methods. In a model where y is the given continuous data and θ is the parameter,

based on Bayes' rule, we have

$$\begin{aligned} p(\theta|y) &= \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} \\ &= \frac{p(y|\theta)p(\theta)}{\int_\theta p(\theta, y)} \\ &= \frac{p(y|\theta)p(\theta)}{\int_\theta p(y|\theta)p(\theta)} \\ &\propto p(y|\theta)p(\theta) \end{aligned}$$

where $p(\theta)$ is called the prior density, $p(\theta|y)$ is referred to as the posterior density and $p(y|\theta)$ is the sampling distribution.

In our project, the marginal density of interest is $p(\Gamma|x)$. We also have a specific sampling distribution for each hospital as discussed in section 1.2.3. The problem became finding a proper prior density for the parameters Γ.

1.3.2 Constructing the Prior

Remembering that our parameter of interest Γ is the odds ratio of the odds of a black patient receiving liver scan to the odds of a white patient receiving liver scan, based on this knowledge, we established our model as follows:
assume p_1, p_2 are independent and $p_i \sim U(0,1), i = 1, 2$, where p_1 and p_2 are the probability of a black patient receiving a liver scan and the probability of a white patient receiving liver scan respectively.

Let $y_1 = \frac{p_1}{1-p_1}, y_2 = \frac{p_2}{1-p_2}$, then the probability density function of y_i $(i = 1, 2)$ are

$$f(y_i) = \begin{cases} \frac{1}{(1+y_i)^2} & \text{if } 0 < y_i < \infty, \\ 0 & \text{if } y_i = 0. \end{cases}$$

Thus, the cumulative density function of $\Gamma = \frac{p_1(1-p_2)}{p_2(1-p_1)}$ is

$$\begin{aligned} F_\Gamma(r) &= P\{\Gamma \le r\} = P\{\frac{y_1}{y_2} \le r\} \\ &= \int_0^\infty \int_0^{ry_2} \frac{1}{(1+y_1)^2} \frac{1}{(1+y_2)^2} dy_1 dy_2 \\ &= \int_0^\infty \frac{1}{(1+y_2)^2} \frac{ry_2}{(1+ry_2)} dy_2. \end{aligned}$$

So the probability density function of Γ is

$$\begin{aligned} f_\Gamma(r) &= \frac{\partial F_\Gamma(r)}{\partial r} \\ &= \frac{\partial}{\partial r}[\int_0^\infty \frac{1}{(1+y_1)^2} \frac{ry_2}{(1+ry_2)} dy_2] \\ &= \int_0^\infty \frac{1}{(1+y_2)^2} \frac{y_2}{(1+ry_2)^2} dy_2. \end{aligned}$$

If we transform y_2 to a bounded variable ϕ, where $y_2 = \frac{\phi}{1-\phi}$ and $dy_2 = \frac{1}{(1-\phi)^2} d\phi$, then the density of the odds ratio Γ could be rewritten as the following:

$$\begin{aligned} f_\Gamma(r) &= \int_0^\infty \frac{1}{(1+y_2)^2} \frac{y_2}{(1+ry_2)^2} dy_2 \\ &= \int_0^1 \frac{1}{(1+\frac{\phi}{1-\phi})^2} \frac{\frac{\phi}{1-\phi}}{(1+r\frac{\phi}{1-\phi})^2} \frac{1}{(1-\phi)^2} d\phi \\ &= \int_0^1 \frac{\phi(1-\phi)}{(1-\phi+r\phi)^2} (1) d\phi \\ &= E[\frac{\phi(1-\phi)}{(1-\phi+r\phi)^2}] \end{aligned}$$

where expectation is taken over $\phi \sim U(0,1)$ and $E[\phi] = 0.5$. Using a first-order Taylor expansion, we have the following approximation,

$$f(r) \approx \frac{E[\phi]E[1-\phi]}{(1-E[\phi]+rE[\phi])^2}]$$
$$= \frac{1}{(1+r)^2}, r \geq 0$$

As a result, we speculated that $\frac{1}{(1+r)^2}$ could be a good candidate for our approximation for the prior density of Γ.

To verify this, we compared the "Riemann summation", a method for approximating integrals, of $\int_0^\infty \frac{1}{(1+y_2)^2}\frac{y_2}{(1+ry_2)^2}dy_2$ to $\frac{1}{(1+r)^2}$. Below is the formula for the "Riemann summation" approximation for our problem.

$$\int_0^1 \frac{\phi(1-\phi)}{(1-\phi+r\phi)^2}d\phi = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{\phi^{(k)}(1-\phi^{(k)})}{(1-\phi^{(k)}+r\phi^{(k)})^2}, \phi^{(k)} = \frac{k}{n}.$$

When $n = 1000, \phi^{(k)} = 0.0005, 0.015, \ldots, 0.9995$.

The difference between the numerical function value of the "Riemann sum"of the integral and that of $\frac{1}{(1+r)^2}$ turned out to be minimal, as is shown in Figure 1. We also plotted the function value of our approximation with respect to the "Riemann sum", which is Figure 2. The plot is almost a straight line with a slope of 1. We also used a Monte Carlo method to check our results which were in accordance with the Riemann summation.

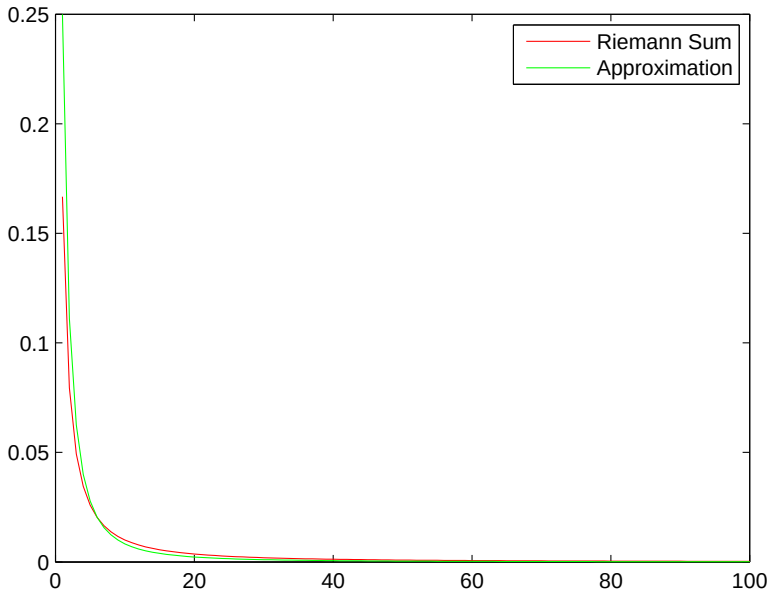


Figure 1: Comparison between Riemann sum and the approximation

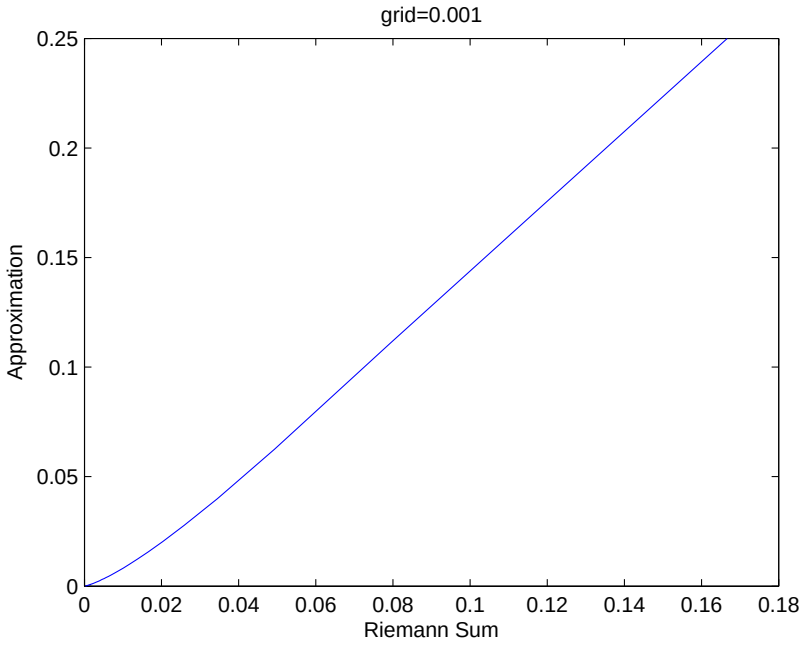


Figure 2: Riemann sum vs approximation

2 Hospital Specific Estimation

In this section, we studied the data from each hospital separately. The goal was to generate random samples of each parameter Γ_i from its respective posterior distribution and make inferences about Γ_i.

2.1 The Grid Method

As discussed in section 1.2.3, our sampling distribution has the following form:

$$f(x|\Gamma) = \frac{\binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x}{\sum_{x\in l}\binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x}, \tag{1}$$

and we constructed the prior distribution for Γ in section 1.3.2, which is $P(\Gamma) = \frac{1}{(1+\Gamma)^2}$.

Bayes' theorem provides that $f(\Gamma|x) \propto f(x|\Gamma)f(\Gamma)$

So the posterior distribution of Γ is:

$$f(\Gamma|x) = \frac{\binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x}{\sum_{x\in l}\binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x}\frac{1}{(1+\Gamma)^2}\ ,\Gamma > 0. \tag{2}$$

For computational convenience, we made a change of variable of Γ. Letting $\theta = \frac{\Gamma}{1+\Gamma}$, so $\theta \in (0,1)$. After Jacobian transformation, the posterior distribution of interest becomes

$$f(\theta|x) = \frac{\binom{n_1}{x}\binom{n_2}{n-x}(\frac{\theta}{1-\theta})^x}{\sum_{x\in l}\binom{n_1}{x}\binom{n_2}{n-x}(\frac{\theta}{1-\theta})^x}\ . \tag{3}$$

From this point on we used θ instead of Γ for computations since θ is more stable than Γ because it is bounded in $(0,1)$.

2.1.1 Data Simulation in Grid Method

The inverse CDF method tells us that the cumulative density function of a random variable follows a Uniform distribution in $[0, 1]$. Because our parameter θ is bounded between $[0, 1]$, we could take advantage of the grid methods to draw a sample of $10,000$ numbers from a specific posterior distribution of θ_i. We started by partitioning the interval of (0,1) into 100 segments and number the middle point of each sub-interval as m_i, for $i = 1, ..., 100$ and calculate the cumulative probability of each m_i, denoted by $F(m_i|x)$ in Table 4. The procedure of drawing numbers goes as follows: we drew a sequence of 10,000 numbers from a uniform distribution u_i, for $i = 1, ..., 10,000$ first. Then we compare each u_i to the cumulative probability $F(n_i)$ at each discrete grid. When $u_i \leq F(\phi_i) < u_{i+1}$, we pick up m_{i+1} and because of the inverse CDF method, this sequence of grids m_i we generated follows the posterior distribution $f(\theta|x)$.

Table 4: The grid method

m_i	$F(m_i\|x)$	u_i
m_1	$F(m_1)$	u_1
m_2	$F(m_2)$	u_2
...	...	...
$m_{10,000}$	$F(m_{10,000}) = 1$	$u_{10,000}$

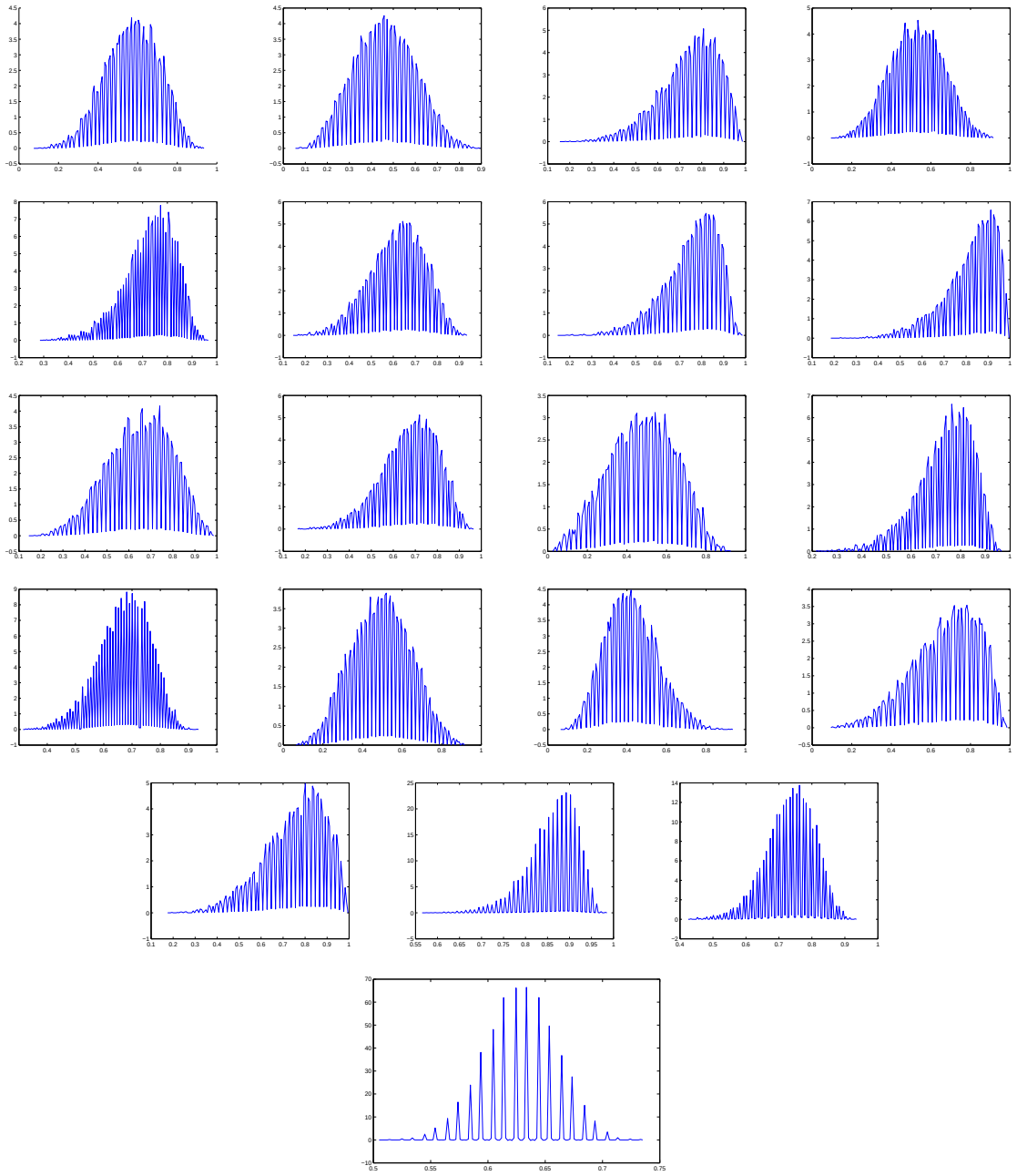
Table 5 lists the means, standard deviations and 95% credible intervals of θ for 19 hospitals. Since our research goal is to find a common distribution for the odds ratio Γ, we combined the data from all the hospitals and used grid method to estimate the overall mean. The summary of the combined data is listed in the last row of the Table 5.

Table 5: Summary of hospital specific estimation using grid method.

Hospital	Mean of θ	Standard deviation of θ	95% Credible Interval
1	0.573	0.136	[0.295, 0.815]
2	0.463	0.139	[0.195, 0.735]
3	0.746	0.135	[0.435, 0.945]
4	0.532	0.137	[0.215, 0.785]
5	0.732	0.104	[0.495, 0.895]
6	0.620	0.126	[0.355, 0.835]
7	0.758	0.123	[0.465, 0.935]
8	0.825	0.119	[0.525, 0.975]
9	0.652	0.146	[0.345, 0.905]
10	0.677	0.127	[0.395, 0.885]
11	0.480	0.170	[0.145, 0.785]
12	0.739	0.116	[0.465, 0.915]
13	0.673	0.093	[0.475, 0.835]
14	0.499	0.148	[0.215, 0.785]
15	0.426	0.131	[0.195, 0.705]
16	0.678	0.165	[0.315, 0.925]
17	0.757	0.139	[0.445, 0.965]
18	0.863	0.058	[0.725, 0.945]
19	0.736	0.074	[0.575, 0.865]
1-19	0.629	0.032	[0.565, 0.685]

The following is the kernel density estimation of histograms of the data of θ for the 19 hospitals and the combined data.

Figure 3: Kernel density plots of θ using hospital specific estimation.



As is shown in the Table 5, hospitals 2, 11, 14, and 15 have a mean less than 0.5. Among the 19 credible intervals for the hospital data, only 3 of them, hospitals 8, 18, 19, did not contain the value 0.5. These three intervals show that the estimates of θ for these hospitals are significantly greater than 0.5, since their lower bound are all greater than 0.5. An estimate for θ greater than 0.5 translates to an estimate of the odds ratio Γ greater than 1. This indicates that these 3 hospitals do in fact have odds ratios greater than 1, which means the two groups in these hospitals have significantly different odds of receiving a liver scan. Specifically, the black patients have higher odds of receiving a liver scan. Also, the credible interval of the combined data from all the hospitals, $[0.5650, 0.6850]$, does not contain 0.5 which suggests that there was overall a difference in treatment between the black patients and the white patients. Hollander and Wolfe (1999) used the same hospital data to perform a Mantel-Haenszel test and got a p-value $P \ll 0.0002$ which is strong evidence that the common odds ratio among the hospitals is not equal to 1. This also suggests that the odds that black patients get a liver scan is greater than the odds that white patients do.

2.2 Validating the Grid Method

We wanted to show that this way of regenerating θ as shown in the previous section is plausible and therefore in this section we discuss the validity of our procedure. We checked the validity of the credible intervals with a frequentist's approach, that is, to generate a sequence of credible intervals and see how many of them contain the posterior mean. Since the true value of posterior means has a probability of 95% that lies in our credible intervals, we want to check if it is true that 95% of the new sequence of intervals would contain the true value.

One way to simulate the data is to draw X and Y separately until their sum is n, when the odds ratio is fixed. This method however will not work for our data simply because this probability is too small to be carried out for our simulation. To illustrate, we tried to find

the maximum of the probability of $P(X+Y=n)$ when $p_2 = a$.

$$\begin{aligned}
\max_{p_2=a} P(X+Y=n) &= \sum_{x\in l} \binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x p_2^n (1-p_2)^{n_2-n}(1-p_1)^{n_1} \\
&= \sum_{x\in l} \binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x a^n (1-a)^{n_2-n}(1-p_1)^{n_1} \\
&= a^n(1-a)^{n_2-n}(1-p_1)^{n_1}\sum_{x\in l} \binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x, \\
&= a^n(1-a)^{n_2-n}(\frac{1-a}{1-a+a\Gamma})^{n_1}\sum_{x\in l} \binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x,
\end{aligned}$$

Since $\sum_{x\in l}\binom{n_1}{x}\binom{n_2}{n-x}\Gamma^x$ does not contain a, it is just a constant with respect to a, so we only need to maximize $a^n(1-a)^{n_2-n}(\frac{1-a}{1-a+a\Gamma})^{n_1}$.

Taking the natural log of this expression,

$$\ln[a^n(1-a)^{n_2-n}(\frac{1-a}{1-a+a\Gamma})^{n_1}] = n\ln a + (n_2-n)\ln(1-a) + n_1\ln(\frac{1-a}{1-a+a\Gamma})$$

Let $f(a) = n\ln a + (n_2-n)\ln(1-a) + n_1\ln(\frac{1-a}{1-a+a\Gamma})$,

$f'(a) = \frac{n}{a} + \frac{n-n_2}{1-a} + n_1\frac{-\Gamma}{(1-a)(1-a+a\Gamma)}$

Setting $f' = 0$, we get a quadratic function:

$[n - n\Gamma + n(n-n_2)(1-\Gamma)]a^2 + [(n-n_1)\Gamma - (n+n_2)]a + n = 0.$

If $\Gamma = 1, a = \frac{n}{n_1+n_2}$;

If $\Gamma \neq 1, a = \frac{(n+n_2)-(n-n_1)\Gamma+\sqrt{[(n-n_1)\Gamma-(n+n_2)]^2-4[n-n\Gamma+n(n-n_2)(1-\Gamma)]}}{2[n-n\Gamma+n(n-n_2)(1-\Gamma)]}$

We used the hospital data to calculate the maximum numerically but the probability is approximately 0 as listed in the Table 6. If this is the case, it will take a long time to get the desired n, so we decided to simulate the data using an iterative method discussed below. This paper also discussed other ways to run the simulation. The reason that we chose the following method is simply other methods are too complicated computationally. [9]

Table 6: Summary of the numerical maximum values

hospital	maximum
1	1.3428e-016
2	2.8095e-027
3	1.9376e-011
4	4.6761e-047
5	1.4129e-031
6	9.6806e-043
7	1.7940e-016
8	8.1939e-021
9	1.1589e-061
10	6.5116e-027
11	7.9693e-021
12	7.7578e-021
13	6.9383e-042
14	3.0186e-030
15	6.4903e-041
16	1.4681e-006
17	1.9217e-048
18	3.3069e-035
19	6.3999e-061

Before we generate a new sequence of θ, we need to simulate the contingency tables as well.The process of generating a new contingency table using the given θ is similar to the way we drew θ in previous section. For each hospital, the value of x has a specific range, which is $\{x : max(0, n - n_2) \leq x \leq min(n_1, n)\}$. For each x_i in the range, we calculated the cumulative probability of x_i, using the equation (1). We calculated the cumulative probability recursively, since the denominator of equation (1) is fixed for each hospital.

For the $i+1^{th}$ iteration, we have

$$f(x_{i+1}|\Gamma) = \frac{\binom{n_1}{x_{i+1}}\binom{n_2}{n-x_{i+1}}\Gamma^{x_{i+1}}}{\sum_{x\in\mathcal{L}}\binom{n_1}{x_{i+1}}\binom{n_2}{n-x_{i+1}}\Gamma^{x_{i+1}}}$$

Taking the ratio of $f(x_{i+1})$ and $f(x_i)$ given in (1), we have

$$\frac{f(x_{i+1}|\Gamma)}{f(x_i|\Gamma)} = \frac{n_1-i}{i+1}\frac{n-i}{n_2-n+i+1}\Gamma$$

where $\Gamma = \frac{\theta}{1-\theta}$.

After calculating the cumulative probability for each x_i, we drew a thousand random numbers $u_i, i = 1, 2, ..., 1000$ from $U \sim (0, 1)$. Using the inverse CDF sampling method, we generated 1000 new contingency tables for each hospital. We used the data from each of the new tables we generated to draw another 10,000 θ values and calculated the 95% credible interval for each hospital. We reproduced each interval 1000 times. We did this is to check if our sampling method was agreeable with frequentist methods since we simulated our data in a Bayesian setting. The percentage of the new credible intervals containing the original θ we generated are listed in Table 7.

With a nominal coverage probability of 0.95, 95% of the 1000 intervals generated should contain the original value of θ. For 17 out of the 19 hospitals, as well as the combined hospital data, the percentages of the new credible intervals containing the original θ was over 95%. This means that the intervals for those hospitals, all but hospitals 2 and 14, were too conservative. The percentage of the new credible intervals containing the original θ for hospital 2 was unreasonable at 78.3%. Only hospital 14 had a percentage of 94.7 which is approximately 95%. This indicates that there is a need for a different approach to estimate

θ, thus leading to our pooled estimation.

Table 7: Percentage of the new intervals containing the original θ.

Hospital	Percentage
1	96.0%
2	78.3%
3	97.1%
4	100%
5	96.2%
6	98.3%
7	96.0%
8	97.5%
9	99.2%
10	98.7%
11	96.7%
12	98.0%
13	97.1%
14	94.7%
15	95.6%
16	99.4%
17	98.9%
18	95.9%
19	95.9%
1-19	96.2%

3 Pooled Hospital Data Estimation

Since most of the credible intervals are too conservative, we modified the model and changed the prior distribution in order to borrow strength from pooling the hospital data together. The new hierarchical model is of the following form:

$$f(x_i|\theta_i) = \frac{\binom{n_{1i}}{x_i}\binom{n_{2i}}{n_i - x_i}\left(\frac{\theta_i}{1-\theta_i}\right)^{x_i}}{\sum \binom{n_{1i}}{x_i}\binom{n_{2i}}{n_i - x_i}\left(\frac{\theta_i}{1-\theta_i}\right)^{x_i}} \qquad 0 < \theta_i < 1 \qquad i = 1, \ldots, 19 \qquad (4)$$

Where $x_i \in$: $\{\mathcal{L}_i : max(0, n_i - n_{2i}) \leq x \leq min(n_{1i}, n_i)\}$.

The prior for θ_i is given by

$$f(\theta_i|\mu, \tau) = Beta(\mu\tau, (1-\mu)\tau) \qquad 0 < \mu < 1$$

For (μ, τ), we assume the following density:

$$f(\mu, \tau) = \frac{1}{(1+\tau)^2} \qquad \tau > 0$$

Thus, the joint density of $\theta_i, \ldots, \theta_{19}, \mu, \tau$ given $x_1, \ldots, x_{19}$ is:

$$f(\theta_1, \ldots, \theta_{19}, \mu, \tau|x_1, \ldots, x_{19}) \propto f(\mu, \tau) \prod_{i=1}^{19} \{f(x_i|\theta_i) f(\theta_i|\mu, \tau)\}$$

$$= \left(\frac{1}{(1+\tau)^2}\right) \prod_{i=1}^{19} \left\{ \frac{\binom{n_{1i}}{x_i}\binom{n_{2i}}{n_i - x_i}\left(\frac{\theta_i}{1-\theta_i}\right)^{x_i}}{\sum \binom{n_{1i}}{x_i}\binom{n_{2i}}{n_i - x_i}\left(\frac{\theta_i}{1-\theta_i}\right)^{x_i}} \frac{\theta_i^{\mu\tau-1}(1-\theta_i)^{(1-\mu)\tau-1}}{B\left(\mu\tau,(1-\mu)\tau\right)} \right\}$$

The conditional marginal distribution of each parameter would be:

$$f(\theta_i|\mu,\tau,\underline{x_i}) \propto \frac{\binom{n_{1i}}{x_i}\binom{n_{2i}}{n-x_i}\left(\frac{\theta_i}{1-\theta_i}\right)^{x_i}}{\sum \binom{n_1 i}{x_i}\binom{n_2 i}{n-x_1 i}\left(\frac{\theta_i}{1-\theta_i}\right)^{x_i}} \theta_i^{\mu\tau-1}(1-\theta_i)^{(1-\mu)\tau-1}$$

θ_i's given (μ, τ, x_i) are assumed to be independent for $i = 1, \ldots, 19$.

$$f(\mu|\tau,\theta_1,\ldots,\theta_{19},\underline{x_i}) \propto \prod_{i=1}^{19} \frac{\theta_i^{\mu\tau-1}(1-\theta_i)^{(1-\mu)\tau-1}}{B\left(\mu\tau,(1-\mu)\tau\right)}$$

$$f(\tau|\mu,\theta_1,\ldots,\theta_{19},\underline{x_i}) \propto \left(\frac{1}{(1+\tau)^2}\right) \prod_{i=1}^{19} \frac{\theta_i^{\mu\tau-1}(1-\theta_i)^{(1-\mu)\tau-1}}{B\left(\mu\tau,(1-\mu)\tau\right)} \qquad 0 < \mu < 1$$

Since $\tau > 0$, we transformed τ to [0,1] for computational convenience and stability.

Let $\tau = \frac{\phi}{1-\phi}$, then

$$f(\phi|\mu, \theta_1, \ldots, \theta_{19}, \underline{x_i}) \propto \prod_{i=1}^{19} \left[\frac{\theta_i^{\mu(\frac{\phi}{1-\phi})-1} (1-\theta_i)^{(1-\mu)(\frac{\phi}{1-\phi})-1}}{B\left(\mu(\frac{\phi}{1-\phi}), (1-\mu)(\frac{\phi}{1-\phi})\right)} \right]$$

3.1 Approximation of the Sampling Distribution

We started from hospital 1 and used the 10000 θ we previously sampled using the grid method to fit a Beta distribution for the sampling distribution (1). Beta distribution is an ideal approximation of our sampling distribution because Beta distribution could take any shape in the interval [0,1]

$\theta_1, ..., \theta_{10000} \sim Beta(\mu_0\tau_0, (1-\mu_0)\tau_0)$

The joint density distribution of $\theta_i, i = 1 \ldots 10000$ is

$$\prod_{i=1}^{10000} \frac{\theta_i^{\mu_0\tau_0-1}(1-\theta_i)^{(1-\mu_0)\tau_0-1}}{B\left(\mu_0\tau_0, (1-\mu_0)\tau_0\right)}$$

This can be observed by doing the steps of maximum likelihood estimation up to taking the derivative of the likelihood function.

$$
\begin{aligned}
&= \prod_{i=1}^{M} \frac{\theta_i^{\mu_0\tau_0-1}(1-\theta_i)^{(1-\mu_0)\tau_0-1}}{B\left(\mu_0\tau_0,(1-\mu_0)\tau_0\right)}\,,\ M = 10000 \\
&= \left[\frac{(\prod_{i=1}^{M}\theta_i)^{\frac{\mu_0\tau_0-1}{M}}(\prod_{i=1}^{M}1-\theta_i)^{\frac{(1-\mu_0)\tau_0-1}{M}}}{B(\mu_0\tau_0,(1-\mu_0)\tau_0)}\right]^M \\
&= \left[\frac{G_1^{\mu_0\tau_0-1}G_2^{(1-\mu_0)\tau_0-1}}{B(\mu_0\tau_0,(1-\mu_0)\tau_0)}\right]^M\,,
\end{aligned}
$$

Where G_1 and G_2 are the geometric means of θ_i and $(1-\theta_i)$ respectively. The next step is to take the log of both sides

$$
\begin{aligned}
&\log\prod_{i=1}^{10000}\frac{\theta_i^{\mu_0\tau_0-1}(1-\theta_i)^{(1-\mu_0)\tau_0-1}}{B\left(\mu_0\tau_0,(1-\mu_0)\tau_0\right)} \\
&= M[(\mu_0\tau_0-1)\log G_1 + ((1-\mu_0)\tau_0-1)\log G_2 - \log\tfrac{\Gamma(\mu_0\tau_0+(1-\mu_0)\tau)}{\Gamma(\mu_0\tau_0)\Gamma((1-\mu_0)\tau_0)}]
\end{aligned}
$$

It is clear that differentiating this cannot be done analytically. Therefore, we implemented a numerical optimization method, Nelder-Mead, which is the algorithm behind a built-in MATLAB function, fminsearch. It is a numerical method used to minimize functions, so we had to take the negative of the function of interest and minimize it with respect to μ_0, τ_0. To pass the constraints $0 < \mu < 1$ and $\tau > 0$, we did the following substitution:

$x_{(1)} = \log\mu - \log(1-\mu)$

$x_{(2)} = \log\tau$

In order for the Nelder-Mead method to converge faster, we calculated two estimators the parameter of interest μ, τ based on the methods of moments to be our initial guess for the optimization. The chosen inital guess are $\mu_0 = \overline{X}$ and $\tau_0 = \frac{\overline{X}(1-\overline{X})}{s^2}$, where $\overline{X}$ and S^2 are

the sample mean and sample variance of $\theta_i, i = 1, ..., 10000$ respectively. For instance, the sample mean and sample variance would be $\overline{X} = 0.5758, S^2 = 0.0186$ for the θ we generated previously from the given hospital 1 data. Thus $\mu_0 = 0.5758$ and $\tau_0 = 13.1636$. Table 8 is the full table of the optimized values of initial guess for μ and τ.

Table 8: Optimized initial values obtained from Nelder-Mead algorithm

hospital	μ,τ
1	0.576, 13.164
2	0.474, 14.421
3	0.744, 10.242
4	0.530, 13.189
5	0.732, 19.332
6	0.618, 14.386
7	0.755, 11.556
8	0.824, 10.522
9	0.651, 10.646
10	0.676, 13.527
11	0.479, 8.888
12	0.738, 14.660
13	0.672, 24.777
14	0.500, 11.233
15	0.426, 14.139
16	0.677, 7.922
17	0.756, 9.620
18	0.865, 35.340
19	0.736, 35.029

3.2 Metropolis-Hastings (M-H) Algorithm

Because the expression of the conditional marginal density is too complicated for θ_i, it is not computationally efficient to generate random numbers from the density using grid method. That is why we chose Metropolis-Hastings Algorithm.

The steps of the algorithm are generally carried out as follows.

Let $V \sim f_V(v)$ and $Y \sim f_Y(y)$ where the density of V is the candidate (or proposal) density and the density of Y is the target density and f_Y and f_V have common support. To generate $Y \sim f_Y$:

0. Generate $V \sim f_V$. Set $Z_0 = V$. For $i = 1, 2, \ldots$:

1. Generate $U_i \sim uniform(0, 1)$, $V_i \sim f_V$ and calculate

$$\rho_i = \min\{\frac{f_Y(V_i)}{f_V(V_i)} \cdot \frac{f_V(Z_{i-1})}{f_Y(Z_{i-1})}, 1\}$$

2. Set

$$Z_i = \begin{cases} V_i & \text{if } U_i \leq \rho_i \\ Z_{i-1} & \text{if } U_i > \rho_i \end{cases}$$

Then as $i \to \infty, Z_i$ converges to Y in distribution [3].

The target density is the 19 conditional density functions of θ_i given the rest of the 18 θ_j for $i \neq j, i, j = 1 \ldots 19$ and μ, τ. The proposal density is the approximated posterior distribution. Take hospital 1 for example:

$$\begin{aligned} f(\theta_1|\mu, \tau, \underline{x_i}) &= Beta(\mu_0\tau_0, (1-\mu_0)\tau_0)\theta_1^{\mu\tau-1}(1-\theta_1)^{(1-\mu)\tau-1} \\ &= \frac{\theta_i^{\mu_0\tau_0-1}(1-\theta_i)^{(1-\mu_0)\tau_0-1}}{B\left(\mu_0\tau_0, (1-\mu_0)\tau_0\right)}\theta_1^{\mu\tau-1}(1-\theta_1)^{(1-\mu)\tau-1} \\ &= Beta(\mu_0\tau_0 + \mu\tau - 1, (1-\mu_0)\tau_0 + (1-\mu)\tau - 1) \end{aligned}$$

Notice that our target density and proposal density have common terms: $\theta_1^{\mu\tau-1}(1-\theta_1)^{(1-\mu)\tau-1}$, so when calculating $\frac{f_Y(V_i)}{f_V(V_i)}$ and $\frac{f_V(Z_{i-1})}{f_Y(Z_{i-1})}$, this term would cancel out.

3.3 Metropolis-Hastings (M-H) Sampler

The griddy Metropolis-Hastings sampler iteratively uses the full set of univariate conditionals which eventually converge to the true posterior distribution. As mentioned previously, we used the M-H Algorithm to draw random samples of parameter $\theta_i, i = 1 \ldots 19$ during each iteration. The grid method was also used within this procedure to more effectively draw samples of μ and τ. The following algorithm is our implementation of griddy M-H Sampler on the set of univariate conditionals that we used:

$$f(\theta_i|\mu,\tau,x_i) = \frac{\binom{n_{1i}}{x_i}\binom{n_{2i}}{n_i - x_{1i}}\left(\frac{\theta_i}{1-\theta_i}\right)^{x_i}}{\sum \binom{n_{1i}}{x_{1i}}\binom{n_{2i}}{n_i - x_i}\left(\frac{\theta_i}{1-\theta_i}\right)^{x_i}}\,\theta_1^{\mu\tau-1}(1-\theta_i)^{(1-\mu)\tau-1}$$

$$f(\mu|\tau,\theta_1,...,\theta_{19},\underline{x_i}) \propto \prod_{i=1}^{19}\frac{\theta_i^{\mu\tau-1}(1-\theta_i)^{(1-\mu)\tau-1}}{B\left(\mu\tau,(1-\mu)\tau\right)} = \left[\frac{G_1^{\mu\tau-1}G_2^{(1-\mu)\tau-1}}{B(\mu\tau,(1-\mu)\tau)}\right]^{19}$$

$$f(\tau|\mu,\theta_1,...,\theta_{19},\underline{x_i}) \propto \prod_{i=1}^{19} \frac{\theta_i^{\mu\tau-1}(1-\theta_i)^{(1-\mu)\tau-1}}{B\left(\mu\tau,(1-\mu)\tau\right)} \left(\frac{1}{(1+\tau)^2}\right)$$
$$= [\frac{G_1^{\mu\tau-1}G_2^{(1-\mu)\tau-1}}{B(\mu\tau,(1-\mu)\tau)}]^{19}$$
$$0<\mu<1, \tau=\frac{\phi}{1-\phi}, 0<\phi<1 \quad ,$$

where G_1 and G_2 are the geometric means of θ_i and $(1-\theta_i)$ respectively.

3.4 Pooled Data Estimation Results

The previously described methods of approximation and simulation used for the pooled hospital data lead to different estimated values for the parameters θ_i. Table 9 shows our pooled data estimation results. For each hospital, the mean of θ was greater than or equal to 0.5 with only the mean of θ for hospital 2 equal to 0.500. The 95% credible intervals for hospitals 18, 19, and the combined data of all 19 hospitals were significant. These three intervals did not contain 0.5 since the lower bound of the interval was greater than 0.5; for example 18's interval of $[0.655, 0.876]$. The combined data had a 95% credible interval of $[0.602, 0.726]$. This means that our estimate for θ for the combined data is significantly greater than 0.5. Recall a value for θ of 0.5 translates to a value of the odds ratio, Γ, of 1. Thus the interval reveals that Γ is significantly greater than 1, once again indicating that overall the odds of a black patient receiving a liver scan are greater than the odds for a white patient.

Table 9: Pooled data estimation results

Hospital	Mean of θ	Standard deviation of θ	95% Credible Interval
1	0.557	0.099	[0.360, 0.741]
2	0.500	0.099	[0.305, 0.688]
3	0.633	0.098	[0.422, 0.808]
4	0.536	0.098	[0.345, 0.723]
5	0.657	0.083	[0.482, 0.805]
6	0.582	0.095	[0.391, 0.755]
7	0.566	0.100	[0.366, 0.756]
8	0.654	0.102	[0.443, 0.835]
9	0.587	0.101	[0.380, 0.773]
10	0.610	0.097	[0.410, 0.789]
11	0.520	0.110	[0.303, 0.722]
12	0.643	0.097	[0.439, 0.814]
13	0.631	0.077	[0.478, 0.773]
14	0.522	0.101	[0.324, 0.717]
15	0.519	0.102	[0.317, 0.714]
16	0.647	0.090	[0.460, 0.809]
17	0.641	0.091	[0.454, 0.810]
18	0.776	0.057	[0.655, 0.876]
19	0.686	0.065	[0.553, 0.805]
1-19	0.662	0.031	[0.602, 0.726]

When implementing Metropolis-Hastings Algorithm, researchers usually are concerned with the following terms: the acceptance rate and the autocorrelation of the simulated data. Ideally, the acceptance rate should be between 0.5 and 0.75. Our result, as shown in Table 10, is not the most optimal but is acceptable. One can improve on the Metropolis-Hastings sampler by adjusting the jumping probabilities via a mixture of beta distributions as a proposal density.

Table 10: Acceptance rate of pooled data from Metropolis-Hastings algorithm.

hospital	jumping probability
1	0.836
2	0.789
3	0.835
4	0.841
5	0.361
6	0.863
7	0.575
8	0.660
9	0.835
10	0.796
11	0.766
12	0.437
13	0.449
14	0.820
15	0.803
16	0.810
17	0.837
18	0.058
19	0.457

As shown in Figure 4, all the θ are bouncing between an interval, which is what we expected since θ is bounded in $[0, 1]$.

Figure 4: Trace plots of θ in the pooled method

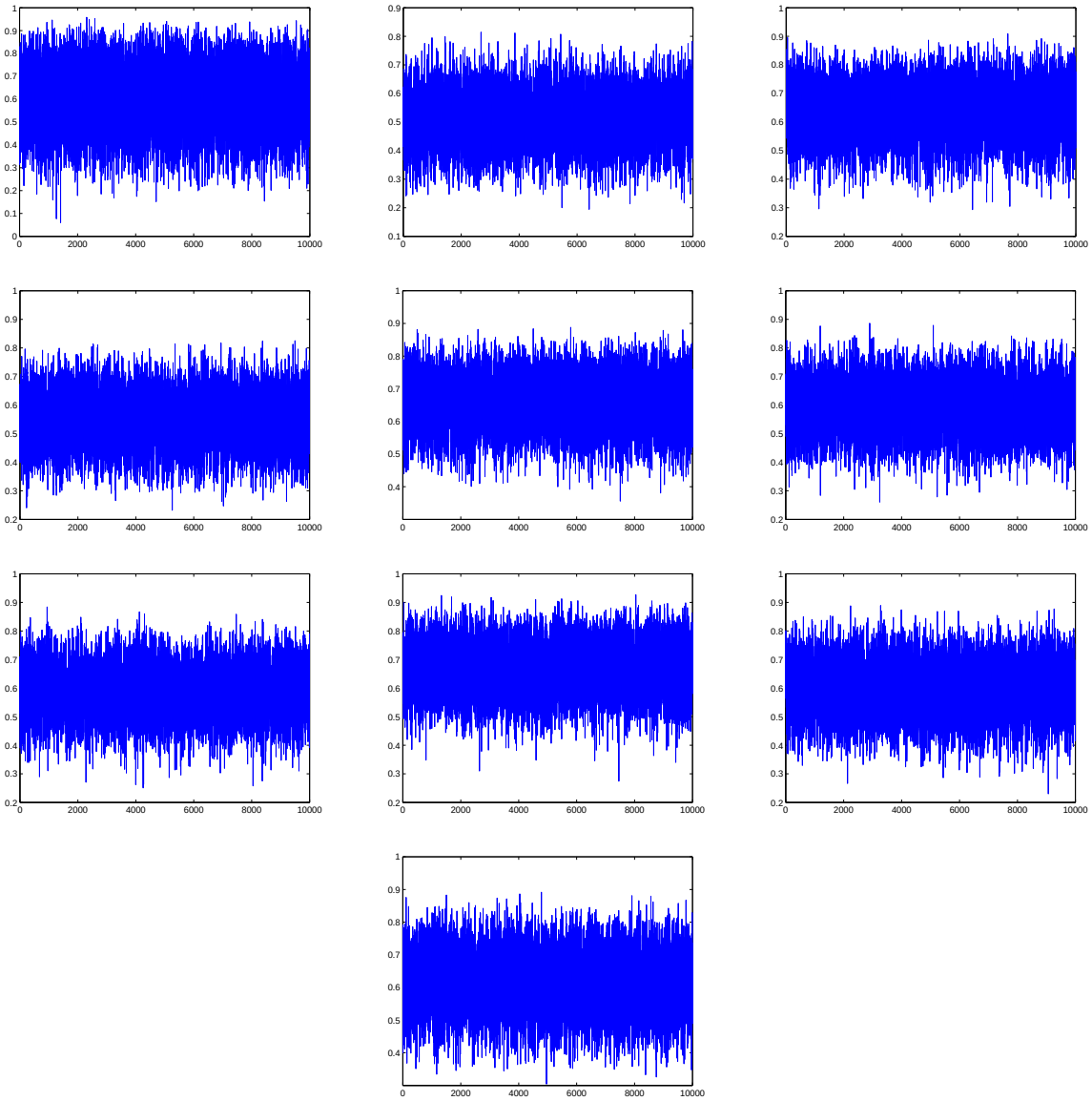
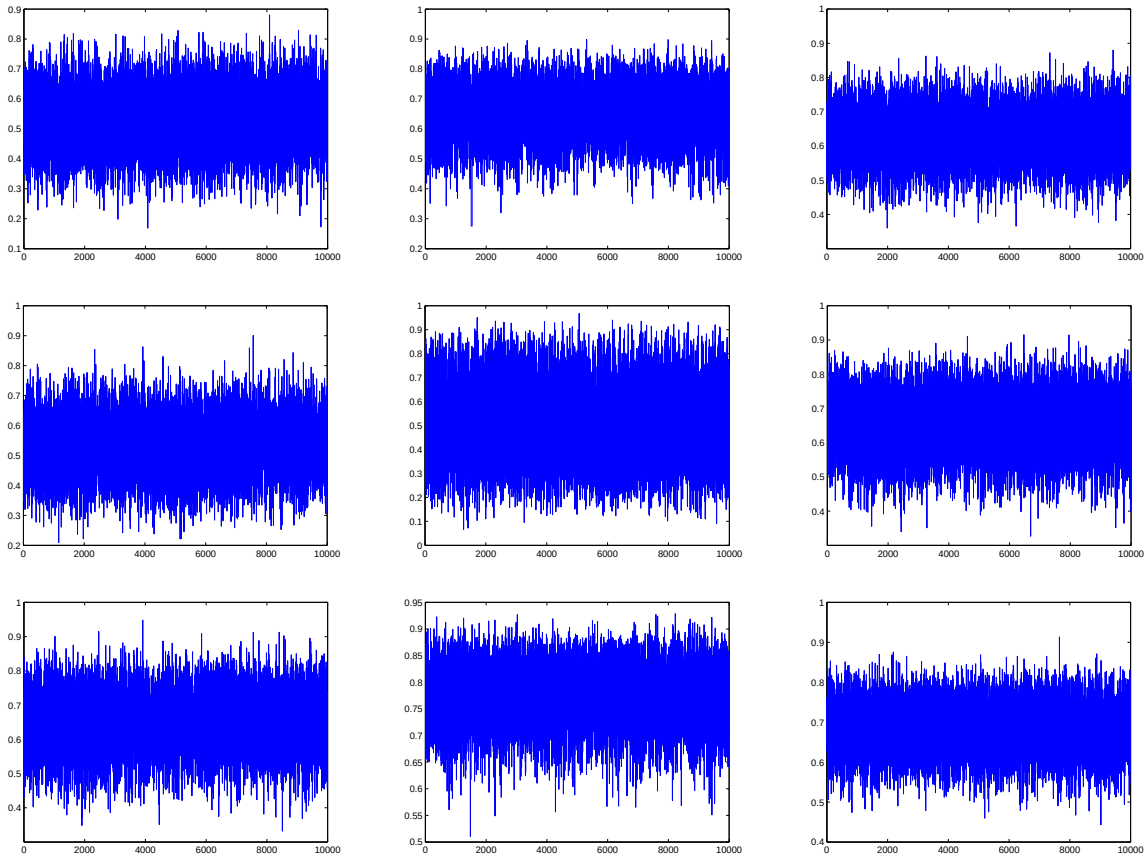
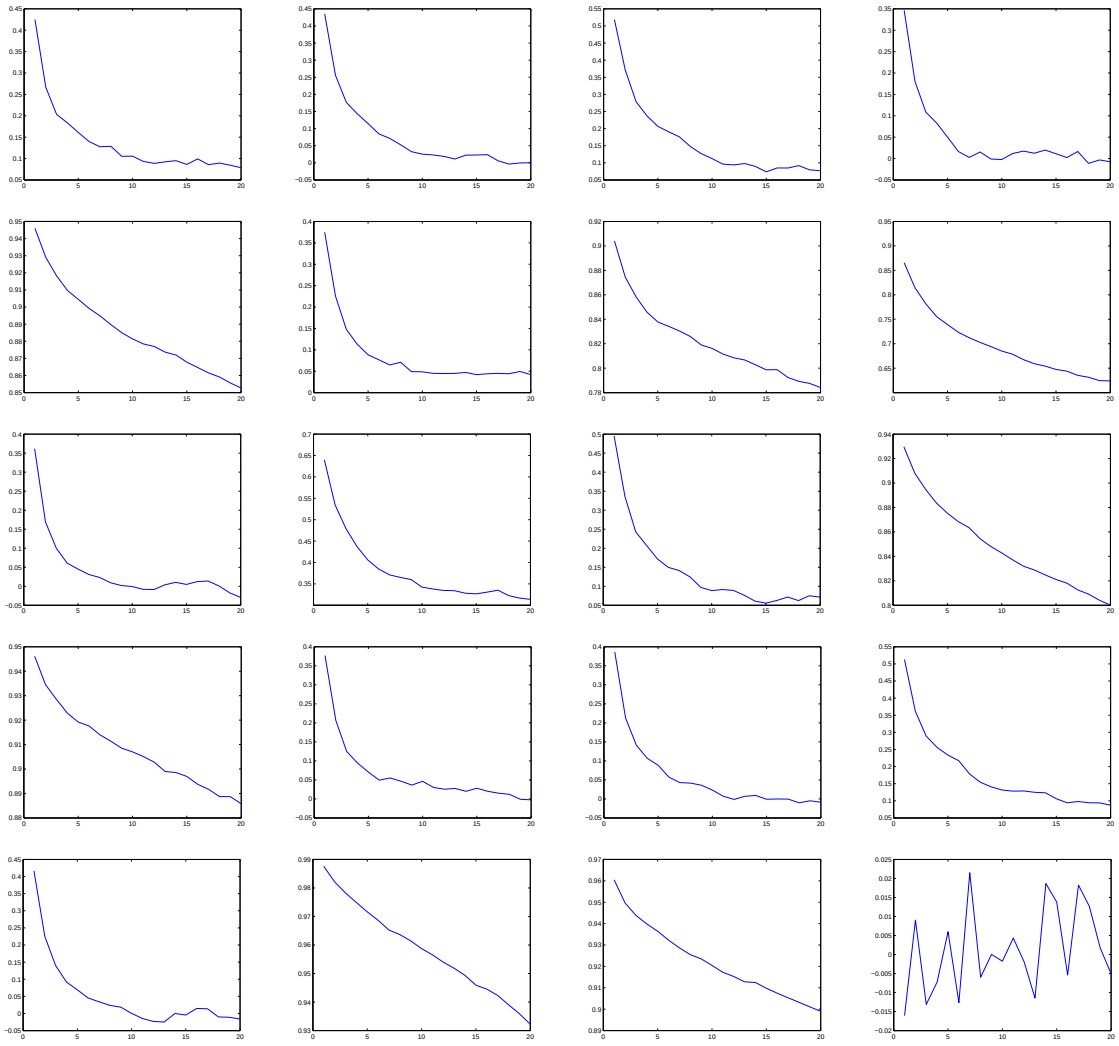


Figure 5: Autocorrelation of θ in the pooled method



We used 20 lags of the data to check the autocorrelation and the result of the pooled data is what we expected. After several lags, the autocorrelation of the data starts to wash off.

Figure 6: Autocorrelation of θ in the pooled method continued



4 Hospital Specific Data vs. Pooled Data Results

For the comparison of the two methods, we transformed the θ values back to Γ values. Significance is determined from analysis of the 95% credible intervals for Γ. First we examined the change in the intervals from the first method to the second.

The hospital specific data estimation method led to wide credible intervals for Γ for several hospitals; for example hospital 8, $[1.1053, 39.000]$, or hospital 17, $[0.7699, 27.5714]$. The pooled method reduced the intervals for hospital's 8 and 17 to $[0.8087, 5.3778]$ and $[0.7462, 4.4577]$ respectively. The pooled method worked similarly for all of the hospitals, creating more reliable intervals for each. The pooled method's intervals are smaller. Table 11 clearly displays the comparison of the intervals from the two methods.

Table 11: Specific vs. pooled data estimation comparison

	Mean of Γ		Std. dev. of Γ		95% Credible Interval	
Hosp.	Specific	Pooled	Specific	Pooled	Specific	Pooled
1	1.647	1.404	1.041	0.610	[0.418, 4.405]	[0.578, 2.876]
2	1.058	1.085	0.673	0.475	[0.290, 2.774]	[0.416, 2.246]
3	4.752	1.947	5.730	1.013	[0.770, 17.182]	[0.702, 4.450]
4	1.391	1.276	0.911	0.571	[0.399, 3.878]	[0.506, 2.71]
5	3.384	2.104	2.044	0.901	[1.062, 8.524]	[0.892, 4.294]
6	2.047	1.551	1.280	0.666	[0.626, 5.452]	[0.645, 3.187]
7	4.466	2.064	3.579	1.038	[0.835, 14.385]	[0.719, 4.660]
8	10.191	2.268	17.845	1.211	[1.105, 39.000]	[0.809, 5.378]
9	2.726	1.592	3.068	0.785	[0.527, 9.526]	[0.616, 3.561]
10	2.707	1.761	1.879	0.800	[0.653, 7.696]	[0.684, 3.755]
11	1.180	1.213	0.958	0.581	[0.183, 3.651]	[0.452, 2.657]
12	3.761	2.087	2.473	1.000	[0.905, 10.765]	[0.766. 4.559]
13	2.337	1.853	1.050	0.687	[0.905, 5.061]	[0.844, 3.489]
14	1.235	1.206	0.916	0.546	[0.2739, 3.444]	[0.474, 2.559]
15	0.853	0.998	0.558	0.438	[0.242, 2.279]	[0.409, 2.085]
16	3.416	1.686	3.723	0.911	[0.439, 12.333]	[0.587, 3.969]
17	6.002	1.958	10.993	1.011	[0.770, 27.571]	[0.746, 4.458]
18	7.834	3.871	4.318	1.638	[2.774, 17.182]	[1.704, 7.825]
19	3.114	2.357	1.270	0.793	[1.353, 6.407]	[1.183, 4.285]
1-19	1.717	2.301	0.238	0.195	[1.299, 2.279]	[1.251, 2.206]

Table 12 displays the percent difference of the standard deviations between the two methods. The pooled data estimation did have smaller standard deviations of the means and thus less variability than the hospital specific data estimation. Reduced variability indicates an improvement of reliability between the two methods. This supports the notion that pooled data estimation is more precise than individual hospital data estimation.

Table 12: Difference of the standard deviations between the two methods

hospital	difference
1	41.4 %
2	29.4%
3	82.3%
4	37.3%
5	55.9%
6	48.0%
7	71.0%
8	93.2%
9	74.4%
10	57.4%
11	39.4%
12	59.6%
13	34.6%
14	40.4%
15	21.5%
16	75.5%
17	90.8%
18	62.1%
19	37.6%
1-19	18.1%

For both methods, the means of Γ for all the hospitals except hospital 15 are greater than 1, thus leading one to intuitively expect that some of the Γ are significantly greater than 1. Based on the intervals from the hospital specific data estimation method, hospitals 5, 8, 18, 19, and the combined data had odds ratios Γ significantly greater than 1. The pooled data estimation method only had three intervals indicating that the odds ratios significantly greater than 1: the intervals from hospitals 18, 19, and the combined data. Since the pooled data method had less variability and thus produced more reliable intervals, one should only assume that individually, only hospitals 18 and 19 were significantly more likely to give a liver scan to black patients than white patients. More importantly however, the combined data in both methods led to Γ significantly greater than 1, supporting the claim that overall

the black patients had higher odds of receiving a liver scan.

5 Discussion

Small area estimation is used to formulate inferences about a larger population based data obtained from small areas such as hospitals. For policy makers to make decisions based on statistical experiments that involve small area estimation, the methods of estimation must be reliable. For example, to draw inferences about the differences in patterns of care between black and white patients in hospitals overall using small area estimates, the estimates must be reliable. Though our data came from a previously conducted and reliable study our main contribution was to illustrate the precision and accuracy of adjusted Bayesian methods in small area estimation.

Both the procedures carried out in Chapter 2 and the procedure carried out in Chapter 3 supported the rejection of the null hypothesis, the common odds ratio $\Gamma_i = 1$, in favor of the alternative hypothesis, the common odds ratio $\Gamma_i \neq 1$. Thus ultimately our results support the claim from the Diehr et al. study that the black breast-cancer patients observed were more likely to receive a liver scan than the white patients, though the liver scan was deemed less appropriate [6]. The main aspect we were concerned with was developing a valid procedure to estimate the parameters Γ_i.

Hospital specific data was used for the first method. MATLAB programs were implemented to draw a sample of 10,000 odds ratios Γ, transformed to $\theta = \frac{\Gamma}{1+\Gamma}$ for computational convenience, from the posterior distribution using the grid method. 4/19 hospitals, as well as the data combined hospital data, had odds ratios significantly greater than 1, indicating higher odds of receiving a liver scan for the black patients. The validity of the process used to generate the 10,000 θ was then assessed and proved to be acceptable, but not optimal since the results were conservative.

Pooled data was used for the second method. Both the Metropolis-Hastings (M-H) algorithm and the grid method were used simultaneously during this procedure with MATLAB. The grid method was used only to more effectively draw samples of μ and τ at each iteration within the M-H algorithm, which was used to draw samples of θ. The pooled data estimation method reduced the variability of the results, producing more reliable intervals for the estimates of the odds ratios.

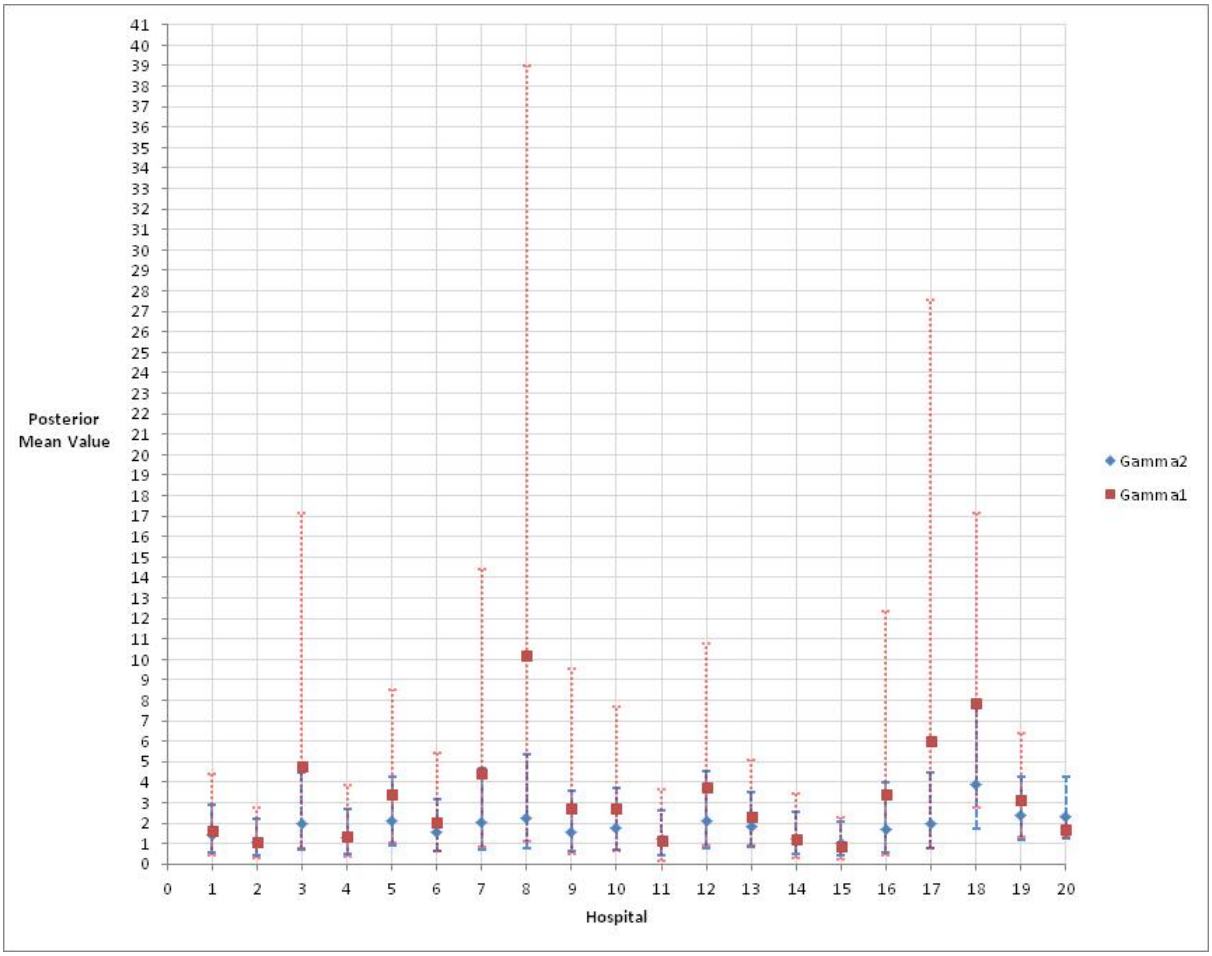
The main difference between the two methods to draw values of θ was the pooling of the hospital data in the second method. The idea of pooling the data comes from the assumption that the data within each hospital are similar. The data was pooled to reduce variance, thus improving precision and the ability to draw inferences about the information contained in the data. As previously stated, the pooled data estimation did prove to have less variability than the individual data.

Though pooling data is done to improve precision, some potential concerns arise when doing so. The pooling of the data pulls all the values closer to the mean, and values with greater variability get pulled more towards the mean. This is one more issue to account for when carrying out precise and accurate small area estimation.

There are always potential concerns when estimating parameters of a distribution using data from small sample sizes. Accurate estimates of parameters of probability distributions are vital components to drawing useful and rational inferences in a statistical study. Precision in the methods used to obtain parameter estimates is equally as important, especially when dealing with statistical studies aimed to evaluate patient care in hospitals. The focus of our project was to evaluate methods of estimating parameters Γ_i of a noncentral hypergeometric distribution using small areas, specifically 19 hospitals. In terms of precision, estimation with pooled data proved to be the more reliable approach. It is useful to use adjusted

Bayesian methods when conducting small area estimation. This Bayesian study was aimed to contribute to what can be known about small area estimation of parameters of a given probability distribution.

Figure 7: Comparison of Γ obtained from two methods



A Hospital Data

Table 13: Hospital data

Hospital 1			Hospital 9			Hospital 17		
4	9	13	7	2	9	6	1	7
12	34	46	77	38	115	45	31	76
16	43	59	84	40	124	51	32	83
Hospital 2			Hospital 10			Hospital 18		
4	6	10	4	6	10	14	10	24
34	33	67	20	70	90	12	70	82
38	39	77	24	76	100	26	80	106
Hospital 3			Hospital 11			Hospital 19		
7	2	9	1	8	9	15	15	30
6	7	13	16	76	92	43	129	172
13	9	22	17	84	101	28	144	202
Hospital 4			Hospital 12					
5	5	10	4	10	14			
59	56	115	10	91	101			
64	61	125	14	101	115			
Hospital 5			Hospital 13					
7	7	14	9	18	27			
22	69	91	27	118	145			
29	76	105	36	136	172			
Hospital 6			Hospital 14					
5	6	11	3	5	8			
41	80	121	35	45	80			
46	86	132	38	50	88			
Hospital 7			Hospital 15					
3	6	9	9	5	14			
8	72	80	69	20	89			
11	78	89	78	25	103			
Hospital 8			Hospital 16			Hospital 1-19		
4	8	12	2	3	5	113	132	245
1	20	21	3	12	15	540	1071	1611
5	28	33	5	15	20	653	1203	1856

References

[1] Brigham and Women's Hospital (2010). *What is a liver scan?* StayWell Custom Communications.

[2] Carr, Wendell E. (1980). *Fisher's Exact Test Extended to More Than Two Samples of Equal Size*, Technometrics **22**(2) , pp. 269-270 Published by: American Statistical Association and American Society for Quality Article Stable URL: http://www.jstor.org/stable/1268467

[3] Casella, George & Berger, Roger L. (2002). *Statistical Inference 2nd Edition.* pp. 253-255, 511. Pacific Grove, CA: Duxbury.

[4] Casella, George & George, Edward (1992). *Explaining the Gibbs Sampler*, The American Statistician, **46**, (3), pp. 167-174 Published by: American Statistical Association

Stable URL: http://www.jstor.org/stable/2685208

[5] Chib, Siddhartha & Greenberg, Edward (1995). *Understanding the Metropolis-Hastings Algorithm*, The American Statistician, **49**, (4), pp. 327-335. Published by: American Statistical Association Article Stable URL: http://www.jstor.org/stable/2684568

[6] Diehr, Yergan, Chu, Feigl, Glaefke, Moe, Bergner, & Rodenbaugh (1989). *Treatment Modality and Quality differences for Black and White Breast-Cancer Patients Treated in Community Hospitals*, Medical Care, **27**,(10) pp. 942-958.

[7] Ghosh, M. & Rao, J. N. K. (1994). *Small Area Estimation: An Appraisal*, Statistical Science **9**,(1), pp. 55-76.

[8] Hollander, M. and Wolfe, D.A. (1999). *Nonparametric Statistical Methods 2nd Edition*. Wiley-Interscience pp. 484-488.

[9] Liao, J. G & Rosen, Ori. (2001). *Fast and Stable Algorithms for Computing and Sampling from the Noncentral Hypergeometric Distribution*, The American Statistician, **55**,(4), 366-369.

[10] Rao, J. N. K. (2003). *Small Area Estimation: Methods and Applications* John Wiley.

[11] Ritter, C. & Tanner, M.A. (1992). *Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler*, **87**,(419), pp. 861-868. Published by: American Statistical Association Stable URL: http://www.jstor.org/stable/2290225

[12] Wittes, J. & Wallenstein, S. (1987). *The Power of the Mantel-Haenszel Test*, Journal of the American Statistical Association, **82**,(400), 1104-1109.