ABSTRACT

Normal Approximation for Bayesian Models with Non-sampling Bias

Jiang Yuan, Ph.D.

Chairpersons: James D. Stamey, Ph.D.

Bayesian sample size determination can be computationally intensive for models where Markov chain Monte Carlo (MCMC) methods are commonly used for inference. It is also common in a large database where the unmeasured confounding presents. We present a normal theory approximation as an alternative to the time consuming MCMC simulations in sample size determination for a binary regression with unmeasured confounding. Cheng et al. (2009) develop a Bayesian approach to average power calculations in binary regression models. They then apply the model to the common medical scenario where a patient's disease status is not known. In this dissertation, we generate simulations based on their Bayesian model with both binary and normal outcomes. We also use normal theory approximation to speed up such sample size determination and compare power and computational time for both.

Unmeasured confounding arises when factors unrelated to the particular study have a hidden effect on observed health outcomes. The potential causal effect estimates would be biased without proper adjustment. In this dissertation, we combine a small sample of validation data to our sample to help address this problem. Simulation studies indicate that both of our methods: Bayesian MCMC method and the normal approximation method have provided good estimates of regression coefficients, variability, and power. The comparison also suggests that the Bayesian model may take advantage of prior information when available, and that it performs similarly to the normal approximation when relatively non-informative priors are used with large sample sizes.

We further explore the performance of the methods by changing the distribution of a continuous response to a skewed distribution. In this dissertation, we consider the gamma distribution due to its popularity and application in health care costs, where different health insurance, treatment modalities or patient characteristics change the cost. (see Manning et al., 2002, 2005; Morteza Khodabina, 2010)

The analysis of cost-effectiveness arises frequently in practice nowadays, and health care policy makers expect evidence supporting the cost-effectiveness of new health care interventions in pharmaceuticals. Hence we apply Bayesian MCMC method in the cost-effectiveness analysis and try to determine whether a normal approximation is feasible to be an time-saving alternative. This becomes the last topic of this dissertation.

Normal Approximation for Bayesian Models with Non-sampling Bias

by

Jiang Yuan, B.S., M.S.

A Dissertation

Approved by the Department of Statistical Science

_____

Jack D. Tubbs, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

_____

James D. Stamey, Ph.D., Chairperson

_____

Dennis A. Johnston, Ph.D.

_____

David J. Ryden, Ph.D.

_____

Jack D. Tubbs, Ph.D.

_____

Dean M. Young, Ph.D.

Accepted by the Graduate School
December 2013

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It is almost six years since I flew half way around the world and landed in Waco, Texas to study at Baylor University. Since then, I have enjoyed every minute of it, and I am going to get my PhD degree. I finally made it.

To study for a graduate degree has opened up a brand new page in my life, thanks to Dr. Bratcher and the statistics faculty, who believed in my potential and offered me this opportunity. I was very fortunate and I have been working very hard to get to where I am now.

First, I want to express my sincere gratitude to all my committee members who unreservedly shared their time and expertise with me. I thank all the faculty members in the Department of Statistical Science for their instructions. I especially would like to thank Dr. James Stamey and Dr. John Seaman Jr, for their patience, motivation, enthusiasm, and immense knowledge. Their supervision and encouragement helped me in all the time of research and the writing of this dissertation. Without their guidance and persistent help, this dissertation would not have been possible. I also want to give my thanks to Dr. Young who always encouraged and comforted me when I was down, to Dr. Tubbs and Dr. Harville, for always being available for my questions and being generous with their time and knowledge. I am heartily thankful to Dr. Johnston, not only for being a great teacher academically but also for caring for us in life. I also want to thank my fellow students for offering me such a friendly studying environment. We discussed homework problems, debugged each other's R codes, went to conferences, and had so much fun at uncountable departmental activities.

I want to thank my parents, who have created every opportunity they can for me to explore the world. They are always there for me: they were the ones seeing

me off at the airport when I moved across the ocean to pursue my dream; they were the ones who came to visit me, cooked for me when I was busy dealing with my schoolwork; and they were the ones who stepped up and took care of Daven when I needed them to. They have never missed any big events in my life, and I am so thankful that I have the best parents in the whole world. There is also a special person that I want to thank, Bowen, the rock of my life. I am grateful to have known him from the first day I came to the U.S. , and he has accompanied me and shared all the laughter and tears with me ever since. I wish him the best of luck with his own dissertation!

Academic work is not everything. I have met many interesting and inspiring people at Baylor, and I have experienced different cultures and traveled to many different places during this time. Life always goes on. I am so grateful to have a balanced personal and academic life. Some of the most important things in my life happened during these six years. I got married to a wonderful man, welcomed my first baby to the world, and will get my highest academic degree from Baylor. But it doesn't end here. Now, I am ready for the new adventure. I hope it is yet another unforgettable chapter in my life.

DEDICATION

To My Family

My Dear Parents and Husband

and Daven, my driving force

CHAPTER ONE

Introduction

It is common to encounter the problem of unmeasured confounding in the large database, especially large health care utilization databases. McCandless et al. (2007a), Corrao et al. (2012) and many others have addressed this issue. Bayesian approaches to accounting for unmeasured confounding have recently drawn increasing interest due to their flexibility with regards to bringing in information on the unmeasured confounder-treatment and unmeasured confounder-response relationship. But Bayesian analysis can be computationally intensive for certain models where MCMC methods are commonly used for inference. So we present in this dissertation a normal theory approximation as an alternative way to the time consuming MCMC simulations. Under most circumstances, these two methods will yield at similar results.

1.1 MCMC vs. Normal Approximation with Unmeasured Cconfounding

When dealing with large databases, one must be particularly aware of the effect of unmeasured confounding on statistical models. Confounding arises when factors unrelated to the particular study have a hidden effect on observed health outcomes. The potential causal effect estimates will be biased without proper adjustment and there is a lot of work studying unmeasured confounding in the epidemiology or other field of healthcare related areas.

There are a few methods to correct for unmeasured confounding. We can depend on external validation data, which is the information on unmeasured confounder from previous studies. But it generally does not contain information about disease and confounder relationship. We can use internal validation data, but we

will need to undertake a large amount of extra effort to ascertain the unmeasured confounder for a randomly selected subset of the main study data.(see Steyerberg et al., 2003, for detailed validation)There is also sensitivity analysis that assumes values of the "bias" parameters are plugged in to determine any potential effect of the unmeasured confounding. Last but not least, we can use informative priors and assign probability distributions to previously assumed fixed values.

Bayesian method provides an operational way to combine validation data and expert opinion in order to estimate parameters, allowing the researcher to control confounding through the inclusion of additional information from independent data sets. In this dissertation, we have used a small sample of validation data to make available more information about unmeasured confounding, and we have focused on a single binary unmeasured confounder because this is the most commonly assumed situation (see Gustafson and McCandless, 2010) and it is the "worst case scenario" of matched pairs. We first study a regression model with binary and normal response outcome and a logit link between the unmeasured confounding and covariate.

It is a standard Bayesian practice to use Markov chain Monte Carlo (MCMC) methods. The increase in generality of these methods comes at the price of requiring an assessment of convergence of the Markov chain to its stationary distribution, and this takes a long time. So we seek an alternative way under the same model.

When the number of data points is fairly large, the likelihood will be quite peaked, and small changes in the priors will have little effect on the posterior distributions. So in this situation, the posterior distribution can be approximated by a normal distribution.

Suppose $X_1, \cdots, X_n \overset{iid}{\sim} f_i(x_i|\boldsymbol{\Theta})$, so we have the likelihood function $f(\boldsymbol{x}|\boldsymbol{\Theta}) = \prod_{i=1}^{n} f_i(x_i|\boldsymbol{\Theta})$. $\pi(\boldsymbol{\Theta})$ is the prior and $f(\boldsymbol{x}|\boldsymbol{\Theta})$ is twice differentiable near $\hat{\Theta}^{\pi}$, the posterior mode of $\boldsymbol{\Theta}$. Then under appropriate regularity conditions, the posterior distribution $p(\boldsymbol{\Theta}|\boldsymbol{x})$ can be approximated by a normal distribution having mean

equal to the posterior mode and covariance matrix equal to minus the inverse Hessian of the log posterior evaluated at the mode. The matrix is also known as the "generalized" observed Fisher information matrix for $\boldsymbol{\Theta}$.

If the prior is considered flat, then the posterior mode $\hat{\Theta}^{\pi}$ can be replaced by the MLE $\hat{\Theta}$, and the covariance matrix by the observed Fisher information matrix. Alternatively, we might replace the posterior mode by the posterior mean, and replace the variance estimate based on observed Fisher information with the posterior covariance matrix, or even the expected Fisher information matrix $I(\hat{\Theta})$.

Just as the Central Limit Theorem enables a broad range of frequentist inference, this enables a broad range of Bayesian inference by showing that the posterior distribution of a continuous parameter $\boldsymbol{\Theta}$ is also asymptotically normal. And for this reason, this property is sometimes referred to as Bayesian Central Limit Theorem.(see Carlin and Louis, 2008)

A prospective study may be used to examine the associations between a binary disease state and various exposures. And a complication when investigating potential risk factors is misclassification of discrete exposure variables, which can occur due to forgetfulness or false reporting. Analyses that ignore this problem may give biased estimates and standard errors that are falsely small. We present a simple example of this kind to illustrate the application of the MCMC and normal approximation approaches.

1.2 MCMC vs. Normal Approximation under Gamma Distributed Data

To further explore the case with a continuous outcome, we study an outcome with a gamma distribution of Bayesian MCMC method and normal approximation analysis.

Health care expenditure data are usually right skewed with variability increasing as the mean costs increases. Many past studies of health care costs and their

responses to health insurance, treatment modalities or patient characteristics indicate that estimates of mean responses may be quite sensitive to how estimators treat the skewness in the outcome and other statistical problems that are common in such data. It has been suggested that those cost data frequently have a log-normal or gamma distribution, so in this dissertation, we consider the gamma response.

We let X and U denote dichotomous random variables taking values 1 or 0 to indicate the whether or not to use a certain medical technology and the unmeasured confounder, respectively. We use the factorization $P(Y,U|X) = P(Y|X,U)P(U|X)$ and model the confounding effect of U using a logistic regression model $\log P(Y|X,U) = \beta_0 + \beta_1 X + \lambda U$ and the underlined relationship between the unmeasured confounder and the covariate is $\text{logit}P(U = 1|x) = \gamma_0 + \gamma_1 X$. Here, the variables X and U are assumed to not interact in their effect on Y. And out of n observations, we have n_1 validation sample points, same as before.

The outcome variable Y is denoted to be health care expenditure, which has a gamma distribution due to the skewness. We keep diffused normal distributions for the independent priors on βs, γs and λ so all the parameters have non-informative priors. We are interested in β_1 as it is the coefficient for x, which represents the impact of the covariate on the health care expenditure.

From the results of the MCMC method and the normal approximation approach, we can see that the difference in the estimators $\hat{\beta}_1$s and the standard error are somehow significant. Hence, the normal theory approximation does not perform as well on skewed data and we should not apply it when dealing with strongly skewed datasets.

1.3 Cost-effectiveness

There is a growing expectation from health care policy makers that evidence supporting the cost-effectiveness of new health care interventions, particularly phar-

maceuticals, be provided along with the customary data on efficacy and safety. One approach of cost-effectiveness analysis(CEA) combines health care utilization data collected on individual patients with the appropriate price weights yield a measure of cost for each patient. Measuring effectiveness and cost at the patient level permits the use of more conventional methods of statistical inference to quantify the uncertainty due to the sampling and the measurement error.

Numerous articles have been published in the area of the statistical analysis of cost-effectiveness data. Initially, efforts were concentrated on providing confidence intervals for incremental cost-effectiveness ratios(ICER), but more recently, the concept of incremental net benefit(INB) has been proposed as an alternative.

In a CEA, whether an ICER or an INB approach is taken, five parameters need to be estimated. Two of the parameters are the differences between treatment arms of mean effectiveness and costs, denoted by Δ_e and Δ_c, respectively. The other three parameters are the variances and covariance of those estimators. And INB has been often used to explore the policy interpretation of CEA. The incremental net benefit is a function of λ, and is defined as $\lambda\Delta_e - \Delta_c$. It is called the incremental net benefit because it is the difference between incremental value ($\lambda\Delta_e$) and incremental cost (Δ_c). Treatment is cost-effective if, and only if, $INB > 0$ (see Willan and Briggs, 2006).

There has been considerable interest in the joint modelling of cost and effectiveness, some of the applications in clinical trials are O'Hagan et al. (2001) where they assume both cost and effectiveness are normally distributed; Negrin et al. (2010) where they also assume both cost and effectiveness are normally distributed, and performed a Bayesian model averaging.

In this dissertation, we follow Grieve et al. (2010) and Thompson and Nixon (2005) to assume gamma distribution for costs and normal distribution for effectiveness since the cost data are often skewed. We mainly focus on both Bayesian

estimation procedures and the normal theory approximation, and also compare their powers. All the models of the Bayesian analysis are straightforward to fit in the freely available package `R2WinBUGS`.

1.4 Outline of Dissertation

The concepts and methods mentioned here will be discussed in more detail in the next chapters. The remainder of the dissertation is organized as follows. Both Bayesian MCMC method and normal approximation method for estimating covariates with the existence of unmeasured confounding for binary and continuous(normal) distributed data are given in Chapter 2. In Chapter 3, we expand the model of normal responses to gamma responses, and assess the performance of both methods through simulation. A further cost-effectiveness analysis presenting incremental net benefit are given via both Bayesian and normal approximation methods in Chapter 4, and the dissertation is concluded with some final remarks in Chapter 5.

CHAPTER TWO

A Comparison of Power in a Normal Theory Approximation to a Bayesian MCMC Procedure

Bayesian sample size determination can be computationally intensive for models where Markov chain Monte Carlo (MCMC) methods are commonly used for inference. In this chapter, we present a normal theory approximation as an alternative way to the time consuming MCMC simulation methods in sample size determination. Cheng et al. (2009) developed a Bayesian approach to average power calculations in binary regression models. They applied it to a common medical scenario, investigating the impact of misclassification. We investigate a normal theory approximation method to a variety of different models, but in this chapter, we focus on the normal theory approximation to speed up sample size determination when an unmeasured confounder exists. We compare the estimation and resulting powers to the ones in the previously mentioned MCMC method. We will do this for both binary and continuous outcomes. The method is applicable to other complicated scenarios as well, such as misclassification and covariate measurement error models.

2.1 Introduction

It is well known that the determination of posterior distributions comes down to the evaluation of complex integrals, and the posterior summaries often involve computing moments or quantiles, which leads to more integration. Some early solutions involved using asymptotic methods to obtain analytic approximations to the posterior density. One simple way is to use a normal approximation to the posterior. It is essentially a Bayesian version of the Central Limit Theorem.(Carlin and Louis, 2008) When the approximate methods are intractable, we resort to numerical inte-

gration. This application is limited to models of low dimensions due to the so-called "curse of dimensionality".

The standard Bayesian MCMC methods are often time consuming. In order to be more efficient while maintaining the power and other features, in this chapter, we apply the normal approximation method from Carlin and Louis (2008) to several different models and try to determine the scenarios where the normal approximation is feasible. Here, we focus on the models with unmeasured confounding and consider other models with non-sampling bias as well.

2.1.1 Unmeasured Confounding

A confounding variable in a statistical model is a prognostic variable that correlates with other variables. One definition of confounder in clinical trial or epidemiology is that the factor must meet the following criteria: First, it must be a risk factor, which is a cause of a disease, or a surrogate measure of a cause in unexposed people. Secondly, it should be correlated, either positively or negatively, with the exposure in the study population. If the study population is classified into exposed and unexposed groups, then this factor must have different distributions (prevalence) in the two groups. Finally, the factor can not be affected by the exposure.

As stated by Vandenbroucke (2002):

> Confounding is the problem of confusing or mixing of exposure effects with other "extraneous" effects: If at the time of its occurrence, an exposure was associated with pre-existing risk for the outcome, its association would reflect at least in part the effect of this baseline association, not the effect of the exposure itself.(P.217)

In this scenario, the portion of the association reflecting this baseline association is confounding and the factors responsible for this confounding (those producing the differences in baseline risk) are confounders.

Dealing with confounders is relatively easy if we know what they are. But confounding from unmeasured variables is a common situation in many observational

studies, and without proper adjustments, the estimates of the effects will be biased. Currently the best defence against unknown confounders is randomization. We can also use Bayesian methods when unmeasured confounders exist because the analyses may involve synthesis of multiple sources of empirical evidence in the form of Meta-analysis. But the models are usually presented along with a family of prior information of a possible unknown unmeasured confounder. In cases where the model for unmeasured confounding is not identifiable, then the standard large sample theory for Bayesian analysis is not applicable. Consequently, the impact of different choices of prior distributions is unknown.(see McCandless et al., 2007b). We focus on the case where validation data does allow for estimation of all parameters.

2.1.2 Diagnostic Testing

In statistics, it is fundamental to model the relationship between explanatory and response variables. To diagnose a patient, the diagnosis, y, is modelled using explanatory variables $X = x_1, x_2, \cdots, x_k$ as $y = f(x_1, x_2, \cdots, x_k)$. Linear regression is often used to find the relationship between a predictor variable and a single response variable. However, the response variable is often not a numerical value. Instead, it can simply be a designation of one of two possible outcomes (a binary response) e.g. alive or dead, cured or not.

Data involving binary responses abound in just about every discipline from the natural sciences, to medicine, to education, etc. There are many examples of binary response data. For instance, using hatching environment(temperature, humidity) during the period when the eggs are incubated to predict the sex of turtles; using various demographic and credit history variables to predict if an individual will be a good or bad credit risk; using tests for trend to learn the toxicity of different doses. All of these involve the idea of prediction of a probability, chance, proportion or percentage. What we are trying to predict is bounded below by 0 and above by

1 (or 100%). The prediction technique we use in the analyses of binary response problems is logistic regression, with model

$$E(y_i|x_i) = \pi_i = \frac{e^{\beta_0+\beta_1 x_1}}{1 + e^{\beta_0+\beta_1 x_1}}.$$

When dealing with a binary response variable, the expected response is more appropriately modelled by curved relationships with the predictor variables. One such curve is given by the logistic model above. The logistic function is bounded between zero and one, which will eliminate the possibility of getting nonsensical predictions of probabilities. Also, there is a linear model hidden inside the function that can be revealed with a proper inverse transformation. Other options including the probit and the complementary log-log models can also be used.

It is common in many disciplines for binary responses to be measured with error. Several examples include mammography, a fallible screening test to detect breast cancer, the Pap smear, a screening test for cervical cancer, cognitive tests for dementia, and Prostate-Specific Antigen Screening for prostate cancer. Ideally, screening tests should have perfect sensitivity so that no opportunities for early intervention are missed. But in reality, no screening tests are perfect. Imperfect sensitivity of a screening test can lead to inappropriate security when false-negative results are obtained. Conversely, imperfect specificity will lead to negative consequences of labeling and unnecessary follow-up tests in healthy patients.

McInturff et al. (2004) evaluated the effects of a smoking cessation program among pregnant women controlling for variables such as age and smoking history. It is well known that smoking during pregnancy has adverse health implications. Moreover, since smoking cessation was self-reported, there was potential for women in the study to falsely report their smoking status. In another study, Roy et al. estimated the proportion of cancer deaths in Japan controlling for radiation exposure, the binary response had the potential to be measured with error in the form of both false positive and false negative misclassification. Accounting for misclassification

in the response of a binomial regression adds another layer of complexity to both data analysis and experimental design, in particular to sample size determination. Cheng et al. (2009) addressed this issue in sample size estimation for binomial and multinomial regression problems using the Bayesian paradigm.

2.1.3 Validation Data and Normal Theory Approximation

In this chapter, in order to correct for bias, we incorporate validation data. Data on a potential "unmeasured confounder" can be obtained for a randomly selected small subset of the original sample, and Bayesian modelling can utilize this additional information to perform analyses in order to quantitatively assess the potential impact of unmeasured confounding. We develop a Bayesian regression model to use the internal validation data as informative prior distributions for all parameters, retaining information on the correlations between the confounder and other covariates.

There are two different types of validation data. If the validation data is a random sample from the current "main study", the data is referred to as internal validation, it provides information on all parameters of interest but is usually expensive. If the subsample is from a previous study or database, it is referred to as external validation, which generally provides information on the unmeasured confounder/exposure relationships, but it requires the assumption of transportability.

It is a standard Bayesian practice to use Markov chain Monte Carlo(MCMC) methods, which operate by sequentially sampling parameter values from a Markov chain whose stationary distribution is exactly the joint posterior distribution of interest as desired. The increase in generality of these methods comes at the price of requiring an assessment of convergence of the Markov chain to its stationary distribution, and this takes time. Hence, we seek an alternative approach for cases when shorter computing time is required.

When we have a fairly large sample, the likelihood will be quite peaked. Small changes in the priors will only result little effect on the posterior distributions. So in this situation, the posterior distribution can be approximated by a normal distribution having mean equal to the posterior mode and covariance matrix equal to minus the inverse Hessian of the log posterior evaluated at the mode. If the prior is flat, the posterior mode can be replaced by the MLE, and the covariance matrix by the observed Fisher information matrix. Just as the Central Limit Theorem enables a broad range of frequentist inference, this enables a broad range of Bayesian inference by showing that the posterior distribution of a continuous parameter $\boldsymbol{\theta}$ is also asymptotically normal.(see Carlin and Louis, 2008, for more details)

2.2 Power Comparison Scheme

2.2.1 The Bayesian Regression Model with Unmeasured Confounding

A Bayesian sensitivity analysis for unmeasured confounding is considered in this section. Here, we assume we have an observational study. Our exposure and response are binary, we also assume a measured confounder and a single binary unmeasured confounder. The association between them can be formulated using a logistic regression model.

Now we denote the outcome variable to be Y. We denote the covariates $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{U}$, respectively. Specifically, Y is a binary response where $Y = 1$ denotes diseased and $Y = 0$ is non-diseased. X is the covariate of interest, generally exposure, $\boldsymbol{Z}$ is a vector of other measured covariates, and U is an unmeasured confounder.

Now we can write a logistic regression model for the probability of disease

$$logit\ P(Y = 1|X, U, Z) = \beta_0 + \beta_1 X + \beta_2 Z + \lambda U \tag{2.1}$$

Since the unmeasured confounder U is binary, we also have a logistic model for U:

$$logit\ P(U = 1|X, Z) = \gamma_0 + \gamma_1 X. \tag{2.2}$$

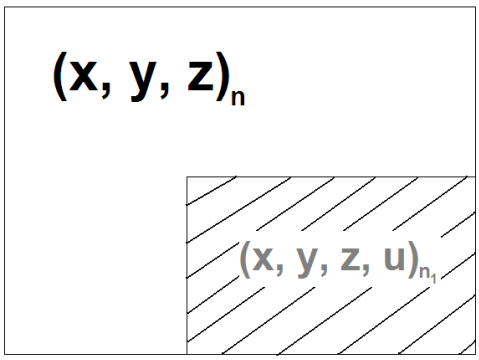

Figure 2.1: Sampling assumption setting: we have a sample of size n sample of (x, y, z) and a sub sample of n_1 of (x, y, z, u).

with all the priors information listed below:

$$\beta_0, \beta_1, \beta_2 \sim N(\mu_\beta, \sigma_\beta^2)$$

$$\gamma_0, \gamma_1 \sim N(\mu_\gamma, \sigma_\gamma^2)$$

$$\lambda \sim N(\mu_\lambda, \sigma_\lambda^2).$$

Here, we assume diffuse priors, specifically, $\mu_\beta = \mu_\gamma = \mu_\lambda = 0, \sigma_\beta^2 = \sigma_\gamma^2 = \sigma_\lambda^2 = 100$. Alternatively, if prior information on these parameters is available, it can be incorporated into the analysis.

In this model, the variables X and U are assumed to not interact. Finally, among the sample size of n observations, we assume there is a subset of n_1 observations that we have additional knowledge of u, the value of the otherwise unmeasured confounder. This type of validation data is referred to as internal. The methodology we develop here can be modified to handle external validation data as well.

Our primary interest is to estimate β_1 as it is the coefficient for x, which represents the impact of the covariate on the probability of getting the disease. The parameter λ represents the impact of the unmeasured confounder on the probability of getting the disease. Since u is unmeasured in general, λ is often referred to as a bias parameter.

The full likelihood function is

$$
\begin{aligned}
&L(\beta_0, \beta_1, \beta_2, \lambda, \gamma_0, \gamma_1 | x, y, z, x_1, y_1, z_1, u_1) \\
&= \prod_{i=1}^{n-n_1} [f(y_i|x_i, u=1, z_i) f(u=1|x_i) + f(y_i|x_i, u=0, z_i) f(u=0|x_i)] \prod_{j=1}^{n_1} [f(y_{1j}|x_{1j}, u1j, z_{1j}) \\
&= \prod_{i=1}^{n-n_1} \left[\frac{\exp\{y_i(\beta_0 + \beta_1 x_i + \beta_2 z_i + \lambda)\}}{1 + \exp\{y_i(\beta_0 + \beta_1 x_i + \beta_2 z_i + \lambda)\}} \frac{\exp\{\gamma_0 + \gamma_1 x_i\}}{1 + \exp\{\gamma_0 + \gamma_1 x_i\}} \right. \\
&\qquad \left. + \frac{\exp\{y_i(\beta_0 + \beta_1 x_i + \beta_2 z_i)\}}{1 + \exp\{y_i(\beta_0 + \beta_1 x_i + \beta_2 z_i)\}} \frac{1}{1 + \exp\{\gamma_0 + \gamma_1 x_i\}} \right] \\
&\quad \times \prod_{j=1}^{n_1} \left[\frac{\exp\{y_{1j}(\beta_0 + \beta_1 x_{1j} + \beta_2 z_{1j} + \lambda u_{1j})\}}{1 + \exp\{y_{1j}(\beta_0 + \beta_1 x_{1j} + \beta_2 z_{1j} + \lambda u_{1j})\}} \frac{\exp\{u_{1j}(\gamma_0 + \gamma_1 x_{1j})\}}{1 + \exp\{\gamma_0 + \gamma_1 x_{1j}\}} \right].
\end{aligned}
\tag{2.3}
$$

The joint posterior is the product of the likelihood and the prior. We use Markov Chain Monte Carlo method for inference. We use the `WinBUGS` software and the `R`(and package `R2WinBUGS`) to perform the simulations.

2.2.2 The Normal Approximation Approach

As mentioned previously, in order to perform a normal approximation, we need the posterior modes along with the Fisher information matrix. Thus we must first derive the log likelihood function of all parameters involved: $\beta_0, \beta_1, \lambda, \gamma_0, \gamma_1$. Since we have n_1 validation data points, the log likelihood function consists of two parts. The first part is $\log L_N$, the log likelihood function of the data where we did not observe the confounding u. The second part is $\log L_{N_1}$, the log likelihood function of the validation data where we observed the "unmeasured" confounder. To make things easier to track, we denote the validation sample as (x_1, y_1, z_1, u_1).

After taking the log of Equation(2.3) and some further simplification, we have the following results:

$$
\begin{aligned}
\log L_N = &\sum \{y(\beta_0 + \beta_1 x + \beta_2 z) - \log[1 + \exp(\gamma_0 + \gamma_1 x)] \\
&+ \log[1 + \exp(\gamma_0 + \gamma_1 x + \lambda y) + \exp(\gamma_0 + \beta_0 + (\beta_1 + \gamma_1)x + \beta_2 z + \lambda y)
\end{aligned}
$$

$$+ \exp(\beta_0 + \beta_1 x + \beta_2 z + \lambda)] - \log\left[1 + \exp(\beta_0 + \beta_1 x + \beta_2 z + \lambda)\right]$$
$$- \log[1 + \exp(\beta_0 + \beta_1 x + \beta_2 z)]\}$$

$$\begin{aligned}\log L_{N_1} = \sum \{ & y_1(\beta_0 + \beta_1 x_1 + \beta_2 z_1 + \lambda u_1) \\ & - \log\left[1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 z_1 + \lambda u_1)\right] \\ & + u_1(\gamma_0 + \gamma_1 x_1) - \log[1 + \exp(\gamma_0 + \gamma_1 x_1)]\}\,.\end{aligned}$$

The total log likelihood function is the two parts above combined,

$$\log L \;\; = \;\; \log L_N + \log L_{N_1}.$$

In order to find the posterior modes, we next differentiate the log-likelihood with respect to the parameter vector and set the resulting gradient vector to zero. Then we solve the system of equations to find extreme. We can also take the second derivative to the function to make sure that we have a maximum rather than a minimum.

First, we denote

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \\ \lambda \end{pmatrix}, \text{ where } \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \text{ and } \boldsymbol{\gamma} = \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix}.$$

Also denote

$$\boldsymbol{P}_i = \begin{pmatrix} 1 \\ x_i \\ z_i \end{pmatrix}, \boldsymbol{Q}_j = \begin{pmatrix} 1 \\ x_{1j} \\ z_{1j} \end{pmatrix}.$$

So we have $\boldsymbol{P}_i'\boldsymbol{\beta} = \beta_0 + x_i\beta_1 + z_i\beta_2,\ \boldsymbol{Q}_j'\boldsymbol{\beta} = \beta_0 + x_{1j}\beta_1 + z_{1j}\beta_2$.

We provide the derivations required for the $\boldsymbol{\beta}$ parameters here and those for $\boldsymbol{\gamma}$ and λ are provided in the appendix. We now take the first derivative of $\log L$ with respect to $\boldsymbol{\beta}$

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \frac{\partial \log L_N}{\partial \boldsymbol{\beta}} + \frac{\partial \log L_{N_1}}{\partial \boldsymbol{\beta}} \tag{2.4}$$

$$\begin{aligned}
\frac{\partial \log L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n} \Bigg\{ \frac{y_i \boldsymbol{P}_i}{\left(\frac{e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}} + \frac{e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}} \right)} \\
&\quad \times \left[-\frac{e^{2y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{\left(1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}\right)^2} + \frac{e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}} \right. \\
&\quad \left. -\frac{e^{2y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{\left(1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}\right)^2} + \frac{e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}} \right] \Bigg\} \\
&\quad + \sum_{j=1}^{n_1} \Bigg\{ \left(y_{1j}\boldsymbol{Q}_j\right) e^{-y_{1j}(\boldsymbol{Q}_j'\boldsymbol{\beta})-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \\
&\quad \times (1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)})(1+e^{\gamma_0+x_{1j}\gamma_1}) \\
&\quad \times \left(-\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2 (1+e^{\gamma_0+x_{1j}\gamma_1})} \right. \\
&\quad \left. + \frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)(1+e^{\gamma_0+x_{1j}\gamma_1})} \right) \Bigg\}
\end{aligned}$$

For this specific individual regression parameters we have

$$\begin{aligned}
\frac{\partial \log L}{\partial \beta_0} &= \sum_{i=1}^{n} \Bigg\{ \frac{y_i}{\left(\frac{e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}} + \frac{e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}} \right)} \\
&\quad \times \left[-\frac{e^{2y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{\left(1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}\right)^2} + \frac{e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}} \right. \\
&\quad \left. -\frac{e^{2y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{\left(1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}\right)^2} + \frac{e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}} \right] \Bigg\} \\
&\quad + \sum_{j=1}^{n_1} \Bigg\{ y_{1j} e^{-y_{1j}(\boldsymbol{Q}_j'\boldsymbol{\beta})-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \\
&\quad \times (1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)})(1+e^{\gamma_0+x_{1j}\gamma_1}) \\
&\quad \times \left(-\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2 (1+e^{\gamma_0+x_{1j}\gamma_1})} \right.
\end{aligned}$$

$$+ \frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\Bigg)\Bigg\}$$

$$\begin{aligned}
\frac{\partial \log L}{\partial \beta_1} &= \sum_{i=1}^{n}\Bigg\{\frac{x_i y_i}{\left(\frac{e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}+\frac{e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}}\right)} \\
&\quad\times\Bigg[-\frac{e^{2y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{\left(1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}\right)^2}+\frac{e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}} \\
&\quad-\frac{e^{2y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{\left(1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}\right)^2}+\frac{e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}}\Bigg]\Bigg\} \\
&\quad+\sum_{j=1}^{n_1}\Bigg\{x_{1j}y_{1j}e^{-y_{1j}(\boldsymbol{Q}_j'\boldsymbol{\beta})-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \\
&\quad\times(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)})(1+e^{\gamma_0+x_{1j}\gamma_1}) \\
&\quad\times\Bigg(-\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \\
&\quad+\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\Bigg)\Bigg\}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \log L}{\partial \beta_2} &= \sum_{i=1}^{n}\Bigg\{\frac{y_i z_i}{\left(\frac{e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}+\frac{e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}}\right)} \\
&\quad\times\Bigg[-\frac{e^{2y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{\left(1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}\right)^2}+\frac{e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}}{1+e^{y_i(\boldsymbol{P}_i'\boldsymbol{\beta})}} \\
&\quad-\frac{e^{2y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{\left(1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}\right)^2}+\frac{e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})+\gamma_0+x_i\gamma_1}}{1+e^{y_i(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta})}}\Bigg]\Bigg\} \\
&\quad+\sum_{j=1}^{n_1}\Bigg\{y_{1j}z_{1j}e^{-y_{1j}(\boldsymbol{Q}_j'\boldsymbol{\beta})-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \\
&\quad\times(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)})(1+e^{\gamma_0+x_{1j}\gamma_1}) \\
&\quad\times\Bigg(-\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}
\end{aligned}$$

$$+ \frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\Bigg)\Bigg\}.$$

The next step is to set the above fractional polynomials to zero and solve for the corresponding βs in order to find the MLE. Since they are not of closed form, it is impossible to do the calculation manually. We will use numerical methods to search for the maximum. Specifically, we use the R function `optim` with default setting `BFGS` method.

For the purpose of finding var($\boldsymbol{\beta}$), we need to calculate the inverse of the Information matrix, and use the fact that

$$\text{var}(\boldsymbol{\Theta}) = [I(\boldsymbol{\Theta})]^{-1}$$

where $\boldsymbol{\Theta}$ is the parameter vector that $\boldsymbol{\Theta} = (\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1, \lambda)$, and $I(\boldsymbol{\Theta})$ is the Fisher's information matrix.

The Information matrix is the negative of the expected value of the Hessian matrix:

$$I(\boldsymbol{\Theta}) = -E[H(\boldsymbol{\Theta})].$$

We now find the Hessian Matrix. The Hessian is the matrix of second derivatives of the likelihood with respect to the parameter:

$$H(\boldsymbol{\Theta}) = \frac{\partial^2 \log L}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}'}$$

Thus, the variance-covariance matrix of maximum likelihood estimator of $\boldsymbol{\Theta}$ is:

$$\begin{aligned}
\text{Var}(\boldsymbol{\Theta}) &= [I(\boldsymbol{\Theta})]^{-1} \\
&= (-E[H(\boldsymbol{\Theta})])^{-1} \\
&= (-E[\frac{\partial^2 \log L}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}'}])^{-1}
\end{aligned}$$

One large sample property of this estimator is that $\boldsymbol{\Theta} \sim N[\hat{\Theta}, I(\hat{\Theta})^{-1}]$ asymptotically. The standard errors of the estimators are the square roots of the diag-

onal terms in the variance-covariance matrix. Because we are interested in $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$, specifically β_1, we also have the property $\boldsymbol{\beta} \sim N[\hat{\beta}, I(\hat{\beta})^{-1}]$ asymptotically. Now we are going to calculate the second derivatives in order to obtain the $\boldsymbol{\beta}$-block in the variance-covariance matrix

$$\begin{bmatrix} \mathrm{Var}(\boldsymbol{\beta}) & \cdots & \cdots & \cdots \\ \mathrm{Cov}(\boldsymbol{\beta}, \lambda) & \mathrm{Var}(\lambda) & \cdots & \cdots \\ \cdots & \cdots & \mathrm{Var}(\gamma_0) & \cdots \\ \cdots & \cdots & \cdots & \mathrm{Var}(\gamma_1) \end{bmatrix},$$

$$\text{where } \mathrm{Var}(\boldsymbol{\beta}) = \begin{bmatrix} \mathrm{Var}(\beta_0) & \cdots & \cdots \\ \mathrm{Cov}(\beta_0, \beta_1) & \mathrm{Var}(\beta_1) & \cdots \\ \cdots & \cdots & \mathrm{Var}(\beta_2) \end{bmatrix}.$$

Similar as before,

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}^2} = \frac{\partial^2 \log L_N}{\partial \boldsymbol{\beta}^2} + \frac{\partial^2 \log L_{N_1}}{\partial \boldsymbol{\beta}^2} \tag{2.5}$$

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \boldsymbol{\beta}^2} = & \sum_{i=1}^{n} \left(y_i^2 \boldsymbol{P}_i^2\right) \left[- \left(- \frac{e^{2y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{\left(1 + e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{1 + e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}} \right.\right. \\ & \left. - \frac{e^{2y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right) + \gamma_0 + x_i\gamma_1}}{\left(1 + e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right) + \gamma_0 + x_i\gamma_1}}{1 + e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right)}} \right)^2 \\ & \times \left(\frac{e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{1 + e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}} + \frac{e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right) + \gamma_0 + x_i\gamma_1}}{1 + e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right)}} \right)^{-2} \\ & + \left(\frac{2e^{3y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{\left(1 + e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^3} - \frac{3e^{2y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{\left(1 + e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{1 + e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}} \right. \\ & + \frac{2e^{3y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right) + \gamma_0 + x_i\gamma_1}}{\left(1 + e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^3} - \frac{3e^{2y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right) + \gamma_0 + x_i\gamma_1}}{\left(1 + e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^2} \\ & \left.\left. + \frac{e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right) + \gamma_0 + x_i\gamma_1}}{1 + e^{y_i\left(\lambda + \boldsymbol{P}_i'\boldsymbol{\beta}\right)}} \right) \end{aligned}$$

$$
\begin{aligned}
&\times \left(\frac{e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{1+e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}} + \frac{e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{1+e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)}} \right)^{-1} \Bigg] \\
&+ \sum_{j=1}^{n_1} \left(y_{1j}^2 \boldsymbol{Q}_j^2\right) \Bigg[e^{-y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \\
&\times \left(-\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2 \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&\left. + \frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \\
&- e^{-y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \\
&\times \left(-\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2 \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&\left. + \frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \\
&+ e^{-y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \\
&\times \left(\frac{2e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+2y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^3 \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&- \frac{3e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2 \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \\
&\left. + \frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \Bigg]
\end{aligned}
$$

Specifically, the second derivative of the log likelihood function with respect to the variable of interest β_1 is

$$
\frac{\partial^2 \log L}{\partial \beta_1^2} = \frac{\partial^2 \log L_N}{\partial \beta_1^2} + \frac{\partial^2 \log L_{N_1}}{\partial \beta_1^2}
$$

$$
\frac{\partial^2 \log L_N}{\partial \beta_1^2} = \sum_{i=1}^{n} - \left(-\frac{e^{2y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)} x_i y_i}{\left(1+e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)} x_i y_i}{1+e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}} \right.
$$

$$
\left.
-\frac{e^{2y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}x_iy_i}{\left(1+e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^2}+\frac{e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}x_iy_i}{1+e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}\right)^2
$$

$$
\times\left(\frac{e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{1+e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}+\frac{e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{1+e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}\right)^{-2}
$$

$$
+\left(\frac{2e^{3y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}x_i^2y_i^2}{\left(1+e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^3}-\frac{3e^{2y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}x_i^2y_i^2}{\left(1+e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^2}+\frac{e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}x_i^2y_i^2}{1+e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}\right.
$$

$$
+\frac{2e^{3y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}x_i^2y_i^2}{\left(1+e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^3}-\frac{3e^{2y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}x_i^2y_i^2}{\left(1+e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)}\right)^2}
$$

$$
\left.+\frac{e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}x_i^2y_i^2}{1+e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}\right)
$$

$$
\times\left(\frac{e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}{1+e^{y_i\left(\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}+\frac{e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{1+e^{y_i\left(\lambda+\boldsymbol{P}_i'\boldsymbol{\beta}\right)}}\right)^{-1}
$$

$$
+\sum_{j=1}^{n_1}x_{1j}^2y_{1j}^2\left[e^{-y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\right.
$$

$$
\times\left(-\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\right.
$$

$$
\left.+\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\right)
$$

$$
-e^{-y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)
$$

$$
\times\left(-\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\right.
$$

$$
\left.+\frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\right)
$$

$$
+e^{-y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)
$$

$$
\times\left(\frac{2e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+2y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^3\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\right.
$$

$$
\left. - \frac{3e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right.
$$

$$
\left. + \frac{e^{y_{1j}\left(\boldsymbol{Q}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{Q}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \Bigg] .
$$

Refer to the Appendix for the second derivatives involving λ and the γs, and all the second derivatives on the off diagonal of Var($\boldsymbol{\Theta}$).

In order to find the approximate variance, take the negative of the Hessian matrix, yielding the observed Information matrix, and calculate the inverse of the Information matrix to obtain the variance-covariance matrix for the MLE of $\boldsymbol{\Theta}$. The standard error for the MLE estimator of β_1 lies in the $\boldsymbol{\beta}$-block of the matrix. It is the square root of the second entry on the diagonal.

Again, we use the R function `optim` to compute the maximum likelihood estimator of the log likelihood function. The methods offered in the `optim` package are "`BFGS`", "`CG`", "`Nelder-Mead`", and "`SANN`". Among these, the "BFGS" method, developed by Broyden, Fletcher, Goldfarb and Shanno, is a quasi-Newton method which uses function values and gradients to build a picture of the surface to be optimized. The "CG" method is a conjugate gradient method, and the "SANN" method uses the Metropolis function for the acceptance probability. We use the default setting method "BFGS" in our simulations for this chapter. In the following chapters, we used a different minimization method developed by Byrd et al. (1995) in the R function. See the relative R code in the appendix.

Sample size determination can be based on a number of criteria such as interval width, posterior variability and power of a hypothesis test. Here, we focus on power, thus we compute posterior probabilities for testing the null hypothesis of $H_0 : \beta_1 = 0$ and the alternative hypothesis of $H_1 : \beta_1 > 0$ with significance level $\alpha = 0.05$. For the normal approximation, we standardize the parameter estimator by calculating

$$
z_0 = \hat{\beta}_1 / sd_{\beta_1},
$$

then obtain the corresponding percentile of z_0, and compare the percentile to $1-\alpha$. Ultimately, the power is computed by finding the proportion of times that z_0 exceeds $1-\alpha$ out of m iterations, namely the probability of making a correct decision of rejecting the null hypothesis.

2.3 Simulation

2.3.1 WinBUGS Simulation Algorithm

Our simulation algorithm is as follows:

(1) First, we generate n_1 values of the covariates x_1 and z_1 from binomial distributions with $p_x = 0.6$ and $p_z = 0.4$, respectively. Then using fixed value of γ_0 and γ_1, we generate n_1 values of u_1 from a binomial distribution according to Equation (2.2). We calculate the outcome y_1 from a binomial distribution according to Equation (2.1) using fixed $\beta_0, \beta_1, \beta_2$ and λ.

(2) We generate the $n - n_1$ values of x, z and u for the main study without validation, calculate the outcome y using the same parameters as in (1). Namely, x and z are binomial(0.6) and binomial(0.4), we use same fixed γ_0, γ_1 and Equation (2.2) to obtain u, and we use the same fixed $\beta_0, \beta_1, \beta_2, \lambda$ and Equation (2.1) to obtain y. Notice here, we do not observe u.

(3) Then, we fit the Bayesian model (2.2) and (2.1) to the data generated above. We use diffuse priors with mean at zero and precision at 0.1 for $\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1$ and λ.

(4) Finally, we approximate the posterior distribution of β_1 using WinBUGS while keeping track of the posterior probability value in each iteration, that is $P(\beta_1 > 0|\text{data})$. For the same data we also approximate this probability using the normal approximation.

(5) We then repeat the whole process for m interactions. We tally the number of times that the posterior probability value exceeds $1-\alpha$ out of m iterations. This approximates the Bayesian power achieved using MCMC method for sample size n.

2.3.2 Simulation Conditions

We consider a single dichotomous covariate x_1 generated from a binomial distribution with probability of success $p = 0.6$. We assume that risk increases with increasing x_1, which is reflected by sampling β_0 from uniform (-1.5, -1) distribution. We fix β_1 at 0.5. The other parameters are generated as follows: β_2 is from Unif(.2, .4), γ_0 is from Unif(-.3, -.2), γ_1 is from Unif(1, 1.2), and λ is from Unif(-1, -.5).

We then generate z and z_1 from a binomial distribution with probability of success $p = 0.4$, and generate u and u_1 from a binomial distribution under Equation (2.2). y and y_1 are from Equation (2.1).

The Bayesian MCMC simulations use 1000 data sets(that is repeat 1000 times) with discrete posterior approximation based on a Monte Carlo sample of 20000 posterior iterates after a 5000 initial burn-in with thinning equals to 2. The computational time varies and depends on different machines. The sample size also played an important role in the length of the simulations. Approximately, the MCMC method generally takes a few days to finish while the normal theory approximation only needs a few hours to run on the same regular PC.

2.3.3 Result for Bayesian MCMC Approach with Unmeasured Confounding

Table 2.1 below provides the result of using the Bayesian MCMC approach when unmeasured confounding is present. We have arranged different total sample sizes from 500 to 1200, with an increase of 100 for each simulation. We have set the validation sample size to be fixed at $n_1 = 300$ and the true β_1 value is set at 0.5.

For different total sample sizes, we record the mean values of β_1, the standard deviations of β_1 and the powers for each simulation. From the result, we can see that the point estimators in all sample size cases are quite close to the truth. The result also produces evidence showing an obvious trend that when we increase the sample size, the standard deviation of the estimate decreases and the power increases. If we have a large enough sample, namely 1200 data points, we can achieve a power over 0.9.

Table 2.1: MCMC result when $n_1 = 300, m = 1000$

n	Estimates(Truth = 0.5)	Standard Deviation	Power
500	0.5089131	0.2521596	0.654
600	0.5025989	0.2331040	0.705
700	0.5156427	0.2183066	0.797
800	0.5114063	0.2092231	0.821
900	0.5124188	0.1999642	0.855
1000	0.5054345	0.1910872	0.856
1100	0.5083408	0.1850259	0.894
1200	0.5158404	0.1799486	0.911

2.3.4 Result for Normal Approximation Approach with Unmeasured Confounding

We next provide the results when using normal approximation approach with unmeasured confounding presented. We use the same sample sizes ranging from 500 to 1200 as the above MCMC case, with exactly the same data generated with fixed $n_1 = 300$ and $\beta_1 = 0.5$. Hence, we can directly compare the results of the two methods. The point estimators in this approach are close to the truth in all sample sizes, and we again see a trend that when we increase the sample size, the power increases as well and the standard deviation of the estimate decreases.

The powers and standard deviations for both the MCMC and normal approximation simulations are quite similar indicating that for future work in sample size determination or other simulation based procedures.

Table 2.2: Normal Approximation when n1 = 300, m = 1000

n	Estimates(Truth = 0.5)	Standard Deviation	Power
500	0.5077775	0.2499366	0.656
600	0.5009355	0.2309771	0.711
700	0.5110463	0.2162722	0.790
800	0.5089363	0.2072116	0.817
900	0.5095178	0.1980417	0.851
1000	0.5032719	0.1892859	0.859
1100	0.5047532	0.1832769	0.896
1200	0.5133611	0.1782327	0.914

We put the powers obtained from the two methods together, and the graph in Figure 2.2 gives a visual representation of the similarity in the results of both methods.

2.4 Continuous Response

2.4.1 MCMC Method for Continuous Response

We next overview how these methods could be applied to a model with a continuous outcome that has a binary unmeasured confounder.

A similar model to the binary regression already considered can be used for a continuous outcome $\boldsymbol{Y}$. We assume the continuous response follows a normal distribution. Thus, y_j is distributed normally with mean μ_j, and standard deviation σ_j^2. All the other parameters remain unchanged: X is still the covariate of interest (generally exposure), $\boldsymbol{Z}$ is the vector of covariates, and U is the unmeasured confounder,

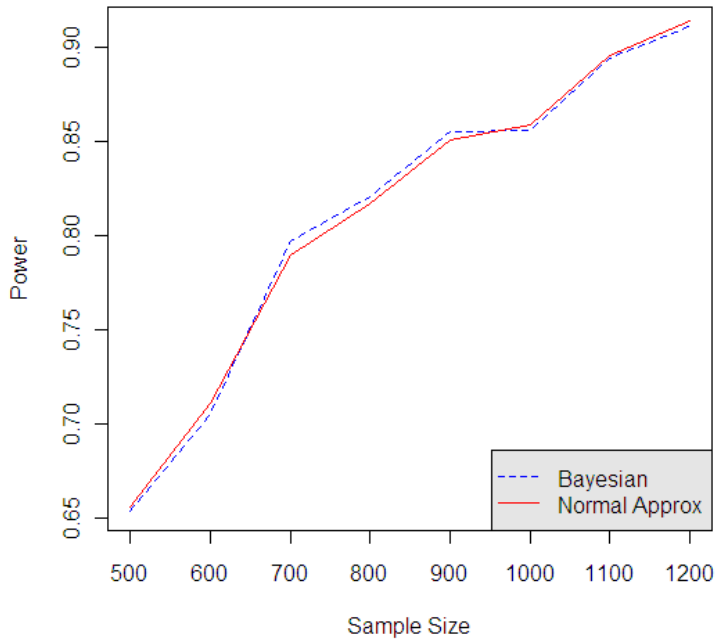

Figure 2.2: The power curves of Bayesian MCMC Method and Normal Approximation of binary responses

all of which stay binary. The mean function is modeled as

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \lambda U \tag{2.6}$$

$$logit\ P(U = 1|X, Z) = \gamma_0 + \gamma_1 X \tag{2.7}$$

where

$$Y_j \sim N(\mu_j, \sigma_j^2)$$

and all the priors' information listed below:

$$\beta_0, \beta_1, \beta_2 \sim N(\mu_\beta, \sigma_\beta^2)$$

$$\gamma_0, \gamma_1 \sim N(\mu_\gamma, \sigma_\gamma^2)$$

$$\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$$

$$\sigma_j \sim \text{Unif}(lo, up).$$

We have $\mu_\beta = \mu_\gamma = \mu_\lambda = 0, \sigma_\beta^2 = \sigma_\gamma^2 = \sigma_\lambda^2 = 100, lo = 0.01, up = 500$. Again, though we give diffuse priors to all the parameters in our simulations, expert opinion and/or prior data can be incorporated.

2.4.2 Normal Approximation

The derivation for the normal theory approximation to the continuous response is similar to the one with a binary outcome that we derived above. The likelihoods for the main study and the validation data are

$$\begin{aligned} L_N(\beta_0,\beta_1,\beta_2,\gamma_0,\gamma_1,\lambda,\sigma|x,y,z) &= \prod_{i=1}^{n}[f(y_i|x_i,u=1,z_i)f(u=1|x_i) \\ &\quad + \quad f(y_i|x_i,u=0,z_i)f(u=0|x_i)] \\ &= \prod_{i=1}^{n}\left[\frac{\exp\left\{\frac{-[y_i-(\beta_0+\beta_1x_i+\beta_2z_i+\lambda)]^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}}\right. \\ &\quad \times\frac{\exp\{\gamma_0+\gamma_1x_i\}}{1+\exp\{\gamma_0+\gamma_1x_i\}} \\ &\quad + \quad \frac{\exp\left\{\frac{-[y_i-(\beta_0+\beta_1x_i+\beta_2z_i)]^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}} \\ &\quad \left.\times\frac{1}{1+\exp\{\gamma_0+\gamma_1x_i\}}\right] \end{aligned}$$

$$\begin{aligned} L_{N_1}(\beta_0,\beta_1,\beta_2,\gamma_0,\gamma_1,\lambda,\sigma|x_1,y_1,z_1,u_1) &= \prod_{j=1}^{n_1}\left[\frac{\exp\left\{\frac{-[y_{1j}-(\beta_0+\beta_1x_{1j}+\beta_2z_{1j}+\lambda u_{1j})]^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}}\right. \\ &\quad \left.\times\frac{\exp\{u_{1j}(\gamma_0+\gamma_1x_{1j})\}}{1+\exp\{\gamma_0+\gamma_1x_{1j}\}}\right] \end{aligned}$$

Taking the log of these functions yields $\log L_N$ and $\log L_{N_1}$. The total log likelihood can be obtained by combining the two parts together.

$$\log L = \log L_N + \log L_{N_1}.$$

We will do the same calculation as previous sections to obtain the MLEs. First we differentiate the likelihood function with respect to the parameter vector and set the resulting gradient vector to zero. We then solve the system of equations to find the extreme.

We denote $\boldsymbol{\beta}, \boldsymbol{P}_i$ and $\boldsymbol{Q}_j$ the same as before:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad \boldsymbol{P}_i = \begin{pmatrix} 1 \\ x_i \\ z_i \end{pmatrix}, \quad \boldsymbol{Q}_j = \begin{pmatrix} 1 \\ x_{1j} \\ z_{1j} \end{pmatrix}.$$

The maximum likelihood estimator vector $\boldsymbol{\Theta}$ is

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \\ \lambda \\ \sigma^2 \end{pmatrix}.$$

By Equation (2.4), we take the first derivative of $\log L_N$ and $\log L_{N_1}$ with respect to $\boldsymbol{\beta}$, and add them together to get the first derivative of the log likelihood function with respect to $\boldsymbol{\beta}$.

$$\begin{aligned} \frac{\partial \log L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n} \left\{ \left(\frac{\boldsymbol{P}_i}{\sigma^2} \right) \right. \\ &\quad \times \left[e^{-\frac{(y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}} (y_i - \boldsymbol{P}_i'\boldsymbol{\beta}) + e^{-\frac{(-\lambda + y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2} + \gamma_0 + x_i\gamma_1} (-\lambda + y_i - \boldsymbol{P}_i'\boldsymbol{\beta}) \right] \\ &\quad \left. \times \left(e^{-\frac{(y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}} + e^{-\frac{(-\lambda + y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2} + \gamma_0 + x_i\gamma_1} \right)^{-1} \right\} \\ &\quad + \sum_{j=1}^{n_1} \boldsymbol{Q}_j \frac{-\lambda u_{1j} + y_{1j} - \boldsymbol{Q}_j'\boldsymbol{\beta}}{\sigma^2} \end{aligned}$$

We then take the first derivative of $\log L$ with respect to σ^2, and get log likelihood function with respect to σ^2.

$$\begin{aligned} \frac{\partial \log L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} \left[e^{-\frac{(y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}} (y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2 \right. \\ &\quad \left. + e^{-\frac{(-\lambda + y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2} + \gamma_0 + x_i\gamma_1} (-\lambda + y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2 \right] \\ &\quad \times \left(e^{-\frac{(y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}} + e^{-\frac{(-\lambda + y_i - \boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2} + \gamma_0 + x_i\gamma_1} \right)^{-1} \end{aligned}$$

$$-\frac{n_1}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{j=1}^{n_1}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{Q}_j'\boldsymbol{\beta}\right)^2$$

Because we are interested in β_1, the first derivative with respect to β_1 is

$$\begin{aligned}\frac{\partial \log L}{\partial \beta_1} &= \sum_{i=1}^{n}\Bigg\{\left(\frac{x_i}{\sigma^2}\right)\\ &\quad\times\left[e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})\right]\\ &\quad\times\left(e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}\right)^{-1}\Bigg\}\\ &\quad+\sum_{j=1}^{n_1}x_{1j}\frac{-\lambda u_{1j}+y_{1j}-\boldsymbol{Q}_j'\boldsymbol{\beta}}{\sigma^2}\end{aligned}$$

To find the MLE, we set the system of the first derivatives to zero and solve for $\boldsymbol{\Theta}$.

Again, we need the second derivatives in order to find the variance-covariance matrix. By Equation (2.5), we can calculate separate parts $\frac{\partial^2 \log L_N}{\partial \boldsymbol{\Theta}^2}$ and $\frac{\partial^2 \log L_{N_1}}{\partial \boldsymbol{\Theta}^2}$ first and then add them together.

$$\begin{aligned}\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}^2} &= \sum_{i=1}^{n}(\boldsymbol{P}_i^2)\Bigg\{-\frac{1}{\sigma^2}\\ &\quad-\left[\frac{e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})}{\sigma^2\left(e^{-\frac{\left(y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}}+e^{-\frac{\left(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}\right)}\right]^2\\ &\quad+\frac{e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{\sigma^4\left(e^{-\frac{\left(y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}}+e^{-\frac{\left(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}\right)}\Bigg\}\\ &\quad-\sum_{j=1}^{n_1}\frac{\boldsymbol{Q}_j}{\sigma^2}\end{aligned}$$

Specifically for β_1, we have

$$\frac{\partial^2 \log L}{\partial \beta_1^2} = \frac{\partial^2 \log L_N}{\partial \beta_1^2}+\frac{\partial^2 \log L_{N_1}}{\partial \beta_1^2}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}(x_i^2)\left\{-\frac{1}{\sigma^2}\right. \\
&\quad -\left[\frac{e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})}{\sigma^2\left(e^{-\frac{\left(y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}}+e^{-\frac{\left(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}\right)}\right]^2 \\
&\quad \left.+\frac{e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{\sigma^4\left(e^{-\frac{\left(y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}}+e^{-\frac{\left(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}\right)}\right\} \\
&\quad -\sum_{j=1}^{n_1}\frac{x_{1j}^2}{\sigma^2}.
\end{aligned}
$$

Now we find $\frac{\partial^2 \log L}{\partial(\sigma^2)^2}$.

$$
\begin{aligned}
\frac{\partial^2 \log L_N}{\partial(\sigma^2)^2} &= \sum_{i=1}^{n}\left\{\frac{3}{4\sigma^4}\right. \\
&\quad +\left(-\frac{1}{2\sigma^2}+\frac{e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^4\left(e^{-\frac{\left(y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}}+e^{-\frac{\left(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}\right)}\right)^2 \\
&\quad -\frac{3}{2\sigma^6}\frac{e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{\left(e^{-\frac{\left(y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}}+e^{-\frac{\left(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}\right)} \\
&\quad \left.+\frac{1}{4\sigma^8}\frac{e^{-\frac{(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}}(y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^4+e^{-\frac{(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta})^4}{\left(e^{-\frac{\left(y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}}+e^{-\frac{\left(-\lambda+y_i-\boldsymbol{P}_i'\boldsymbol{\beta}\right)^2}{2\sigma^2}+\gamma_0+x_i\gamma_1}\right)}\right\} \\
&\quad +\sum_{j=1}^{n_1}-\frac{1}{4\sigma^4}\left(1-\frac{(-\lambda u_{1j}+y_{1j}-\boldsymbol{Q}_j'\boldsymbol{\beta})^2}{\sigma^2}\right)^2 \\
&\quad +\frac{3}{4\sigma^4}-\frac{3(-\lambda u_{1j}+y_{1j}-\boldsymbol{Q}_j'\boldsymbol{\beta})^2}{2\sigma^6}+\frac{(-\lambda u_{1j}+y_{1j}-\boldsymbol{Q}_j'\boldsymbol{\beta})^4}{4\sigma^8}
\end{aligned}
$$

All the other off-diagonal values can be found in the appendix.

As in the logistic regression case, the variance is obtained by taking the negative of the inverse of the Hessian matrix.

2.4.3 Simulation Algorithm

The simulation steps we use are similar to the binary outcome scenario, with some minor changes.

(1) First, we generate the covariates for the measured confounder: x_1, z_1 from Binomial distributions $p_x = 0.6$ and $p_z = 0.4$ respectively. Then using fixed values of γ_0 and γ_1, we generate the covariates for measured confounder u_1 from a binomial distribution according to Equation (2.7). Now we generate the outcome y_1 from a normal distribution with mean $\beta_0 + \beta_1 X + \beta_2 Z + \lambda u_1$ using fixed $\beta_0, \beta_1, \beta_2, \lambda$ and standard deviation at fixed σ by Equation (2.6).

(2) Secondly, we generate the $n - n_1$ covariates x, z and u for the main study without validation. We calculate the outcome y using the same parameters to the last step, Namely, x and z are binomial(0.6) and binomial(0.4). Using same fixed γ_0, γ_1 and Equation (2.7), we obtain u, and using the same fixed $\beta_0, \beta_1, \beta_2, \lambda$ together with u and σ, we obtain y under Equation (2.6). Notice we do not actually observe u.

(3) Then, we fit the Bayesian model (2.6) and (2.7) to the generated data above. We use the diffuse normal priors with mean at zero and precision at 0.1 on $\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1$ and λ, and use a uniform(0.01, 500) prior on σ which is also flat.

(4) Finally, we approximate the posterior distribution of β_1 using WinBUGS while keeping track of $P(\beta_1 > 0|\text{data})$, the posterior probability values of β_1 greater than zero in each iteration. We also compute the normal approximation for this same data set.

(5) We then repeat the whole process for m iterations. We tally the number of times that posterior probability value exceeds $1-\alpha$ out of m iterations. This approximates the Bayesian power achieved with the MCMC method for sample size n. Similarly, the power can be computed for the normal approximation.

For the simulation, we generated 1000 data sets with discrete posterior approximations based on a Monte Carlo sample of 20000 posterior iterates after a 5000 initial burn-in with thinning equals to 2.

2.4.4 Continuous Response Example

In this section, we will see both methods' application in a single data set of a continuous response.

We let $\boldsymbol{Y}$ to be a continuous distributed variable which is generated from a normal distribution. The covariate of interest $\boldsymbol{X}$ and the unmeasured confounder U are both binary variables. We have a logistic regression model for $\boldsymbol{Y}$ and $\boldsymbol{U}$:

$$Y = \beta_0 + \beta_1 X + \lambda U, \quad logit\ P(U = 1|X) = \gamma_0 + \gamma_1 X$$

We generate β_0 from a uniform distribution between 100 and 150, fix the parameter of interest β_1 at 30, and we generate λ from a uniform distribution between -30 and -20. Also, the standard deviation σ of y is randomly generated from a uniform distribution between 150 and 175. The linear coefficients from Equation (2.7) are generated from uniform distributions (-0.3, -0.2) and (1.1, 1.3) respectively.

We assume to have a total of 500 samples within which 300 are validation samples, where we observe the unmeasured confounder u. In the MCMC method, we assume the priors of β_0, β_1 and λ to be a diffuse normal centered at zero with precision 0.0000001. The priors for γ_0 and γ_1 are diffuse normal centered at zero with precision 0.1, and the prior for σ is a flat uniform(0.01, 500).

For the normal approximation method, we find the MLE and the standard error from the likelihood function

$$\begin{aligned} L(\beta_0,\beta_1,\gamma_0,\gamma_1,\lambda,\sigma|x,y) &= \prod_{i=1}^{n}[f(y_i|x_i,u=1)f(u=1|x_i)+f(y_i|x_i,u=0)f(u=0|x_i)] \\ &\quad \times \prod_{j=1}^{n_1} f(y_{1j}|x_{1j},u_{1j}) \\ &= \prod_{i=1}^{n}\left[\frac{\exp\left\{\frac{-[y_i-(\beta_0+\beta_1x_i+\lambda)]^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}}\right. \\ &\quad \times \frac{\exp\{\gamma_0+\gamma_1x_i\}}{1+\exp\{\gamma_0+\gamma_1x_i\}} \\ &\quad \left. + \frac{\exp\left\{\frac{-[y_i-(\beta_0+\beta_1x_i)]^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}[1+\exp(\gamma_0+\gamma_1x_i)]}\right] \\ &\quad \times \prod_{j=1}^{n_1}\left[\frac{\exp\left\{\frac{-[y_{1j}-(\beta_0+\beta_1x_{1j}+\lambda u_{1j})]^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}}\right. \\ &\quad \left. \times \frac{\exp\{u_{1j}(\gamma_0+\gamma_1x_{1j})\}}{1+\exp\{\gamma_0+\gamma_1x_{1j}\}}\right] \end{aligned}$$

The results we get are listed below in Table 2.3.

Table 2.3: Result of a continuous response data

Method	Estimator of β_1	Standard Error	Posterior Prob
MCMC	30.94636	14.76807	0.98080
Norm.Approx	30.217106	14.380454	0.982191

We can see that both of the methods perform well in estimating the true value and getting a high posterior probability. Their estimators are both very close to 30, the standard errors in both methods are close to each other at around 14, and their posterior probabilities are also close to each other at around 0.98. Hence the results present evidence justifying that the normal theory approximation can also be applied as an alternative method with respect to the Bayesian MCMC approach.

2.4.5 Misclassification Example

Cheng et al. (2009) introduced a Bayesian MCMC approach for regression models with misclassified outcomes. Suppose we have one imperfect test with the simplest model where each subject is tested by one diagnostic instrument that has unknown sensitivity and specificity. Let $D = 1$ denote diseased, and 0 denote healthy, π_j denotes the prevalence for the jth subject, $\boldsymbol{X}_j$ is covariate vector, then we have

$$\pi_j = Pr(D = 1|\boldsymbol{X}_j).$$

π_j depends on $\boldsymbol{X}_j$ through a logit link function,

$$\pi = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_i x_i+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_i x_i+\cdots+\beta_n x_n}}$$

where β_1 is the regression parameter of primary interest with corresponding covariate x_1 and $\beta_i s$ for $i \neq 1$ are the other regression parameters including an intercept term (β_0).

Let $y = 1$ represent a positive test result and 0 denote a negative test result for subject j. There is a single test sensitivity

$$S = Pr(y = 1|D = 1),$$

which is the probability of testing a diseased patient correctly. And a single specificity

$$C = Pr(y = 0|D = 0),$$

which is the probability of testing a healthy patient correctly.

The probability of observing a positive test result, $p_j = Pr(y_j = 1|\boldsymbol{X}_j)$ can be found as

$$p_j = Pr(y_j = 1|\boldsymbol{X}_j) = \pi_j S + (1 - \pi_j)(1 - C).$$

The data are then modelled as independent with a Bernoulli(p_j) distribution. Here we consider a single example with $n = 800$.

We generate the true value S from beta(80,20) distribution with mean 0.8, and C from beta(92,8) distribution with mean 0.92. We assume β_0 is from a normal distribution with mean -1 and standard deviation 0.2 and fix β_1 at 0.8.

We then generate one data set and compare the MCMC and normal approximation methods. We fit the data using the Bayesian model described above with prior distributions of β_0, β_1 to be both normal centered at zero, of a precision 0.1. And we obtain the MLE and the standard error of the MLE, considering them as the center and the standard deviation of the normal theory approximation. The likelihood we use in order to find the MLE and the standard error is:

$$L(\beta_0, \beta_1, s, c|\boldsymbol{x}, \boldsymbol{y}) = \prod [\pi s + (1-\pi)(1-c)]^{y_i} [\pi(1-s) + (1-\pi)c]^{(1-y_i)} s^{0.8}(1-s)^{0.2} c^{0.92}(1-c)^{0.08}$$

where

$$\pi = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}.$$

The $\hat{\beta}_1$ we obtain from the MCMC method is 0.7172, with a standard error at 0.2205. And the $\hat{\beta}_1$ we obtain from the normal approximation method is 0.7183, with a standard error at 0.2229.

We also estimated the posterior probability of $\beta_1 > 0$ given the data in the MCMC method, and correspondingly, we compute a z-score using $\hat{\beta}_1/\hat{\sigma}$ in order to obtain the probability of observing an equivalent or more extreme case. The posterior probability of the MCMC method is 1 and the probability of the normal theory approximation is 0.9994.

From this example we see the normal approximation is very close to the results using MCMC, but further simulation should be run to verify the exact sample sizes where the normal approximation is highly accurate.

2.5 Conclusions and Discussion

When unmeasured confounding exists, to have validation data on hand is a helpful way to compensate because it can provide more information about unmea-

sured confounding. In this chapter, Bayesian regression models are developed to utilize the internal validation data as informative prior distributions for all parameters, first to a binary response and later to a continuous(normal) response. They have retained information on the correlation between the confounder and other covariates. The Bayesian MCMC approach adjusting for unmeasured confounders works well when there are only few covariates. The use of Bayesian modelling provides consistent results, suggesting that the lack of data on the unmeasured confounder does not have a strong impact on the original analysis, due to the relatively weak correlation between the confounder and the outcome variable.

We have also provided a normal theory approximation approach allowing much faster power studies for a Bayesian regression model with unmeasured confounding for both binary and normal outcomes. The results we have are generally similar to the more time consuming Bayesian MCMC approach, which is exactly what we expected. Based upon this information, it is reasonable to conclude that normal approximation is a feasible alternative to the MCMC method in this case.

We do want to mention here that in order to achieve similar power, we need to be careful in choosing appropriate priors for the Bayesian MCMC approach. The priors have to be diffuse and non-informative, and the precision of any parameter should be no larger than 0.1 for the normal response and should be much smaller for other continuously distributed responses.

Sometimes, internal validation may not be available, and the parametric model to adjustment may work well only when there are few covariates and the variables are dichotomous. For multiple unmeasured confounders, model based adjustment is more difficult. Both continuous and categorical variables may be correlated. So under those circumstances, more simulations will be needed to demonstrate performance of external validation and informative prior approach. Bayesian modelling with informative priors may be useful tools in such situations for unmeasured con-

founding sensitivity analyses. However, the need for further research remains in order to understand the operating characteristics of those methods in a variety of situations.

CHAPTER THREE

Bayesian analysis and normal theory approximation under Gamma Distributed Data

3.1 Introduction

The use of retrospective observational studies as tools for medical decision making has been growing in recent years. The use of such data is sometimes challenged due to the potential for unmeasured confounding. Even when no or limited additional data on the unmeasured confounders isavailable, Bayesian modeling with informative priors can still be utilized to assess the sensitivity to the unmeasured confounding.

In the last chapter, we discussed Bayesian MCMC and normal approximation analyses for binomial and normal outcomes when unmeasured confounding exists. To further explore the case with a continuous outcome, in this chapter we study an outcome with a gamma distribution, which is often seen in cost data. Cost data is well known to often be skewed. Several classes of models can be used to address the problems caused by skewness in data commonly encountered in health care applications. Manning et al. (2005) has presented a method using the three parameter generalized Gamma (GGM) distribution, which includes several of the standard alternatives as special cases, for instance OLS with a normal error, OLS for the log-normal, the standard Gamma and exponential with a log link.

Our work focuses on situations in which internal validation data are available. Additional data on the "unmeasured confounder" are obtained for a small subset of the original sample. The advantage of the Bayesian approach for this type of problem is the ability to combine validation data with informative priors in a straightforward operational way to perform estimation.(see Faries et al., 2013)

Bayesian methods provide a flexible approach to studying unmeasured confounding. It can utilize all sources of information while simultaneously modeling all uncertainty. In this work, we are interested in what advantages a normal theory approximation brings when the data are skewed. In this chapter, we present a Bayesian MCMC method with the skewed cost outcome and compare the results with another approach using the normal theory approximation, and see whether the normal approximation approach is reasonable.

3.1.1 Gamma Distributed Data

Health care expenditure data are usually right skewed with variability increasing as the mean cost increases. Many past studies of health care costs and their responses to health insurance, treatment modalities or patient characteristics indicate that estimation of mean responses may be quite sensitive to how estimators account for the skewness in the outcome and other statistical problems that are common in such data. It has been suggested that cost data can be modeled with the log normal or gamma distributions. The generalized gamma distribution has one scale parameter and two shape parameters. This form is also referred to as the family of generalized gamma distributions because the standard gamma, Weibull, exponential and the log normal are all special cases of this distribution. Hence, it provides a convenient form to identify the data generating mechanism of the dependent variable and in turn helps to select the best distribution. (see Manning et al., 2002)

The probability density of the generalized gamma distribution ($\mathrm{GG}(\alpha, \tau, \lambda)$) is given by Morteza Khodabina (2010),

$$f(y|\alpha, \tau, \lambda) = \frac{\tau}{\lambda\Gamma(\alpha)}(\frac{y}{\lambda})^{\alpha\tau-1}e^{-(\frac{y}{\lambda})^{\tau}}, \;\; y \geq 0, \tau, \alpha, \lambda > 0,$$

where $\Gamma(.)$ is the gamma function, α and τ are shape parameters, and λ is the scale parameter. $E(y|x)$ is often the primary quantity of interest in many health

economics applications. For example, predicted costs by treatment, predicted health care utilization by patient characteristics, etc.

The GG family is flexible in that it includes several special cases: exponential($\alpha = \tau = 1$), standard gamma for($\tau = 1$), and Weibull for($\alpha = 1$). The log normal distribution is also obtained as a limiting distribution when α goes to infinity. In this chapter, we will focus on the standard gamma model. Like the log normal, the gamma distribution has a variance function that is proportional to the square of the mean function, a property that characterizes many health care data sets.

The probability density function (PDF) of the standard gamma distribution is:

$$f(y|\alpha,\mu) = \frac{1}{\Gamma(\alpha)}(\frac{\alpha}{\mu})^{\alpha} y^{\alpha-1} \exp(-\frac{\alpha y}{\mu}), \;\; y \geq 0, \alpha, \mu > 0$$

In this parameterization, α is the shape parameter and $\mu = E[Y]$ is the mean. If $0 < \alpha < 1$, then the density has a pole at the origin and decreases monotonically as y increases. If $\alpha = 1$, then this is an exponential distribution. If $\alpha > 1$, then the density is zero at the origin, with a maximum at $\mu - \mu/\alpha$. There are some other common parameterizations of the gamma distribution. We can define it in terms of shape and scale, where scale = μ/α. We can also define it in terms of shape and rate, where rate = α/μ. The MLE of μ is the sample mean, but to get the MLE of α requires an iterative approximation. Since the gamma distribution is a generalized linear model, α and μ can be estimated using standard statistical packages by setting the systematic component of the model to be an intercept term alone. The MLEs of the intercept and the dispersion parameters are the MLEs of μ and $1/\alpha$, respectively.

3.2 Bayesian Regression and Normal Approximation Model with Unmeasured Confounding

3.2.1 Model Description

Let X be a binary random variable where values 1 and 0 indicate whether or not a certain medical technology is used, and let U be a binary variable that is an unmeasured confounder. Following McCandless et al. (2007b), we use the factorization $P(Y, U|X) = P(Y|X, U)P(U|X)$ and model the confounding effect of U using the regression models:

$$\log P(Y|X, U) = \beta_0 + \beta_1 X + \lambda U \tag{3.1}$$

$$\text{logit}\, P(U = 1|x) = \gamma_0 + \gamma_1 X \tag{3.2}$$

Here, the effects of the variables X and U on Y are assumed to not interact. Out of n observations, we assume that we have n_1 validation sample points, as in Chapter 2.

The outcome variable Y is assumed to have a gamma distribution

$$Y \sim \text{Gamma}(\alpha, s), \text{ where } s = \alpha/e^{\beta_0+\beta_1 X+\lambda U}$$

and all the priors' information is listed below:

$$\beta_0, \beta_1 \sim N(\mu_\beta, \sigma_\beta^2)$$

$$\gamma_0, \gamma_1 \sim N(\mu_\gamma, \sigma_\gamma^2)$$

$$\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$$

$$\alpha \sim \text{Unif}(lo, up)$$

We have $\mu_\beta = \mu_\gamma = \mu_\lambda = 0, \sigma_\beta^2 = 10000, \sigma_\gamma^2 = \sigma_\lambda^2 = 100, lo = 0.001, up = 50$. We use diffuse normal distributions as the independent priors on the βs, γs and λ, and a uniform distribution as the prior on the α in our simulations, although we give diffuse priors to all the parameters in our simulations, expert opinion and/or prior data can be incorporated as well. We are still primarily interested in β_1 as it

is the coefficient for x, which represents the impact of the treatment on the health care expenditure.

To perform a normal theory approximation to this gamma regression with unmeasured confounding estimation problem, we first acquire the full likelihood function:

$$\begin{aligned} L_N(\beta_0,\beta_1,\gamma_0,\gamma_1,\lambda|x,y) =& \prod_{i=1}^{n}[f(y_i|x_i,u=1)f(u=1|x_i) \\ & + f(y_i|x_i,u=0)f(u=0|x_i)] \\ =& \prod_{i=1}^{n}\left[\frac{y_i^{\alpha-1}e^{\frac{-y_i\alpha}{\exp(\beta_0+\beta_1x_i+\lambda)}}}{(\frac{\exp(\beta_0+\beta_1x_i+\lambda)}{\alpha})^{\alpha}\Gamma(\alpha)}\right. \\ & \times\frac{\exp\{\gamma_0+\gamma_1x_i\}}{1+\exp\{\gamma_0+\gamma_1x_i\}} \\ & + \frac{y_i^{\alpha-1}e^{\frac{-y_i\alpha}{\exp(\beta_0+\beta_1x_i)}}}{(\frac{\exp(\beta_0+\beta_1x_i)}{\alpha})^{\alpha}\Gamma(\alpha)} \\ & \left.\times\frac{1}{1+\exp\{\gamma_0+\gamma_1x_i\}}\right] \end{aligned}$$

$$\begin{aligned} L_{N_1}(\beta_0,\beta_1,\gamma_0,\gamma_1,\lambda|x_1,y_1,u_1) =& \prod_{j=1}^{n_1} f(y_{1j}|x_{1j},u_{1j}) \\ =& \prod_{j=1}^{n_1}\left[\frac{y_{1j}^{\alpha-1}e^{\frac{-y_{1j}\alpha}{\exp(\beta_0+\beta_1x_{1j}+\lambda u_{1j})}}}{(\frac{\exp(\beta_0+\beta_1x_{1j}+\lambda u_{1j})}{\alpha})^{\alpha}\Gamma(\alpha)}\right. \\ & \left.\times\frac{\exp\{u_{1j}(\gamma_0+\gamma_1x_{1j})\}}{1+\exp\{\gamma_0+\gamma_1x_{1j}\}}\right] \end{aligned}$$

Here $\log L$ is the total log likelihood by combining the log of the two parts together,

$$\log L \;=\; \log L_N + \log L_{N_1}.$$

The first and second derivative used to perform the maximation and the estimation of the variances are provided in the appendix.

3.2.2 Simulation Algorithm and Conditions

We now discuss the simulation procedure used to investigate the normal approximation.

First, we generate the covariate x_1 from a binomial distribution with $p_x = 0.6$. Then using fixed value of γ_0 and γ_1, we generate the covariates for the confounder u_1 from a binomial distribution according to Equation (3.2). Next we generate the outcome y_1 from a gamma distribution with log mean $\beta_0 + \beta_1 X + \lambda u_1$ using fixed $\beta_0, \beta_1, \lambda$ and shape parameter α by Equation (3.1).

Next, we generate the $n - n_1$ values for the main study without validation in a similar way. Namely, x is generated from binomial(0.6), and we use same fixed γ_0, γ_1 and Equation (3.2) to obtain u, and the same fixed $\beta_0, \beta_1, \lambda$ together with u and α to obtain y from Equation (3.1). Notice we do not actually observe u for the main study data.

Then, we fit the Bayesian models (3.1) and (3.2) to the generated data above. We use diffuse normal priors, with mean of zero and precision of 0.01 for β_0 and β_1 and a precision of 0.1 for γ_0, γ_1 and λ. We use a uniform(0.001, 50) prior on α.

Finally, we approximate the posterior distribution of β_1 using WinBUGS while keeping track of the posterior probability values of $P(\beta_1 > 0|\text{data})$ for each iteration. For the normal approximation, the test statistic $z = \frac{\hat{\beta}_1}{\text{SE}_{(\hat{\beta}_1)}}$ is used to determine inference on the relationship.

We then repeat this process for m iterations, and tally the number of times that the posterior probability value exceeds $1 - \alpha$ out of m iterations. This value approximates the Bayesian power achieved for a sample size n.

Table 3.1: MCMC vs. Normal Approximation with Gamma distribution under unmeasured confounding

n	Norm's β_1	Norm's SD	MCMC's β_1	MCMC's SD
500	0.7960963	0.1607586	0.7941613	0.14975661
600	0.7893362	0.1489264	0.7995130	0.1370918
700	0.8017156	0.1404461	0.7902297	0.1265650
800	0.7933256	0.1332436	0.7932495	0.1185082
900	0.8011129	0.1271441	0.8092028	0.1116286
1000	0.7964141	0.1223493	0.8011267	0.1057505

We generated $m = 500$ data sets with discrete posterior approximations based on a MC sample of 5000 iterates after a 1000 burn-in and thinning of 2.

3.2.3 Results

We now describe the results of one simulation experiment. Under Equation (3.1) and (3.2), our outcome $\boldsymbol{Y}$ takes on a gamma distribution. We fix β_1 at 0.8. The other parameters are fixed as follows: $\beta_0 = -0.5$, $\alpha = 0.4$, $\gamma_0 = -0.2$, $\gamma_1 = 1$ and $\lambda = -0.6$. We also fixed n_1, the validation sample size, to 200, and let the total sample size to go from 500 to 1000. For $m = 500$ different data sets, the powers of both methods all turn out to be very close to 1.

As we can see from the results, both Bayesian MCMC and the normal theory approximation methods are doing quite well in estimating the coefficient of x_1. Both of the methods have estimated the true value (0.8) within a bias of only 0.01. The MCMC method has smaller standard errors as compared to the normal theory approximation for all sample sizes.

Since the differences in the standard errors is quite large, we investigated whether the increasing the total sample sizes would reduce the difference. For the true β_1 still fixed at 0.8, and total sample sizes being from 1500 to 2000, the standard errors from the normal approximation method are from 0.094 to 0.084, and the standard errors from the Bayesian MCMC method are from 0.087 to 0.075. The

differences in the standard errors between the two methods are still the same when we increase the sample size. We also test some other true values of β_1, and the standard errors of the two methods do not fluctuate much at all. All of the results above indicate that the differences in the standard errors are large enough for us to conclude that there is a significant difference between the two methods, hence we do not recommend using the normal theory approximation as an alternative quicker way to the Bayesian MCMC approach in this case. We believe that the reason why we cannot use normal approximation is due to the apparent skewness of the sample data.

3.3 Conclusions and Discussion

In this chapter, we have considered the common issue of modeling skewed cost data in observational health care studies. This type of data often suffers from the problem of unmeasured confounding. Here, we have assumed internal validation data are available where the confounder was not completely unmeasured. Like in (see Faries et al., 2013), additional data on this potential unmeasured confounder are obtained for a small subset of the original sample.

After running simulations assuming a gamma response with one binary covariate and one binary unmeasured confounder using both MCMC and normal approximations to obtain estimators, we obtained some interesting results. Not surprisingly, the averages of the posterior means for both methods were nearly unbiased. The posterior variability however is significantly smaller for the MCMC approach than the normal approximation. This is a surprising result given that generally the normal approximation would be expected to underestimate the uncertainty for smaller samples sizes.

Cost data usually have very skewed distributions and can be difficult to model. According to (see Thompson and Nixon, 2005), conclusions from cost-effectiveness

analyses are sensitive to choice of distribution and, in particular, to how the upper tail of the cost distribution beyond the observed data is modeled. How well a distribution fits the data is an insufficient guide to model choice. In the future, we would like to investigate whether the choice of distribution can make a difference to the power of the method of choosing, and to explore the importance in selecting the correct model to fit the response data.

CHAPTER FOUR

Bayesian Cost-Effectiveness Analyses and their Normal Theory Approximation

The field of health economics is growing rapidly, and there is an increasing interest from health providers of many countries in assessing the evidence of economic value along with clinical efficacy for new drugs and treatments. Suppose we need to compare two therapies aimed at the same medical condition and try to determine which one of these can be judged as "better". Physicians frequently need to base their daily treatment decisions on this type of comparative effectiveness research. But instead of referring only to "effectiveness research", which is in the clinical realm, we also consider "comparative cost-effectiveness analysis", which also accounts for differential cost between treatments. This cost-effectiveness analysis seeks to establish which of several alternative strategies capable of achieving a given therapeutic goal is the best. The literature has revealed a lot of variation in the methodology and the reporting of these analyses. Improving the quality of these studies is very important to these decision makers.

4.1 Introduction

Cost-effectiveness analyses of clinical trial data are based on assumptions about the distributions of costs and effectiveness.

An important problem is the comparison of two treatments using data from a clinical trial. Both cost and effectiveness are measured on each patient in each of the two treatment groups. Recent research about this problem has been brought up by several authors. For instance, Willan and O'Brien (1996) presented a procedure for the statistical analysis of cost-effectiveness data, with specific application to those studies for which effectiveness is measured as a binary outcome, using Fieller's The-

orem to calculate confidence intervals for the incremental cost-effectiveness ratio. Normality of the underlying cost and effectiveness data was a common assumption among many researchers even though the cost data are typically non-normally distributed, with a high skewness. (see Willan and O'Brien, 1996; Laska et al., 1997; Zethraeus and Johannesson, 1998; Stinnett and Mulahy, 1998.; Heitjan et al., 1999; Briggs and Fenn, 1998; O'Hagan et al., 2000)

However, the volume of literature already dealing with the normal case testified to how common this assumption was made. Most previous work adopts essentially a frequentist approach, although the Bayesian approach has been applied. Briggs (1999) provides an outline, without technical details or references to any particular data, of how the Bayesian approach would work, while O'Hagan et al. (2000) presented a simplified Bayesian analysis with non-informative prior information in a Bayesian cost-effectiveness analysis from clinical trial data, and O'Hagan et al. (2001) is the first to present explicit analysis making use of substantial prior information, as well as demonstrating the value of such information in a practical case study.

A move toward economic evaluation studies being conducted alongside clinical trials requires that individual patient data be available. We also face an issue of analyzing uncertainty due to sampling variability in that case. In the health economics literature, the incremental net benefit (INB) or incremental cost-effectiveness ratio (ICER) have been the main focus of interest and various methods for computing confidence intervals around the INB and the ICER have been proposed and discussed extensively.

4.2 Cost and Effectiveness Distribution

4.2.1 Distributional Assumptions

The cost of medical resources is often recorded for each patient in clinical studies in order to inform decision making. Although cost data are generally skewed to the right, our interest is still in making inferences about the population mean cost. The Central Limit Theorem ensures the sample mean is a consistent estimator. We choose an alternative estimator only when there are sufficient data to permit detailed modeling, otherwise the sample mean remains the best estimator.

Here we assume our cost data are distributed as a gamma distribution. Then the maximum likelihood estimator of the population mean is the sample mean. Notice here all parametric assumptions are just approximations, and incorrect assumptions may lead to misleading conclusions. Hence we will gain efficiency if an appropriate distribution such as the gamma is chosen to fit the data.

Thompson and Nixon (2005) used data from a low back pain trial to analyze the cost-effectiveness. Patients were recruited from Washington State, USA. In this study, 190 patients were randomized to an investigation by rapid magnetic resonance imaging (rMRI), and the other 190 to a standard X-ray investigation. The issue being addressed by the trial was whether rMRI would allow better diagnosis and treatment, or simply lead to unnecessary treatment without improvement in symptoms.

We set up our model following the results of Thompson and Nixon (2005). We want to know if a normal approximation for cost-effectiveness data is reasonable when the costs are skewed. Hence, to represent the usual skewness in cost data, we use a gamma distribution for costs, and a normal distribution for effectiveness. Specifically, we assume:

$$E_i \sim Normal(\mu_{E_i}, \sigma_E)$$

$$
\begin{aligned}
C_i &\sim Gamma(\mu_C, \rho_C) \\
\mu_{E_i} &= \mu_E + \beta(C_i - \mu_C)
\end{aligned}
$$

where E_i represents effectiveness and C_i represents costs. The gamma distribution for the costs above is parametrized by its mean and shape. The parameter β accounts for the correlation between costs and effectiveness.

4.2.2 Incremental Net Benefit

After looking at the estimation of β using both a normal approximation and Bayesian MCMC methods, we are able to determine whether using the normal approximation is reasonable. Generally we are interested in is the Incremental Net Benefit(INB).

We denote by C_{ij} a random variable that is the total cost for individual $j = 1, \cdots, n_i$ who is given treatment i, where i is either treatment(t) or standard(s). The number of individuals given treatment i is denoted as n_i. We assume that the costs for those individuals given treatment i are from the distribution with mean μ_{ci} and variance σ^2_{ci}, and the costs from the two groups are independent. E_{ij} denotes the random variable for health outcome for individual j given treatment i, and it has mean μ_{ei} and variance σ^2_{ei}. The population mean differences in costs and in effects are denoted by $\Delta_C = \mu_{ct} - \mu_{cs}$ and $\Delta_E = \mu_{et} - \mu_{es}$ respectively.

We use the incremental net benefit (INB) of the treatment comparing to the standard to summarize the results. Let λ represents the decision makers' willingness to pay for one unit gain in health outcome. Methods for estimating willingness to pay are discussed in O'Brien (1998); O'Brien and Gafni (1996). The INB is defined to be

$$
\begin{aligned}
INB(\lambda) &= \lambda\Delta_E - \Delta_C && (4.1) \\
&= \lambda(\mu_{et} - \mu_{es}) - (\mu_{ct} - \mu_{cs}) && (4.2)
\end{aligned}
$$

This quantity is the net benefit expressed in dollars, of giving a patient the treatment rather than the standard, with positive difference favoring the treatment and a negative difference favoring the standard.

Note that

$$\begin{aligned} INB(\lambda) &= \lambda(\mu_{et} - \mu_{es}) - (\mu_{ct} - \mu_{cs}) \\ &= (\lambda\mu_{et} - \mu_{ct}) - (\lambda\mu_{es} - \mu_{cs}) \\ &= NB_t - NB_s \end{aligned}$$

The value of λ is generally unknown, so it is common to plot the estimated value of INB(λ) for various values of λ. Also note that $INB(0) = -\Delta_C$ illustrates that the cost minimization is a special case of incremental net benefit.

A treatment is considered cost-effective if, and only if, the INB is greater than zero, namely $INB(\lambda) = \lambda\Delta_E - \Delta_C > 0$. See for example Figure 4.1.

The treatment is cost-effective when $\lambda\Delta_E > \Delta_C$, which is equivalent to

$$\lambda > \frac{\Delta_C}{\Delta_E} \equiv R(\Delta_E > 0) \text{ or } \lambda < \frac{\Delta_C}{\Delta_E} \equiv R(\Delta_E < 0),$$

R is referred to as the incremental cost-effectiveness ratio (ICER), and can be seen as additional cost to realize an extra unit of effectiveness from using the treatment rather than the standard.

Observing that $INB(R) = 0$ which demonstrates the connection between the ICER and the INB. Therefore, in a cost-effectiveness analysis, one need only to estimate $INB(\lambda)$ and its confidence limits, and to graph them as a function of λ. These curves cross the vertical axis at minus the cost difference and the horizontal axis at the ICER, defining the respective estimates and the corresponding confidence intervals.

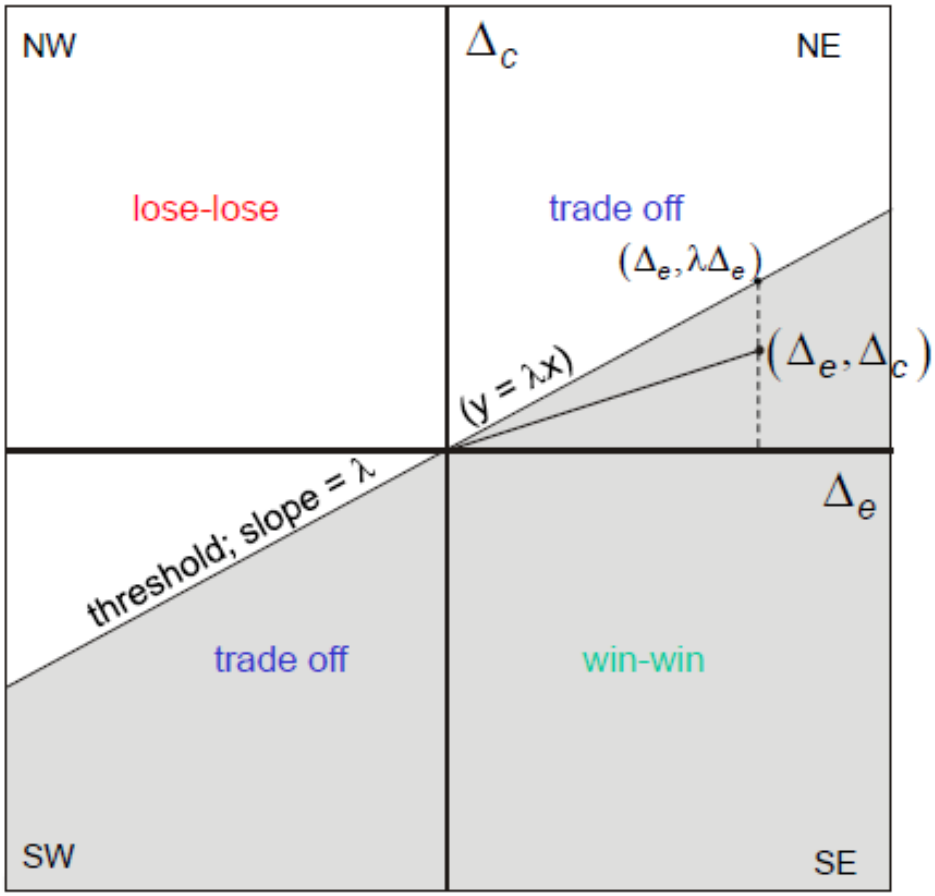

Figure 4.1: the Cost-Effective Plane

4.2.3 Estimation

The expected value and variance of $INB(\lambda)$ are given by

$$\begin{aligned} E[\hat{INB}(\lambda)] &= \lambda E(\hat{\Delta}_E) - E(\hat{\Delta}_C) \\ &= \lambda E(\hat{\mu}_{et} - \hat{\mu}_{es}) - E(\hat{\mu}_{ct} - \hat{\mu}_{cs}) \\ &= \lambda(\mu_{et} - \mu_{es}) - (\mu_{ct} - \mu_{cs}) \end{aligned}$$

$$\begin{aligned} Var[\hat{INB}(\lambda)] &= \lambda^2 Var(\hat{\Delta}_E) + Var(\hat{\Delta}_C) - 2\lambda Cov[\hat{\Delta}_E, \hat{\Delta}_C] \\ &= \lambda^2 \sigma^2_{\Delta_E} + \sigma^2_{\Delta_C} - 2\lambda\sigma_{\Delta EC}, \end{aligned}$$

where

$$\begin{aligned} \sigma^2_{\Delta_C} &= \frac{\sigma^2_{ct}}{n_t} + \frac{\sigma^2_{cs}}{n_s} \\ \sigma^2_{\Delta_E} &= \frac{\sigma^2_{et}}{n_t} + \frac{\sigma^2_{es}}{n_s} \end{aligned}$$

Here $\sigma^2_{\Delta_C}$ and $\sigma^2_{\Delta_E}$ are the variances of the estimated population mean cost differences and effectiveness differences respectively. The quantities σ^2_{ci} and $\sigma^2_{ei}, i = t, s$ are the variances of the distributions from which the cost and effectiveness data are sampled.

4.2.4 Advantages of INB over ICER

There are many advantages to using INB over the ICER. The main advantage we will gain is that the INB analysis provides cost minimization while the ICER analysis is a special case of it. Also, the INB considers the value that is given to a unit of effectiveness as well as the cost: the INB is the difference between value and the cost, whereas ICER is the cost of an extra unit of effectiveness.(see Willan and Lin, 2001)

Negative ICERs are difficult to interpret and are not properly ordered. For example, in some special cases, the upper limit of the ICER can be less than the lower limit. This appears misleading and may causes investigators to mistakenly reverse the limits and reach incorrect conclusions. Also, two totally opposite results could even have the same ICER. The situation will be more clear when using the INB.

The lower limit of INB is always negative, implying that no value ascribed to a unit of effectiveness would lead to rejection of the null hypothesis that $INB = 0$ in favor of $INB > 0$. This means that no matter how much one values say, a year of life, there is no evidence that the treatment is cost-effective compared to the standard.

Although the fact that ICER's limits include the positive vertical axis amounts to the same conclusion, it is less obvious to realize.

The confidence intervals for the ICER sometimes include undefined values. For the INB analysis, this situation is characterized by neither INB limit crossing the horizontal axis. Thus with zero being in the confidence interval, no matter how much or how little one values a unit of effectiveness, there is no evidence that either the standard or the treatment is more favored than the other.

There are other advantages as well. INB has the ability to generalize to more than one measures of effectiveness, for instance, in a trial of antithrombotic medication, one might be interested in deaths, strokes and blood clots at the same time.

4.2.5 Bayesian Approach

In this section, we discuss the Bayesian approach to estimation of cost-effectiveness, implemented using Markov chain Monte Carlo (MCMC) in the software R calling the package `R2WinBUGS`. The Bayes approach gives a natural interpretation of cost-effectiveness providing the posterior probability of the parameter of interest not being zero, given the data.

Specifically, we assume cost and effectiveness are both continuous where effectiveness has a normal distribution and the cost has a gamma distribution. We denote the effectiveness variable to be Y and the cost variable to be X. Then Y is a normal response with mean μ_Y and the standard deviation σ_Y. X is a gamma distributed variable with mean μ_X, and shape parameter α. We assume a linear function between the mean of effectiveness and the residual of the cost. Also, note that we have $i = 2$ groups, the treatment group and the standard group, respectively. This can be summarized as:

$$Y_i \sim Normal(\mu_{Y_i}, \sigma_{Yi}) \tag{4.3}$$

$$X_i \sim Gamma(\alpha_i, \mu_{Xi}) \tag{4.4}$$

$$\mu_{Y_i} = \beta_{0i} + \beta_1(X_i - \mu_{Xi}) \tag{4.5}$$

where in the gamma distribution, the shape parameter is α, and the rate parameter satisfies $rate = \alpha/\mu_X$. We denote baseline effectiveness μ_Y to be β_0. Because the Bayesian MCMC method requires prior distributions for all the parameters in the model, we give diffuse priors, which are intended to be approximately noninformative, to those parameters so that our inferences essentially only depend on the data. We use wide uniform(0.01,500) priors for μ_X and α in the gamma distribution and the same prior for σ_Y in the normal distribution because those parameters are bounded to be positive. Since β_{0i} is the baseline effectiveness, and β_1 is our coefficient of interest in the analysis, and they could be either positive or negative, we use diffuse normal distributions center at 0 with precision 0.00001 as the prior for these parameters.

Posterior distributions of quantities of interest for estimation and the inferences about cost-effectiveness were derived from $m = 10000$ MCMC iterations, 1000 initial burn-in iterations were discarded to ensure convergence. These posterior distributions are summarized by means, standard deviations and powers.

Our primary interest is the incremental net benefit, $INB(\lambda) = \lambda\Delta Y - \Delta X$, and specifically, the $P(\text{INB}(\lambda) > 0|\text{data})$.

4.2.6 Normal Theory Approximation

We now present the details of a normal theory approximation. The likelihood functions of the parameters are:

$$\begin{aligned} L_{N1}(\beta_0, \beta_1, \beta_2, \alpha_1, \sigma_{Y_1}|x_1, y_1) &= \prod [f(\boldsymbol{y}_1|\boldsymbol{x}_1) f(\boldsymbol{x}_1)] \\ &= \prod_{i=1}^{n} \left[\frac{\exp\left\{ \frac{-\{y_{1i} - [\beta_{01} + \beta_1(x_{1i} - \mu_{X1})]\}^2}{2\sigma_{Y1}^2} \right\}}{\sqrt{2\pi\sigma_{Y1}^2}} \frac{x_{1i}^{\alpha_1 - 1} e^{\frac{-x_{1i}\alpha_1}{\mu_{X1}}}}{(\frac{\mu_{X1}}{\alpha_1})^{\alpha_1}\Gamma(\alpha_1)} \right] \end{aligned}$$

$$L_{N2}(\beta_0, \beta_1, \beta_2, \alpha_2, \sigma_{Y_2}|x_2, y_2) = \prod [f(\boldsymbol{y}_2|\boldsymbol{x}_2) f(\boldsymbol{x}_2)]$$

$$= \prod_{i=1}^{n} \left[\frac{\exp\left\{ \frac{-\{y_{2i}-[\beta_{02}+\beta_1(x_{2i}-\mu_{X2})]\}^2}{2\sigma_{Y2}^2} \right\}}{\sqrt{2\pi\sigma_{Y2}^2}} \frac{x_{2i}^{\alpha_2-1} e^{\frac{-x_{2i}\alpha_2}{\mu_{X2}}}}{(\frac{\mu_{X2}}{\alpha_2})^{\alpha_2}\Gamma(\alpha_2)} \right]$$

We take a log transformation of the likelihood functions to get $\log L_{N1}$ and $\log L_{N2}$, and the total log likelihood function will be the sum of these,

$$\log L = \log L_{N1} + \log L_{N2}.$$

Like before, the first and second derivatives used to find the MLEs and the Hessian matrix are provided in the appendix.

4.3 Simulation Algorithm

4.3.1 WinBUGS Simulation and Conditions

We now describe the simulation algorithm. First, we generate the variable x_i from a gamma distribution with mean μ_{Xi} and shape parameter α_i for some fixed value of μ_{Xi}s and α_is where i is the standard or the treatment using Equation (4.4).

Secondly, we generate the variable y_i from a normal distribution with mean $\beta_{0i} + \beta_1(x_i - \mu_{Xi})$ and standard deviation σ_{Yi} using the $\boldsymbol{X}$ value obtained from the gamma distribution and fixed value of β_{0i}s where $i = 1, 2$ by Equation (4.3). Hence, the true INB value of this set of data is INB $= \lambda(\beta_{01} - \beta_{02}) - (\mu_{X1} - \mu_{X2})$, with a fixed value at λ.

Then, we use WinBUGS to fit the Bayesian model (4.5) to the generated data above using Markov chain Monte Carlo method. Here we also calculate a function of INB: $\lambda(y_1 - y_2) - (x_1 - x_2)$, with the same value of λ.

Finally, we approximate the posterior distribution of β_{0i}, β_1, α_i, μ_{Xi}, σ_{Yi} and INB for i = 1, 2 using MCMC method in WinBUGS while keeping track of the power by recording the posterior probability values in each iteration.

Table 4.1: Results of NA and MCMC methods for n = 100, $\text{MCMC}_{\text{power}} = 0.19$, $\text{NA}_{\text{power}} = 0.18$

	MCMC's Est	MCMC's SD	NA's Est	NA's SD
$\alpha_1(\mathbf{2})$	2.088	0.274	2.068	0.272
$\alpha_2(\mathbf{4})$	4.156	0.562	4.116	0.561
$\beta_1(\mathbf{1.5})$	1.477	0.087	1.477	0.085
$\mu_{X1}(\mathbf{8})$	8.094	0.573	8.017	0.352
$\mu_{X2}(\mathbf{4})$	4.027	0.201	4.007	0.560
$\sigma_{Y1}(\mathbf{5})$	5.085	0.369	4.972	0.199
$\sigma_{Y2}(\mathbf{5})$	5.064	0.367	4.116	0.352
$\beta_{01}(\mathbf{4})$	4.171	0.992	4.059	0.968
$\beta_{02}(\mathbf{2})$	2.051	0.590	2.023	0.578
INB(**2**)	2.292	3.016	2.098	2.944

We repeat the whole process for $m = 100$ iterations, and tally the number of times that posterior probability value exceeds 0.95 out of m iterations. This value approximates the Bayesian power achieved for sample size n.

We do the simulation for a Monte Carlo sample of 5000 posterior iterates after a 1000 initial burn-in with a thinning equals to 2. All simulations have converged to the supporting distribution by this point, and mixed well. The fitting was achieved by WinBUGS through package `R2WinBUGS` in `R` to perform Bayesian estimation. The necessary WinBUGS code is given in the appendix.

4.4 Results

For the simulation, we set $\beta_{01} = 4$, $\beta_{02} = 2$, and β_1 is fixed at 1.5. The other parameters are fixed as follows: $\alpha_1 = 2$, $\alpha_2 = 4$, $\mu_{X1} = 8$, $\mu_{X2} = 4$, $\sigma_{Y1} = \sigma_{Y2} = 5$ and $\lambda = 3$. We let the total sample size n range from 100 to 700.

The results are displayed in Tables 4.1-4.7. The parameters of interest are provided in the first column, with the true values in bold located between the parentheses.

Table 4.2: Results of NA and MCMC methods for n = 200, $\text{MCMC}_{\text{power}} = 0.26$, $\text{NA}_{\text{power}} = 0.23$

	MCMC's Est	MCMC's SD	NA's Est	NA's SD
$\alpha_1(\mathbf{2})$	2.037	0.189	2.028	0.189
$\alpha_2(\mathbf{4})$	4.087	0.392	4.068	0.400
$\beta_1(\mathbf{1.5})$	1.501	0.060	1.502	0.060
$\mu_{X1}(\mathbf{8})$	8.081	0.404	8.041	0.250
$\mu_{X2}(\mathbf{4})$	4.022	0.142	4.012	0.391
$\sigma_{Y1}(\mathbf{5})$	5.049	0.256	4.993	0.141
$\sigma_{Y2}(\mathbf{5})$	5.036	0.255	4.068	0.250
$\beta_{01}(\mathbf{4})$	4.123	0.704	4.062	0.698
$\beta_{02}(\mathbf{2})$	2.048	0.416	2.035	0.412
INB($\mathbf{2}$)	2.166	2.132	2.052	2.113

Table 4.3: Results of NA and MCMC methods for n = 300, $\text{MCMC}_{\text{power}} = 0.36$, $\text{NA}_{\text{power}} = 0.35$

	MCMC's Est	MCMC's SD	NA's Est	NA's SD
$\alpha_1(\mathbf{2})$	2.053	0.156	2.046	0.155
$\alpha_2(\mathbf{4})$	4.035	0.316	4.022	0.325
$\beta_1(\mathbf{1.5})$	1.498	0.049	1.498	0.048
$\mu_{X1}(\mathbf{8})$	8.067	0.327	8.041	0.203
$\mu_{X2}(\mathbf{4})$	4.007	0.116	4.000	0.316
$\sigma_{Y1}(\mathbf{5})$	5.014	0.207	4.978	0.115
$\sigma_{Y2}(\mathbf{5})$	5.016	0.207	4.022	0.204
$\beta_{01}(\mathbf{4})$	4.125	0.571	4.085	0.566
$\beta_{02}(\mathbf{2})$	2.046	0.338	2.036	0.336
INB($\mathbf{2}$)	2.177	1.731	2.107	1.718

The first two columns are the results from the MCMC method. We have the estimator and standard deviation of the parameters. The corresponding power in each case is listed in the table title. The last two columns are results from the normal theory approximation. To illustrate our findings better, we also plot those results of both methods by each parameters, and put them side by side to simplifies the comparison.

Table 4.4: Results of NA and MCMC methods for n = 400, $\text{MCMC}_{\text{power}} = 0.46$, $\text{NA}_{\text{power}} = 0.44$

	MCMC's Est	MCMC's SD	NA's Est	NA's SD
$\alpha_1(\mathbf{2})$	2.030	0.133	2.025	0.133
$\alpha_2(\mathbf{4})$	4.023	0.273	4.014	0.283
$\beta_1(\mathbf{1.5})$	1.503	0.042	1.503	0.041
$\mu_{X1}(\mathbf{8})$	8.069	0.285	8.049	0.175
$\mu_{X2}(\mathbf{4})$	4.015	0.101	4.010	0.273
$\sigma_{Y1}(\mathbf{5})$	4.978	0.177	4.951	0.100
$\sigma_{Y2}(\mathbf{5})$	5.011	0.179	4.014	0.176
$\beta_{01}(\mathbf{4})$	4.130	0.495	4.099	0.493
$\beta_{02}(\mathbf{2})$	2.037	0.293	2.031	0.292
INB($\mathbf{2}$)	2.226	1.499	2.168	1.493

Table 4.5: Results of NA and MCMC methods for n = 500, $\text{MCMC}_{\text{power}} = 0.44$, $\text{NA}_{\text{power}} = 0.42$

	MCMC's Est	MCMC's SD	NA's Est	NA's SD
$\alpha_1(\mathbf{2})$	2.010	0.118	2.006	0.118
$\alpha_2(\mathbf{4})$	4.024	0.244	4.015	0.253
$\beta_1(\mathbf{1.5})$	1.502	0.038	1.502	0.037
$\mu_{X1}(\mathbf{8})$	8.028	0.254	8.011	0.158
$\mu_{X2}(\mathbf{4})$	4.001	0.089	3.997	0.244
$\sigma_{Y1}(\mathbf{5})$	5.026	0.160	5.004	0.089
$\sigma_{Y2}(\mathbf{5})$	4.975	0.158	4.015	0.157
$\beta_{01}(\mathbf{4})$	4.044	0.443	4.019	0.441
$\beta_{02}(\mathbf{2})$	2.041	0.260	2.035	0.259
INB($\mathbf{2}$)	1.984	1.340	1.938	1.334

Considering the posterior means in columns two and four, we see both methods exhibit very little bias. The posterior standard deviations in columns three and five agree quite nicely for all the parameters.

Now if we focus on the standard deviation of each parameter of interest, and just look at one parameter separately across different sample sizes, we find that the standard deviations of all parameters tends to decrease when the sample size increases. And also the results from the normal theory approximations are generally similar to the results from the Bayesian MCMC method.

Table 4.6: Results of NA and MCMC methods for n = 600, $\text{MCMC}_{\text{power}} = 0.52$, $\text{NA}_{\text{power}} = 0.52$

	MCMC's Est	MCMC's SD	NA's Est	NA's SD
$\alpha_1(\mathbf{2})$	2.009	0.108	2.006	0.108
$\alpha_2(\mathbf{4})$	4.004	0.222	3.997	0.231
$\beta_1(\mathbf{1.5})$	1.503	0.034	1.503	0.034
$\mu_{X1}(\mathbf{8})$	8.021	0.232	8.008	0.144
$\mu_{X2}(\mathbf{4})$	3.998	0.082	3.994	0.222
$\sigma_{Y1}(\mathbf{5})$	5.013	0.145	4.995	0.082
$\sigma_{Y2}(\mathbf{5})$	5.000	0.145	3.997	0.144
$\beta_{01}(\mathbf{4})$	4.037	0.404	4.019	0.403
$\beta_{02}(\mathbf{2})$	2.009	0.238	2.004	0.238
INB($\mathbf{2}$)	2.061	1.222	2.030	1.220

Table 4.7: Results of NA and MCMC methods for n = 700,$\text{MCMC}_{\text{power}} = 0.52$, $\text{NA}_{\text{power}} = 0.52$

	MCMC's Est	MCMC's SD	NA's Est	NA's SD
$\alpha_1(\mathbf{2})$	1.993	0.099	1.991	0.099
$\alpha_2(\mathbf{4})$	4.036	0.207	4.030	0.214
$\beta_1(\mathbf{1.5})$	1.499	0.032	1.499	0.031
$\mu_{X1}(\mathbf{8})$	7.995	0.215	7.983	0.134
$\mu_{X2}(\mathbf{4})$	3.997	0.075	3.994	0.207
$\sigma_{Y1}(\mathbf{5})$	5.012	0.135	4.996	0.075
$\sigma_{Y2}(\mathbf{5})$	4.993	0.134	4.030	0.133
$\beta_{01}(\mathbf{4})$	3.987	0.374	3.970	0.372
$\beta_{02}(\mathbf{2})$	1.983	0.221	1.979	0.219
INB($\mathbf{2}$)	2.015	1.131	1.985	1.127

Finally, we look at the powers of each method. Not only do the powers we obtain from the different methods yield the same value, but also both of them show a regular pattern of increasing when the sample size increases. The rate of increase tends to slow down when the sample size went over 600.

In order to have a more intuitive understanding of the simulation results and easier comparison of the normal approximation to the Bayesian MCMC method, we choose to plot two of the parameters of interest out of the ten parameters we have recorded. Here is the box plot of INB and β_1 respectively from sample size 100 to 700.

The circles in the middle of each vertical lines are the means of the estimation for normal approximation(on the left) and the Bayesian MCMC method(on the right), and the whiskers show the upper and lower 15 percentile. The horizontal lines across all the sample sizes are the true values of INB and β_1.

Based upon information from the graphs, it is reasonable to say that as the sample sizes increases, the estimations are getting closer to the true values, and the standard deviations of those estimations are decreasing. However, we can also see from the graphs that the standard errors for estimating INBs are relatively large. When the sample size is not large enough(n = 100 and 200), the estimation could even be negative, which will definitely mislead our conclusion.

4.5 Conclusions and Discussion

In this chapter, we have applied the Bayesian MCMC and the normal theory approximation method onto a cost-effective analysis. We have provided a Bayesian regression model and derived a normal theory approximation power study for known cost and efficacy data. The conclusion is similar to what we have in Chapter 2 and 3, which also agrees to what we expected.

We explored the application in cost-effective data where it is usually presented in gamma distribution. The results of both methods are fairly close. It showed that fitting with normal approximation when the data is log normal or gamma distributed can speed up the power study and we think it will also benefit the sample size determination and the sensitivity analysis.

We do want to mention that in order to get better and closer power, we need to think much of the prior selection stage in the Bayesian approach. The priors needs to be really diffuse and non-informative, and the precision of a parameter should be no larger than 0.0001.

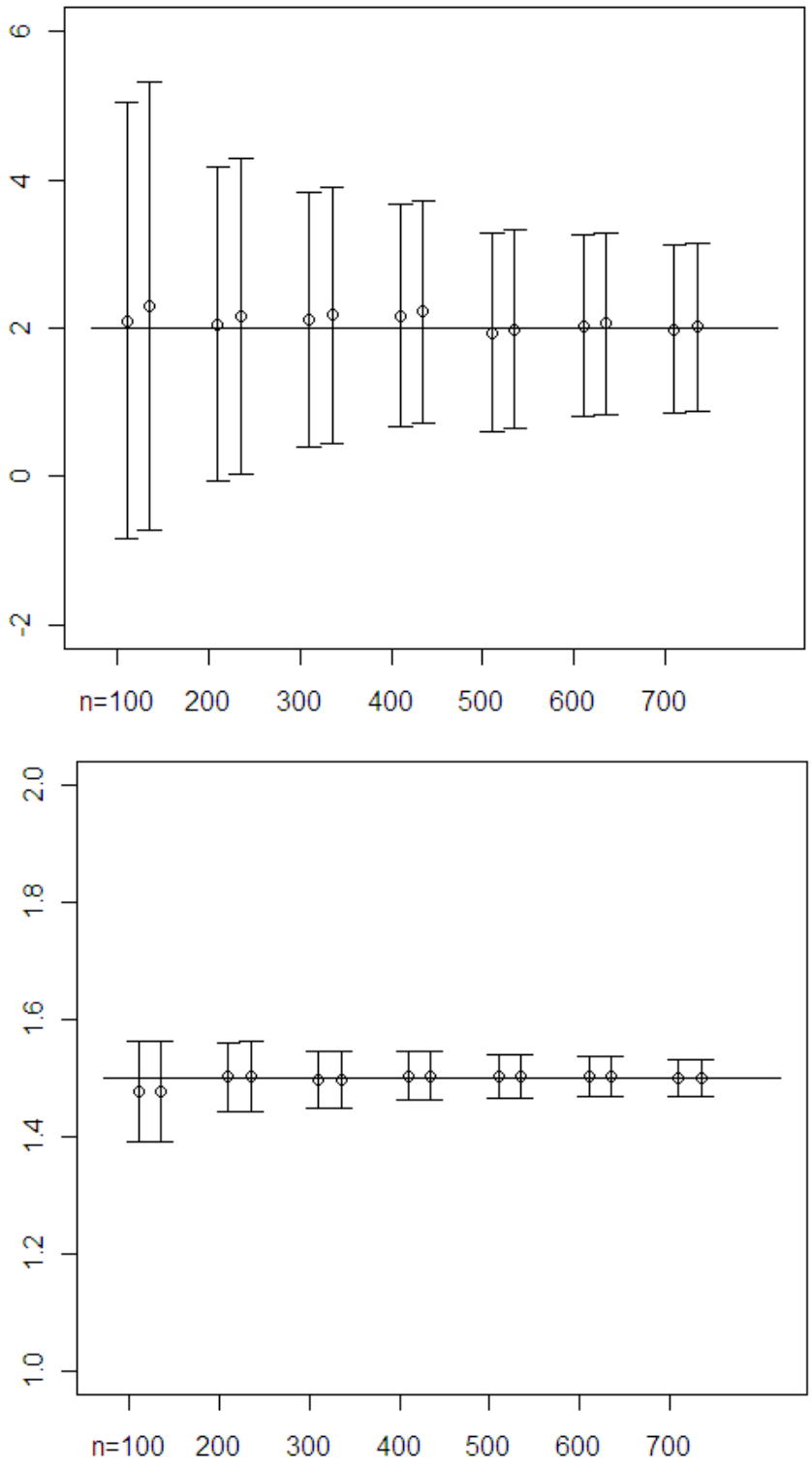

Figure 4.2: Plot of INB and β_1 from Normal Approximation and MCMC methods

In the future, we may introduce some unmeasured confounders into the cost-effectiveness analysis, because failing to account for the uncontrolled confounding can result in a biased estimation in cost-effectiveness studies, and is sometimes overlooked. We are very interested to see how Bayesian MCMC method performs compared to the normal theory approximation when an unmeasured confounder exists in cost benefit and cost-effectiveness analyses of the health field.

CHAPTER FIVE

Final Comments

In this dissertation, we mainly focused on comparisons of the Bayesian MCMC approach with the normal theory approximation approach under various circumstances in parameter estimation. We explored different scenarios under discrete and continuous responses with and without an unmeasured confounder. We also derived the corresponding normal approximations and determined whether it is feasible to use the normal approximation in different scenarios. We learned about the advantages and disadvantages of both methods during the process.

In Chapter two, we presented a logistic model with binary responses and covariates when unmeasured confounding exists. We showed that a small sample of validation data can be helpful in providing more information about the unmeasured confounding. Through simulations, we confirmed that both of our methods did a good job of estimating true parameter values. While the Bayesian MCMC method achieves smaller variation, the normal theory approximation leads to much faster power studies.

In Chapter three, we expanded the model with an associated normal response to a model with a gamma distributed response. We constructed only one binary covariate and one unmeasured confounder in the model, and then fit a Bayesian regression and a normal theory approximation on the gamma response. We discovered that the results of the Bayesian MCMC method are generally the same as the results of the more time efficient normal theory approximation. However, due to the significantly larger standard errors from the normal approximation, we do not recommend applying normal theory approximation in this scenario. Also, the likelihood function that is required in the normal theory approximation method cannot

always be easily derived, so the Bayesian MCMC method has some advantages over the normal approximation due to the incorporation of some prior information.

In Chapter four, we focused on the incremental net benefit in a cost-effectiveness analysis using both methods to estimate the true parameters. We provided a Bayesian regression model and derived normal theory approximation power studies for known gamma cost and normal efficacy. The results of both methods are fairly close. It showed that fitting with normal approximation when the data is log-normal or gamma distributed can speed up the power study. It can also benefit the sample size determination as well as the sensitivity analyses. In the future, we are also interested in looking at unmeasured confounders in the cost-effectiveness analysis with both Bayesian and normal theory approximation approaches.

APPENDICES

APPENDIX A

Chapter Two Calculation Details

This appendix contains the details results from Section 2.2.2 and 2.4.2.

A.1 The Second Derivatives in Section 2.2.2

$$\frac{\partial^2 \log L_N}{\partial\beta_0\partial\beta_1} = \frac{\partial^2 \log L_N}{\partial\beta_1\partial\beta_0} = \sum_{i=1}^{n} -\frac{(y_i A)\,(x_i y_i B)}{E} + \frac{x_i y_i^2 (C+D)}{F}$$

$$\frac{\partial^2 \log L_N}{\partial\beta_0\partial\beta_2} = \frac{\partial^2 \log L_N}{\partial\beta_2\partial\beta_0} = \sum_{i=1}^{n} -\frac{(y_i A)\,(y_i z_i^2 B)}{E} + \frac{y_i z_i^2 (C+D)}{F}$$

$$\frac{\partial^2 \log L_N}{\partial\beta_1\partial\beta_2} = \frac{\partial^2 \log L_N}{\partial\beta_2\partial\beta_1} = \sum_{i=1}^{n} -\frac{(x_i y_i A)\,(y_i z_i B)}{E} + \frac{x_i y_i^2 z_i (C+D)}{F}$$

where

$$A = -\frac{e^{2y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{\left(1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}} - \frac{e^{2y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{\left(1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}$$

$$B = -\frac{e^{2y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{\left(1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}} - \frac{e^{2y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{\left(1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}$$

$$C = \frac{2e^{3y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{\left(1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}\right)^3} - \frac{3e^{2y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{\left(1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}$$

$$D = \frac{2e^{3y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{\left(1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}\right)^3} - \frac{3e^{2y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{\left(1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}\right)^2} + \frac{e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}$$

$$E = \left(\frac{e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}} + \frac{e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}\right)^2$$

$$F = \frac{e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}{1+e^{y_i\left(\boldsymbol{M}_i'\boldsymbol{\beta}\right)}} + \frac{e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)+\gamma_0+x_i\gamma_1}}{1+e^{y_i\left(\lambda+\boldsymbol{M}_i'\boldsymbol{\beta}\right)}}$$

$$\begin{aligned}
\frac{\partial^2 \log L_{N_1}}{\partial\beta_0\partial\beta_1} &= \frac{\partial^2 \log L_{N_1}}{\partial\beta_1\partial\beta_0} \\
&= \sum_{j=1}^{n_1} x_{1j}y_{1j}^2 \left[e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \right. \\
&\quad \times \left(-\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&\quad \left. + \frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \\
&\quad - e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \\
&\quad \times \left(-\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&\quad \left. + \frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \\
&\quad + e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \\
&\quad \times \left(\frac{2e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+2y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^3\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&\quad - \frac{3e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \\
&\quad \left. \left. + \frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \log L_{N_1}}{\partial\beta_0\partial\beta_2} &= \frac{\partial^2 \log L_{N_1}}{\partial\beta_2\partial\beta_0} \\
&= \sum_{j=1}^{n_1} y_{1j}^2 z_{1j} \left[e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)} \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \right.
\end{aligned}$$

$$
\begin{aligned}
&\times \left(-\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&\left. +\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \\
&-e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \\
&\times \left(-\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&\left. +\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \\
&+e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \\
&\times \left(\frac{2e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+2y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^3\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&-\frac{3e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \\
&\left. +\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L_{N_1}}{\partial\beta_1\partial\beta_2} &= \frac{\partial^2 \log L_{N_1}}{\partial\beta_2\partial\beta_1} \\
&= \sum_{j=1}^{n_1} x_{1j}y_{1j}^2 z_{1j} \left[e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \right. \\
&\quad \times \left(-\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right. \\
&\quad \left. +\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right) \\
&\quad -e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right) \\
&\quad \times \left(-\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)} \right.
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\Bigg)\\
&+e^{-y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)-u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\\
&\times\left(\frac{2e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+2y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^3\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\right.\\
&-\frac{3e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)^2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\\
&\left.+\frac{e^{y_{1j}\left(\boldsymbol{N}_j'\boldsymbol{\beta}\right)+u_{1j}(\lambda y_{1j}+\gamma_0+x_{1j}\gamma_1)}}{\left(1+e^{y_{1j}\left(\lambda u_{1j}+\boldsymbol{N}_j'\boldsymbol{\beta}\right)}\right)\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)}\right)\Bigg]
\end{aligned}
$$

A.2 The Second Derivatives in Section 2.4.2

$$
\frac{\partial^2\log L_{N_1}}{\partial\beta_0\partial\beta_1} = -\sum_{j=1}^{n_1}\frac{x_{1j}}{\sigma^2}
$$

$$
\frac{\partial^2\log L_{N_1}}{\partial\beta_0\partial\beta_2} = -\sum_{j=1}^{n_1}\frac{z_{1j}}{\sigma^2}
$$

$$
\frac{\partial^2\log L_{N_1}}{\partial\beta_1\partial\beta_2} = -\sum_{j=1}^{n_1}\frac{x_{1j}z_{1j}}{\sigma^2}
$$

$$
\begin{aligned}
\frac{\partial^2\log L_{N_1}}{\partial\beta_0\partial\sigma^2} = &\sum_{j=1}^{n_1}\left\{e^{\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}-u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\frac{-\sqrt{2\pi}}{\sigma}\right.\\
&\times\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)\\
&\times\left(-\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^3}\right.\\
&\left.+\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^5}\right)\\
&+e^{\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}-u_{1j}(\gamma_0+x_{1j}\gamma_1)}
\end{aligned}
$$

$$
\begin{aligned}
&\times \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sqrt{\sigma^2} \\
&\times \left[-\frac{3e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^5} \right. \\
&\left. +\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^3}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^7} \right] \Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L_{N_1}}{\partial\beta_1\partial\sigma^2} &= \sum_{j=1}^{n_1} x_{1j}\left\{ e^{\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}-u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\frac{-\sqrt{2\pi}}{\sigma}\right. \\
&\times \left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right) \\
&\times \left(-\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^3} \right. \\
&\left. +\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^5} \right) \\
&+e^{\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}-u_{1j}(\gamma_0+x_{1j}\gamma_1)} \\
&\times \left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sqrt{\sigma^2} \\
&\times \left[-\frac{3e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^5} \right. \\
&\left. \left. +\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^3}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^7} \right] \right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L_{N_1}}{\partial\beta_2\partial\sigma^2} &= \sum_{j=1}^{n_1} z_{1j}\left\{ e^{\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}-u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\frac{-\sqrt{2\pi}}{\sigma}\right. \\
&\times \left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right) \\
&\times \left(-\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^3} \right. \\
&\left. \left. +\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^5} \right) \right.
\end{aligned}
$$

$$+e^{\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}-u_{1j}(\gamma_0+x_{1j}\gamma_1)}$$

$$\times\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sqrt{\sigma^2}$$

$$\times\left[-\frac{3e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^5}\right.$$

$$\left.+\frac{e^{-\frac{\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^2}{2\sigma^2}+u_{1j}(\gamma_0+x_{1j}\gamma_1)}\left(-\lambda u_{1j}+y_{1j}-\boldsymbol{N}_j'\boldsymbol{\beta}\right)^3}{2\left(1+e^{\gamma_0+x_{1j}\gamma_1}\right)\sqrt{2\pi}\sigma^7}\right]\Bigg\}$$

APPENDIX B

Chapter Two code

This appendix contains the R program and WinBUGS model used for the simulation presented in Chapter Two.

B.1 Normal Approximation of Binary Response

```
norm.approx.ber<-function(m,n, n1, a1,b1, a2, b2, a3, b3, a4, b4,
  a5, b5, a6, b6, s)
 {
# sets the seed so simulation is repeatable
   set.seed(s)
# Set up vectors to store results from the simulation runs

   mean.b1<-rep(NA, m)
    sd.b1<-rep(NA, m)
   power <-rep(NA, m)

# compute sample size for subjects with unmeasured confounder

   n.main<-n-n1
   for(i in 1:m)
   {
# design prior parameter generation, one of these sets should
# be commented out and recall the parameters are either means
# and standard deviations or bounds for the uniform distributions

B0<-runif(1,a1,b1)
  B1<-runif(1,a2,b2)
  B2<-runif(1,a3,b3)
  G0<-runif(1,a4,b4)
G1<-runif(1,a5,b5)
L<-runif(1,a6,b6)

# Generating the data for the main study portion.  Top is
#"reduced model" and bottom is "expanded model"
# Select the model of interest and comment the other one out.

# z is an observed confounder
# x is the treatment variable
```

```
# u is the unobserved confounder, function of only x
# y is the outcome variable
# only z, x, and y are used in the data analysis

  z <- rbinom(n.main,1,0.4)
x <- rbinom(n.main,1,0.6)
u <- rbinom(n.main, 1, 1/(1 + exp(-(cbind(1,x)
%*%c(G0, G1)))))
y <- rbinom(n.main, 1, 1/(1 + exp(-(cbind(1,x,u,z)
%*%c(B0, B1, L, B2)))))

# Generating the data for the validation data portion
# z1 is an observed confounder
# x1 is the treatment variable
# u1 is the unobserved confounder, function of only x
# y1 is the outcome variable
# all four data vectors are used in the analysis

z1 <- rbinom(n1,1,0.4)
x1 <- rbinom(n1,1,0.6)
u1 <- rbinom(n1, 1, 1/(1 + exp(-(cbind(1,x1) %*%
c(G0, G1)))))
y1 <- rbinom(n1, 1, 1/(1 + exp(-(cbind(1,x1,u1,z1)
%*% c(B0, B1, L, B2)))))
##########################################

v.data<-cbind(x1,z1,y1,u1)

data<-cbind(x,z,y)

loglike <- function (beta) {
beta0 <- beta[1]
beta1 <- beta[2]
beta2 <- beta[3]
lambda <- beta[4]
gamma0<-beta[5]
gamma1<-beta[6]

logN <-sum(
y*(beta0 + beta1*x + beta2*z) - log(1 + exp(gamma0 +
 gamma1*x)) + log(1 + exp(gamma0 + gamma1*x + lambda*y) +
 exp(gamma0 + beta0 + (beta1 + gamma1)*x + beta2*z + lambda*y)
 + exp(beta0 + beta1*x + beta2*z+ lambda))-
 log(1 + exp(beta0 + beta1*x + beta2*z + lambda)) -
log(1 + exp(beta0 + beta1*x + beta2*z))
```

```
)

logN1 <- sum(y1*(beta0 + beta1*x1 + beta2*z1 + lambda*u1)
 - log(1 + exp(beta0 + beta1*x1 + beta2*z1 + lambda*u1))+
u1*(gamma0 + gamma1*x1) - log(1 + exp(gamma0 + gamma1*x1)))

return(-logN-logN1)
}

result<-optim(c(-1,1,1,-1, 0, 1),loglike,method="BFGS",hessian=T)

beta1.hat<-result$par[2]
as.var.inv<-result$hessian
as.var<-solve(as.var.inv)
post.prob<-pnorm(beta1.hat/as.var[2, 2]^.5)
#################################################################

  mean.b1[i]<-beta1.hat
    sd.b1[i]<-as.var[2, 2]^.5
    power[i]<-ifelse(post.prob>.95, 1, 0)
   }

   ave_mean<-mean(mean.b1)  ## calcualate average mean
   ave_sd<-mean(sd.b1)  ##
   ave_power<-mean(power)

   return(c(ave_mean,ave_sd,ave_power))
 }
```

B.2 Bayesian MCMC of Binary Response

```
Model
{
for(i in 1 : n.main) {
y[i] ~ dbern(p[i])
U[i]~dbern(q[i])
logit(p[i]) <- beta0 + beta1 * x[i]  +beta2*z[i] + lambda*U[i]
logit(q[i]) <- gamma0+gamma1 * x[i]
}

for(j in 1 : n1) {
y1[j] ~ dbern(p1[j])
```

```
u1[j]~dbern(q1[j])
logit(p1[j]) <- beta0 + beta1 * x1[j]+beta2*z1[j]+lambda*u1[j]
logit(q1[j]) <- gamma0+gamma1 * x1[j]
}

beta0 ~ dnorm(0.0,.1)
beta1 ~ dnorm(0.0,.1)
beta2 ~ dnorm(0.0,.1)
p.value<-step(beta1)
gamma0~dnorm(0.0, .1)
gamma1~dnorm(0.0, .1)
lambda~dnorm(0.0, .1)
}



bayes.ber<-function(m,n, n1, a1,b1, a2, b2, a3, b3, a4, b4,
a5, b5, a6, b6, s)
 {
# sets the seed so simulation is repeatable
   set.seed(s)

# Set up vectors to store results from the simulation runs
   mean.b1<-rep(NA, m)
   sd.b1<-rep(NA, m)

# compute sample size for subjects with unmeasured confounder
   n.main<-n-n1

   for(i in 1:m)
   {

B0<-runif(1,a1,b1)
        B1<-runif(1,a2,b2)
        B2<-runif(1,a3,b3)
        G0<-runif(1,a4,b4)
G1<-runif(1,a5,b5)
L<-runif(1,a6,b6)
  z <- rbinom(n.main,1,0.4)
x <- rbinom(n.main,1,0.6)
u <- rbinom(n.main, 1, 1/(1 + exp(-(cbind(1,x) %*%
c(G0, G1)))))
y <- rbinom(n.main, 1, 1/(1 + exp(-(cbind(1,x,u,z)
%*% c(B0, B1, L, B2)))))
```

```
z1 <- rbinom(n1,1,0.4)
x1 <- rbinom(n1,1,0.6)
u1 <- rbinom(n1, 1, 1/(1 + exp(-(cbind(1,x1) %*%
c(G0, G1)))))
y1 <- rbinom(n1, 1, 1/(1 + exp(-(cbind(1,x1,u1,z1)
%*% c(B0, B1, L, B2)))))

# This next portion of the code gets the data ready to be sent
# to WinBUGS for this study we assume beta1 is the parameter
# of primary interest. If other parameters are also of interest,
# they can be added here as well, but vectors collecting the
# output also need to be added.

# For reduced model

parameters<-list("beta0", "beta1", "beta2", "gamma0",
"gamma1",  "lambda", "p.value")

# this is the data for WinBUGS

  data<-list("n.main", "n1", "z", "x",  "y", "z1", "x1",
"y1", "u1")

# This vector is for initial values.  Initial values are
# important for this model since the unmeasured confounding
# yields a lack of identifiability that is remedied by
# the validation data, but if the validation data sample
# size is small poor mixing can result if "bad" initial
# values are used.  We recommend the means of the
# design priors

    inits<-list(beta0=-2, beta1=.5, beta2=.2, lambda = -.5,
gamma0=-.2, gamma1=1.1)
    inits<-list(inits)

# this line calls WinBUGS, note that n.iter is TOTAL
# iterations (including burnin)
# "model int" should be used for interaction model and
# "model red" should be used for reduced

  ss.sim<-bugs(data,inits,parameters,"model red.txt",
n.chains=1, n.burnin=5000,n.iter=20000,  n.thin=2, debug=TRUE)

   mean.b1[i]<-ss.sim$summary[2, 1]
```

```
   sd.b1[i] <- ss.sim$summary[2, 2]
       }

   ave_mean<-mean(mean.b1)  ## calcualate average mean
   ave_sd  <-mean(sd.b1)  ## calcualate average sd
   return(c(ave_mean, ave_sd))
 }
```

B.3 Misclassification Example

```
#### The function "ss" calculates average Bayesian power when
#### one imperfect diagnostic test is used and one covariate is
#### included in the binomial regression model.
## x: covariate
## y: response
## beta0: intercept coefficient
## beta1: slope coefficient
## s: sensitivity
## c: specificity
## m: number of population
## n: sample size
## as,bs: prior parameters of s
## ac,bc: prior parameters of c
## a0,b0: prior parameters of beta0
## s : the random number generator seed

library(R2WinBUGS)

ss<-function(m,n,as,bs,ac,bc,a0,b0,s)
{
# sets the seed so simulation is repeatable
  set.seed(s)

  mean.b1 <- rep(NA, m)
  sd.b1 <- rep(NA, m)
  power<-rep(NA,m)

  for(i in 1:m)
  {

    x<-rbinom(n,1, .5)  ## simulate the predictor
    B0<-rnorm(1,a0,b0)
    B1<-0.3
```

```
    pi<-exp(B0+B1*x)/(1+exp(B0+B1*x))
    s<-rbeta(1,as,bs)
    c<-rbeta(1,ac,bc)
    p<-pi*s+(1-pi)*(1-c)
    y<-rbinom(n,1,p)  ## simulate the binomial response

    data<-list("n","x","y")  ## save data

    parameters<-list("beta1","p.value")

# This vector is for initial values.
    inits<-list(beta0=-0.8,beta1=0.5,se=0.5,sp=0.5)
    inits<-list(inits)


# this line calls WinBUGS, note that n.iter is TOTAL iterations.
  ss.sim<-bugs(data,inits,parameters,"model1.txt", n.chains=1,
n.burnin=1000,n.iter=5000, bugs.directory = "D:/Software/WinBUGS14/", debug=F)

  mean.b1[i]<-ss.sim$summary[1, 1]
   sd.b1[i] <- ss.sim$summary[1, 2]
 power[i]<-ss.sim$summary[2, 1]
   }

   ave_mean <- mean(mean.b1)  ## calcualate average mean
   ave_sd  <- mean(sd.b1)  ## calcualate average sd
   ave_p <- mean(power)  ## calcualate average Bayesian power

   return(c(ave_mean, ave_sd,ave_p))
 }
ss(1,800,80,20,92,8,-1,0.2,26)


############################
##       model1.txt       ##
############################

model
{
  for(i in 1:n)
  {
    y[i]~dbern(p[i])
    p[i]<-pi[i]*se+(1-pi[i])*(1-sp)
    logit(pi[i])<-beta0+beta1*x[i]
```

```
  }

  beta0~dnorm(0,0.1)
  beta1~dnorm(0,0.1)

p.value<-step(beta1)

  se~dbeta(80,20)
  sp~dbeta(92,8)
}

#############################################################


norm.approx<-function(m,n,as,bs,ac,bc,a0,b0,s)
{

set.seed(s)

  mean.b1 <- rep(NA,m)
  sd.b1 <- rep(NA,m)
  power <- rep(NA,m)
  for(i in 1:m)
  {
    x<-rbinom(n,1, .5)  ## simulate the predictor
    B0<-rnorm(1,a0,b0)
    B1<-0.3
    Pi<-exp(B0+B1*x)/(1+exp(B0+B1*x))
    S<-rbeta(1,as,bs)
    C<-rbeta(1,ac,bc)
    p<-Pi*S+(1-Pi)*(1-C)
    y<-rbinom(n,1,p)  ## simulate the binomial response


 loglike <- function(beta) {

  beta0 <- beta[1]
  beta1 <- beta[2]
  s <- beta[3]
  c <- beta[4]
  ppi<-exp(beta0+beta1*x)/(1+exp(beta0+beta1*x))
 l <- rep(NA,3)
 l[1] <-prod((ppi*s + (1-ppi)*(1-c))^y)
 l[2] <-prod((ppi*(1 - s) + (1-ppi)*c)^(1-y))
```

```
 l[3] <-s^as*(1 - s)^bs*c^ac*(1 - c)^bc

  llk <- sum(log(l))
  if (is.na(llk))
          llk <- -1e+6
     if (llk == -Inf)
       llk <- -1e+6
    return(-llk) }

result<-optim(c(-0.8,0.5,0.5,0.5),loglike,method="L-BFGS-B",
lower = c(-2,-100,0,0),upper = c(0,100,1,1),hessian=T)


beta1.hat<-result$par[2]
as.var.inv<-result$hessian
as.var<-solve(as.var.inv)
power[i] <-pnorm(beta1.hat/as.var[2, 2]^.5)
 }

   ave_mean<-mean(mean.b1)  ## calcualate average mean
   ave_sd<-mean(sd.b1)  ##
   ave_power <- mean(power)
   return(c(ave_mean,ave_sd,ave_power))
 }
norm.approx(1,800,80,20,92,8,-1,0.2,26)
```

B.4 Continuous Outcome Example

```
######Normal Approximation########
 norm.approx.ber<-function(m,n, n1, a1,b1, a2, b2, a3, b3, a4,
 b4,  a5, b5, a6, b6, a7, b7,s)
  {
 # sets the seed so simulation is repeatable
     set.seed(s)

 # Set up vectors to store results from the simulation runs
     mean.b1<-rep(NA, m)
     sd.b1<-rep(NA, m)
    power<-rep(NA,m)

 # compute sample size for subjects with unmeasured confounder
```

```
    n.main<-n-n1

   for(i in 1:m)
   {

 B0<-runif(1,a1,b1)
 B1<-runif(1,a2,b2)
 G0<-runif(1,a4,b4)
 G1<-runif(1,a5,b5)
 L<-runif(1,a6,b6)
 sigma.y<-runif(1, a7, b7)

 x <- rbinom(n.main,1,0.6)
 u <- rbinom(n.main, 1, 1/(1 + exp(-(cbind(1,x) %*% c(G0, G1)))))
 y <- rnorm(n.main, cbind(1,x,u,z) %*% c(B0, B1, L, B2), sigma.y)


 x1 <- rbinom(n1,1,0.6)
 u1 <- rbinom(n1, 1, 1/(1 + exp(-(cbind(1,x1) %*% c(G0, G1)))))
 y1 <- rnorm(n1, cbind(1,x1,u1,z1) %*% c(B0, B1, L, B2), sigma.y)

 ##########################################
 v.data<-cbind(x1,y1,u1)
 data<-cbind(x,y)

 loglike <- function (beta) {
 beta0 <- beta[1]
 beta1 <- beta[2]
 lambda <- beta[3]
 gamma0<-beta[4]
 gamma1<-beta[5]
 sigma<-beta[6]

 mu1<-beta0+beta1*x+lambda
 mu0<-beta0+beta1*x

 logN <-sum(log(dnorm(y, mu1, sigma)*1/(1+exp(-gamma0-gamma1*x))+
dnorm(y, mu0, sigma)*1/(1+exp(gamma0+gamma1*x))))

 mu11<-beta0+beta1*x1+lambda*u1

 logN1 <- sum(log(dnorm(y1, mu11, sigma))+
   u1*log(1/(1+exp(-gamma0-gamma1*x1)))+
(1-u1)*log(1-1/(1+exp(-gamma0-gamma1*x1))))
```

```
 return(-logN-logN1)
 }

result<-optim(c(100, 30, -50, -.2, 1, 100),loglike,
method="BFGS",hessian=T)

beta1.hat<-result$par[2]
as.var.inv<-result$hessian
as.var<-solve(as.var.inv)

mean.b1[i]<-beta1.hat
sd.b1[i]<-as.var[2, 2]^.5

post.prob<-pnorm(beta1.hat/as.var[2, 2]^.5)
power[i]<- post.prob
}
ave_mean<-mean(mean.b1)  ## calcualate average mean
ave_sd<-mean(sd.b1)  ## calcualate average standard error
ave_p<-mean(power)  ## calcualate average Bayesian power
return(c(ave_mean,ave_sd,ave_p))
  }
norm.approx.cts(1, 500, 300, 100, 150, 30, 30, 20, 30, -.3,
-.2,  1.1, 1.3, -30, -20, 150, 175, 265)




##########MCMC Method####################
library(R2WinBUGS)
ss.counfound.1<-function(m,n, n1, a1,b1, a2, b2, a3, b3, a4,
 b4, a5, b5, a6, b6, a7, b7, s)
 {
   set.seed(s)
   mean.b1<-rep(NA, m)
   sd.b1<-rep(NA, m)
   n.main<-n-n1
for(i in 1:m)
   {
    B0<-runif(1,a1,b1)
    B1<-runif(1,a2,b2)
    G0<-runif(1,a4,b4)
G1<-runif(1,a5,b5)
L<-runif(1,a6,b6)
sigma.y<-runif(1, a7, b7)
```

```
x <- rbinom(n.main,1,0.6)
u <- rbinom(n.main, 1, 1/(1 + exp(-(cbind(1,x) %*% c(G0, G1)))))
y <- rnorm(n.main, cbind(1,x,u) %*% c(B0, B1, L), sigma.y)

x1 <- rbinom(n1,1,0.6)
u1 <- rbinom(n1, 1, 1/(1 + exp(-(cbind(1,x1) %*% c(G0, G1)))))
y1 <- rnorm(n1, cbind(1,x1,u1) %*% c(B0, B1, L), sigma.y)

parameters<-list("beta0", "beta1","gamma0","gamma1","lambda",
"sig","p.value")

  data<-list("n.main", "n1", "x", "y", "x1", "y1", "u1")

     inits<-list(beta0=B0, beta1=B1, lambda = L, gamma0=G0,
gamma1=G1,sig=sigma.y,u=rep(1,n.main),u1=rep(1,n1))
     inits<-list(inits)

    ss.sim<-bugs(data,inits=inits,parameters,"contin.txt",
bugs.directory="C:/Users/jiang_yuan/Desktop/winbugs14/WinBUGS14/",
            n.chains=1,n.burnin=5000,n.iter=10000, n.thin=2,debug=F)
    mean.b1[i]<-ss.sim$summary[2, 1]
    sd.b1[i]<-ss.sim$summary[2, 2]
    post.prob<-ss.sim$summary[7, 1]
   power[i]<-post.prob
   }

   ave_p<-mean(power)  ## calcualate average Bayesian power
   ave_mean<-mean(mean.b1)  ## calcualate average mean
  ave_sd <- mean(sd.b1)
   return(c(ave_p, ave_mean,ave_sd))
 }
ss.counfound.1(1, 500, 300, 100, 150, 30, 30, 20, 30, -.3, -.2,
  1.1, 1.3, -30, -20, 150, 175, 265)

############################
##       contin.txt       ##
############################
Model
{
for(i in 1 : n.main) {
y[i] ~ dnorm(mu[i], tau)
u[i] ~ dbern(q[i])
mu[i] <- beta0 + beta1 * x[i]  + lambda * u[i]
logit(q[i]) <- gamma0+gamma1 * x[i]
```

```
}

for(j in 1 : n1) {

y1[j] ~ dnorm(mu1[j], tau)
u1[j] ~ dbern(q1[j])

mu1[j] <- beta0 + beta1 * x1[j]+ lambda *u1[j]
logit(q1[j]) <- gamma0+gamma1 * x1[j]
}

beta0 ~ dnorm(0.0,.0000001)
beta1 ~ dnorm(0.0,.0000001)

gamma0~dnorm(0.0, .1)
gamma1~dnorm(0.0, .1)

lambda~dnorm(0.0, .0000001)
tau<-1/(sig*sig)
sig~dunif(0.01, 500)
p.value<-step(beta1)
}
```

APPENDIX C

Chapter Three Code

This appendix contains the R program and WinBUGS model used for the simulation presented in Chapter Three.

C.1 Normal Approximation of Gamma Response

```
norm.approx.gam<-function(m,n, n1, a1,b1, a2, b2, a3, b3, a4, b4,
  a5, b5, a6, b6, s)
 {
# sets the seed so simulation is repeatable
   set.seed(s)

# Set up vectors to store results from the simulation runs

   mean.b1<-rep(NA, m)
    sd.b1<-rep(NA, m)
     power <-rep(NA, m)

# compute sample size for subjects with unmeasured confounder
   n.main<-n-n1

   for(i in 1:m)
   {


B0<-runif(1,a1,b1)
        B1<-runif(1,a2,b2)
        Alpha<-runif(1,a3,b3)
        G0<-runif(1,a4,b4)
G1<-runif(1,a5,b5)
L<-runif(1,a6,b6)

# Generating the data for the main study portion.
# Top is "reduced model" and bottom is "expanded model"
# Select the model of interest and comment the other one out.

# x is the treatment variable
# u is the unobserved confounder, function of only x
# y is the outcome variable
```

```
# only x, and y are used in the data analysis

x <- rbinom(n.main,1,0.6)
u <- rbinom(n.main, 1, 1/(1 + exp(-(cbind(1,x) %*%
c(G0, G1)))))
y <- rgamma(n.main, Alpha, Alpha * exp(-(cbind(1,x,u)
%*% c(B0, B1, L))))

# Generating the data for the validation data portion

x1 <- rbinom(n1,1,0.6)
u1 <- rbinom(n1, 1, 1/(1 + exp(-(cbind(1,x1) %*%
c(G0, G1)))))
y1 <- rgamma(n1, Alpha, Alpha * exp(-(cbind(1,x1,u1)
%*% c(B0, B1, L))))

##########################################

v.data<-cbind(x1,y1,u1)
data<-cbind(x,y)


loglike <- function (beta) {
beta0 <- beta[1]
beta1 <- beta[2]
alpha <- beta[3]
lambda <- beta[4]
gamma0<-beta[5]
gamma1<-beta[6]

mu1<-exp(beta0+beta1*x+lambda)
mu0<-exp(beta0+beta1*x)

logN <-sum(log(dgamma(y, alpha, alpha/mu1)*1/(1+exp(-gamma0-
gamma1*x))+dgamma(y, alpha, alpha/mu0)*1/(1+exp(gamma0+gamma1*x))))

mu11<-exp(beta0+beta1*x1+lambda*u1)

logN1 <- sum(log(dgamma(y1, alpha,alpha/mu11))+
u1*log(1/(1+exp(-gamma0-gamma1*x1)))+(1-u1)*log(1-1/(1+
exp(-gamma0-gamma1*x1))))

return(-logN-logN1)
}
```

```
result<-optim(c(-0.5,0.5,4,-.5, 1, -0.6),loglike,
method="BFGS",hessian=T)

beta1.hat<-result$par[2]
as.var.inv<-result$hessian
as.var<-solve(as.var.inv)
post.prob<-pnorm(beta1.hat/as.var[2, 2]^.5)

#################################################################


   mean.b1[i]<-beta1.hat
    sd.b1[i]<-as.var[2, 2]^.5
         power[i]<-ifelse(post.prob>.95, 1, 0)
}


   ave_mean<-mean(mean.b1)  ## calcualate average mean
   ave_sd<-mean(sd.b1)  ##
   ave_power <- mean(power)
   return(c(ave_mean,ave_sd,ave_power))
 }
```

C.2 Bayesian MCMC Model of Gamma Response

```
model
{
# validation study data
for (i in 1:n1) {
y1[i] ~ dgamma(alpha,tau1[i])
tau1[i] <- alpha/mu1[i]

u1[i] ~ dbern(q1[i])
log(mu1[i]) <- beta0 + beta1 * x1[i] + lambda * u1[i]
logit(q1[i]) <- gamma0 + gamma1 * x1[i]
}

# Main study data
for (j in 1:n.main) {
y[j] ~ dgamma(alpha,tau[j])
tau[j] <- alpha/mu[j]
```

```
u[j] ~ dbern(q[j])
log(mu[j]) <- beta0 + beta1 * x[j] + lambda * u[j]
logit(q[j]) <- gamma0 + gamma1 * x[j]
}

beta0 ~ dnorm(0.0,0.01)
beta1 ~ dnorm(0.0,0.01)
p.value<-step(beta1)
  gamma0~dnorm(0.0, .1)
gamma1~dnorm(0.0, .1)
lambda~dnorm(0.0, .1)
  alpha ~ dunif(.001, 50)
}
```

APPENDIX D

Chapter Four Code

D.1 Normal Approximation of the Cost-effectiveness

```
library(stats4)
norm.approx<-function(m,n,s)
 {
# sets the seed so simulation is repeatable
   set.seed(s)

# Set up vectors to store results from the simulation runs

  mean.ym1<-rep(NA, m)
   sd.ym1<-rep(NA, m)
 mean.ym2<-rep(NA, m)
   sd.ym2<-rep(NA, m)
 mean.b1<-rep(NA, m)
   sd.b1<-rep(NA, m)
 mean.alpha1<-rep(NA, m)
   sd.alpha1<-rep(NA, m)
 mean.alpha2<-rep(NA, m)
   sd.alpha2<-rep(NA, m)
 mean.mux1<-rep(NA, m)
   sd.mux1<-rep(NA, m)
 mean.mux2<-rep(NA, m)
   sd.mux2<-rep(NA, m)
 mean.sig1<-rep(NA, m)
   sd.sig1<-rep(NA, m)
 mean.sig2<-rep(NA, m)
   sd.sig2<-rep(NA, m)
 mean.inb<-rep(NA, m)
   sd.inb<-rep(NA, m)
covy1x1<-rep(NA, m)
covy1x2<-rep(NA, m)
covy2x1<-rep(NA, m)
covy2x2<-rep(NA, m)
p.inb <- rep(NA, m)

   for(i in 1:m)
```

```
   {
YM1 <- 4
YM2 <- 2
B1<- 1.5
Alpha1 <- 2
Alpha2 <- 4
#Alpha<-runif(1,a2,b2)
#Mux <- runif(1,a3,b3)
Mux1 <- 8
Mux2 <- 4
Sigma.y1<-5
Sigma.y2<-5

INB <- function(lambda, E, C){
lambda * E - C
}

lambda = 3


x1 <- rgamma(n, Alpha1, Alpha1 / Mux1)
y1 <- rnorm(n,cbind(1,x1-Mux1) %*% c(YM1, B1), Sigma.y1)
x2 <- rgamma(n, Alpha2, Alpha2 / Mux2)
y2 <- rnorm(n,cbind(1,x2-Mux2) %*% c(YM2, B1), Sigma.y2)
Tinb <- INB(3, YM1 - YM2, Mux1 - Mux2)

##########################################


loglike <- function (ym1,ym2,beta1,alpha1,alpha2,mux1,
mux2,sig1,sig2) {

logN1 <- sum(log(dgamma(x1,alpha1,alpha1/mux1))+log(dnorm(
y1,ym1+beta1*(x1-mux1),sig1)))
logN2 <- sum(log(dgamma(x2,alpha2,alpha2/mux2))+log(dnorm(
y2,ym2+beta1*(x2-mux2),sig2)))
return(-logN1-logN2)
}

est<-mle(minuslogl=loglike,method = "L-BFGS-B", lower =
rep(0.000001, 9),list(ym1 = 4,ym2=1,beta1 = 1,alpha1= 1,
alpha2 = 1, mux1= 1,sig1 = 1, mux2= 1,sig2 = 1))

mean.ym1[i]<-est@coef[1]
```

```
 sd.ym1[i] <- est@vcov[1,1]^.5

    mean.ym2[i]<-est@coef[2]
        sd.ym2[i] <- est@vcov[2, 2]^.5

 mean.b1[i]<-est@coef[3]
        sd.b1[i] <- est@vcov[3,3]^.5

 mean.alpha1[i]<-est@coef[4]
        sd.alpha1[i] <-est@vcov[4,4]^.5

 mean.alpha2[i]<-est@coef[5]
        sd.alpha2[i] <-est@vcov[5,5]^.5

 mean.mux1[i]<-est@coef[6]
        sd.mux1[i] <-est@vcov[6,6]^.5

 mean.mux2[i]<-est@coef[7]
        sd.mux2[i] <-est@vcov[7,7]^.5

 mean.sig1[i]<-est@coef[8]
        sd.sig1[i] <-est@vcov[8,8]^.5

 mean.sig2[i]<-est@coef[5]
        sd.sig2[i] <-est@vcov[9,9]^.5
  covy1x1[i] <-est@vcov[1,6]^.5
  covy2x2[i] <-est@vcov[2,7]^.5
  covy1x2[i] <-est@vcov[1,7]^.5
  covy2x1[i] <-est@vcov[2,6]^.5
   }
   mean.inb <- lambda *(mean.ym1 - mean.ym2) - (mean.mux1 -
mean.mux2)

sd.inb <-sqrt(lambda^2*sd.ym1^2 + lambda^2*sd.ym2^2 +
sd.mux1^2 + sd.mux2^2 - 2*lambda*covy1x1 - 2*lambda*covy2x2
 + 2*lambda*covy1x2 + 2*lambda*covy2x1)

     p.inb<-ifelse(pnorm(mean.inb/sd.inb)>.95, 1, 0)

means <- rbind(mean.ym1,mean.ym2,mean.b1,mean.alpha1,mean.alpha2,
mean.mux1,mean.mux2,mean.sig1,mean.sig2,mean.inb)

 sds <- rbind(sd.ym1,sd.ym2,sd.b1,sd.alpha1,sd.mux1,sd.sig1,
sd.alpha2,sd.mux2,sd.sig2,sd.inb)
```

```
ave_mean<-rowMeans(means)  ## calcualate average mean
   ave_sd  <-rowMeans(sds)
 ave_p_inb <- mean(p.inb)

return(cbind(ave_mean,ave_sd,rep(Tinb,10),rep(ave_p_inb,10)))
 }
```

D.2 Bayesian MCMC Model of the Cost-effectiveness

```
Model
{

for(i in 1 : n) {

x1[i] ~ dgamma(alpha1,taux1)
y1[i] ~ dnorm(mu1[i], tau1)
mu1[i] <- ym1 + beta1 * (x1[i] - mux1)

    x2[i] ~ dgamma(alpha2,taux2)
y2[i] ~ dnorm(mu2[i], tau2)
mu2[i] <- ym2 + beta1 * (x2[i] - mux2)

  }

ym1 ~ dnorm(0, .00001)

  ym2 ~ dnorm(0, .00001)

beta1 ~ dnorm(0, .00001)

  taux1 <- alpha1/mux1

  taux2 <- alpha2/mux2

  alpha1 ~ dunif(0.01, 500)

mux1 ~ dunif(0.01, 500)

  alpha2 ~ dunif(0.01, 500)
```

```
mux2 ~ dunif(0.01, 500)

tau1<-1/(sig1*sig1)

sig1~dunif(0.01, 500)

tau2<-1/(sig2*sig2)

sig2~dunif(0.01, 500)

  inb <- lambda * (ym1 - ym2) - (mux1 - mux2)

p.value<-step(inb)
}
```

BIBLIOGRAPHY

Briggs, A. (1999), "A Bayesian approach to stochastic cost-effectiveness analysis," *Health Economics*, 8, 257–261.

Briggs, A. and Fenn, P. (1998), "Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane," *Health Economics*, 7, 723–740.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995), "A Limited Memory Algorithm for Bound Constrained Optimization," *SIAM Journal on Scientific and Statistical Computing*, 16(5), 119–1208.

Carlin, B. P. and Louis, T. A. (2008), *Bayesian Methods for Data Analysis*, CRC Press.

Cheng, D., Stamey, J. D., and Branscum, A. J. (2009), "Bayesian approach to average power calculations for binary regression models with misclassified outcomes," *Statistics in Medicine*.

Corrao, G., Nicotra, F., Parodi, A., Zambon, A., Soranna, D., Heiman, F., Merlino, L., and Mancia, G. (2012), "External adjustment for unmeasured confounders improved drug-outcome association estimates based on health care utilization data," *Journal of Clinical Epidemiology*, Nov;65(11), 1190–9.

Faries, D., Peng, X., Pawaskar, M., Price, K., Jr, J. W. S., and Stamey, J. D. (2013), "Evaluating the Impact of Unmeasured Confounding with Internal Validation Data: An Example Cost Evaluation in Type 2 Diabetes," *VALUE IN HEALTH*, 16, 259–266.

Grieve, R., Nixon, R., and Thompson, S. (2010), "Bayesian hierarchical models for cost-effectiveness analyses that use data from cluster randomized trials," *Medical Decision Making*, 30(2), 163–75.

Gustafson, P. and McCandless, L. (2010), "Probabilistic approaches to better quantifying the results of epidemiological studies," *International Journal of Environmental Research and Public Health*, 7, 1520–1539.

Heitjan, D., Moskowitz, A., and Whang, W. (1999), "Bayesian estimation of cost-effectiveness ratios from clinical trials," *Health Economics*, 8, 191–201.

Laska, E., Meisner, M., and Siegel, C. (1997), "Statistical inference for cost-effectiveness ratios," *Health Economics*, 6, 229–242.

Manning, W. G., Basu, A., and Mullahy, J. (2002), *Modeling Costs with Generalized Gamma Regression*.

— (2005), "Generalized modeling approaches to risk adjustment of skewed outcomes data," *Journal of Health Economics*, 24, 465–488.

McCandless, L. C., Gustafson, P., and Levy, A. (2007a), "Bayesian sensitivity analysis for unmeasured confounding in observational studies," *Statistics in Medicine*.

— (2007b), "Bayesian sensitivity analysis for unmeasured confounding in observational studies," *STATISTICS IN MEDICINE*, 26, 2331–2347.

McInturff, P., Johnson, W., Gardner, I. A., and Cowling, D. (2004), "Modelling risk when binary outcomes are subject to error," *STATISTICS IN MEDICINE*, 23(7), 1095–109.

Morteza Khodabina, A. A. (2010), "Some properties of generalized gamma distribution," *Mathematical Sciences*, 4, 9–28.

Negrin, M., Vozquez-Polo, F., Martel, M., Moreno, E., and Girn, F. (2010), "Bayesian variable selection in cost-effectiveness analysis," *International Journal of Environmental Research and Public Health*, 7(4), 1577–96.

O'Brien, B. (1998), "Willingness-to-pay in cost-benefit analysis of health care programs," *Encyclopedia of Biostatistics*, 6, 4750–4755.

O'Brien, B. and Gafni, A. (1996), "When do the 'dollars' make sense? Toward a conceptual framework for contingent evaluation studies in health care," *Medical Decision Making*, 16, 288–299.

O'Hagan, A., Stevens, J., and Montmartin, J. (2000), "Inference for the C/E Acceptability Curve and C/E Ratio," *PharmacoEconomics*, 17, 339–349.

O'Hagan, A., Stevens, J. W., and Montmartin, J. (2001), "Bayesian cost-effectiveness analysis from clinical trial data," *Statistics in Medicine*, 20, 733–753.

Steyerberg, E. W., Bleeker, S. E., Moll, H. A., Grobbee, D. E., and Moons, K. G. (2003), "Internal and external validation of predictive models: A simulation study of bias and precision in small samples," *Journal of Clinical Epidemiology*, Volume 56, Issue 5, 441447.

Stinnett, A. and Mulahy, J. (1998.), "Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis," *Medical Decision Making*, 18, S68–S80.

Thompson, S. G. and Nixon, R. M. (2005), "How Sensitive Are Cost-Effectiveness Analyses to Choice of Parametric Distributions?" *Medical Decision Making*, 25, 416.

Vandenbroucke, J. P. (2002), "The history of confounding," *Sozial- und Prventivmedizin*, 47, 216–224.

Willan, A. and O'Brien, B. (1996), "Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem," *Health Economics*, 5, 297–305.

Willan, A. R. and Briggs, A. H. (2006), *Statistical Analysis of Cost-Effectiveness Data*, John Wiley and Sons.

Willan, A. R. and Lin, D. Y. (2001), "Incremental net benefit in randomized clinical trials," *STATISTICS IN MEDICINE*, 20, 1563–1574.

Zethraeus, M. T. N. and Johannesson, M. (1998), "A note on confidence intervals in cost-effectiveness analysis," *International Journal of Technology Assessment in Health Care*, 14, 467–471.