ABSTRACT

Bayesian Approach to Partially Validated Binary Regression with Response and Exposure Misclassification

Katrina Anderson, Ph.D.

Chairperson: James D. Stamey, Ph.D.

Misclassification of epidemiological and observational data is a problem that commonly arises and can have adverse ramifications on the validity of results if not properly handled. Considerable research has been conducted when only the response or only the exposure are misclassified, while less work has been done on the simultaneous case. We extend previous frequentist work by investigating a Bayesian approach to dependent, differential misclassification models. Using a logit model with misclassified binary response and exposure variables and assuming a validation sub-sample is available, we compare the resulting confidence and credible intervals under the two paradigms. We compare the results under varying percentages of validation subsamples, 100% (ideal scenario), 25%, 15%, 10%, 5%, 2.5%, and 0% (naive scenario) of the overall sample size. We extend this work further be examining scenarios for which the assumptions may falter; we assume independent, differential misclassification, increase the overall sample size, and vary the influence of our priors from diffuse to concentrated.

Finally, we examine the scenario in which the response variable is correlated over time and differentially misclassified. We compare four different models: a model that assumes the differential misclassification is correlated within subjects and correlated with the response model; a model that assumes independent response processes but with differential misclassification having correlation within the subject; a model with independent response processes and differential misclassification that is uncorrelated within the subject; and, lastly, a model that assumes independent response processes and that the non-differential misclassification is uncorrelated within the subject. We present the Bayesian approach, in addition to the previous frequentist work, to each of these models. We compare not only the two approaches via estimation bias and precision, but also the ability of each approach to select the "correct" model (assuming differential misclassification process is correlated with the response model and within subject).

Bayesian Approach to Partially Validated Binary Regression with Response and Exposure Misclassification

by

Katrina Anderson, B.S., M.S.

A Dissertation

Approved by the Department of Statistical Science

_____

Jack 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

_____

Jack D. Tubbs, Ph.D.

_____

Dean M. Young, Ph.D.

_____

Phil D. Young, Ph.D.

Accepted by the Graduate School
August 2018

_____

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

I would like to thank my committee for providing thoughtful improvements for this dissertation. I appreciate the time that each of you has taken to help me through this process. I want to thank the entire department for welcoming me and encouraging me to finish my Ph.D. Dr. Young, I would like to thank you for not only providing for opportunities of intellectual rigor, but for also allowing fun to resonate through the office; your comedic relief was greatly appreciated.

I would like to thank Dr. Stamey for the never failing, and never-ending support you offered over the years. You have been a great role-model and a constant mentor to me during this time in my career. I appreciate how easily you seemed to roll with the punches when we expanded our family, twice, during my time at Baylor. I appreciate the flexibility you offered me and the structure you forced, when needed.

I would also like to thank the students who labored through classes, exams, and presentations with me. Your kindness to help prepare me for prelims and your friendship to help get us through courses, will be cherished. I thank Chris and Morgan for helping to instill fun during all of our time at Baylor. I will forever be grateful for Clay's ability to lighten the mood, keep things happy, and for his unwavering example of what a good parent looks like. I am indebted to Sarah, who not only helped me get through coursework but helped me to find the perfect testing spot for prelims, finish a half marathon (the toughest in Texas), re-connect all of my remote computers, and listen to my struggles these past two years while finishing this remotely. I came to Baylor to learn and I leave not only with a great education, but with life-long friends and supportive colleagues.

Lastly, but most importantly, I would like to thank my family. To my wonderful daughters, Maddelyn and Abigail, thank you for motivating me to complete this. To my in-laws, I am grateful for the support you have shown me, the help with watching my girls

while I work, and the love you have shared with my family and I. Kandice, thank you for the uplifting texts, chances to vent, and the words of encouragement; you kept me thinking positively, feeling loved, and helped me to laugh through my struggles. Mom, thanks for always being ready for a phone call, helping to fly me all over the U.S. when I was home sick or needed a break, and helping support our family during this endeavor. Obviously, I thank you also for the years of support getting me through all of my schooling and forcing me back on track when I drifted off course. You were right, one day, I would thank you for all of the tough love. Thank you, so much. Mark, thank you for being right by my side these past couple of years, encouraging me to apply to schools, spending the first (ish) year of our marriage apart, helping me move states even though I was hangry, giving me lots of snacks during two pregnancies, and constantly encouraging me, even when I made it difficult! You have helped me immeasurably and I won't be able to thank you enough, but I will try. Also, Modulus.

To those I have mentioned, and those I have not, I thank you all for your support. These years have been difficult, but I am so blessed to have had the opportunity to experience them. I could not have done so without you all. Thank you.

DEDICATION

For my Dad

When good men die, their goodness does not perish. -Euripides

CHAPTER ONE

Introduction

Data from diagnostic tests are often subject to misclassification. Suppose we want to screen a large number of patients for a disease, but using a "gold-standard" (i.e. highly accurate) test is prohibitively expensive in either time, labor, or cost. Suppose also we have an error prone test that while not accurate can be readily given due to it's ease of use, low cost, or lack of time required to give it. Misclassification arises when the error-prone test gives a different result than the gold-standard test, which is assumed to be the truth. To understand the relationship between the error prone test and the gold-standard test, we validate a small subsample of our data. Thus all participants are tested with the error prone instrument, while only a faction of the sample is also given the gold-standard test. In this dissertation, a Bayesian approach to misclassification with partially validated data is developed and compared to a frequentist approach.

1.1 Motivating Examples

Specifically, two distinct scenarios involving misclassification with partially validated data are examined; the first, studies the relationship when both the response and the exposure variable are misclassified, while the second studies the relationship when the response is not only misclassified but is measured repeatedly, and thus correlated by time.

1.1.1 Misclassified Binary Response and Exposure Variables

Suppose that in addition to our response we also have a variable (referred to as the exposure) that is related to the disease and is also very expensive to measure. Like our response, this variable also has two tests, an error prone and gold-standard test. In this setting, both our exposure and response variable are binary. Our validated patients will

receive both tests, for both variables while our main data will only receive the error prone tests for both variables.

The logistic regression that uses only the error prone measurements is known to be biased. It is an open question whether a Bayesian approach may perform better than a frequentist approach. Chapter Two will further present this scenario and explore the differences between the two approaches; Chapter Three will further expand upon the differences between the two approaches.

1.1.2 Misclassified Response in Longitudinal Data

Suppose that the tests for our response can be given at multiple doctor's visits. More specifically we follow patients over J visits and test them for a disease using a readily available, but error-prone test for the disease. For a small percentage of the patients we "validate" their error-prone test with a more expensive and error free, gold-standard test. Since each of the patients is tested each visit, there is correlation between visits.

In Chapter Four we develop methods to adequately predict for the presence of the disease dealing with the misclassification and the correlation between visits.

1.2 Literature Review

A common problem in inference for observational studies is the existence of errors. For discrete variables, misclassification, in which the observed value is different than the true value of the variable is often encountered. Errors of this kind are widespread and can lead to biased and inefficient results; refer to Barron (1977), Copeland et al. (1977), Neuhaus (1999), and Carroll et al. (2006) for discussions of such problems. We can see misclassification in data pertaining to engineering, business, and health sciences. In the medical field misclassification is often seen in the diagnosis of disease; for example, a patient is given a positive test result for a disease, however the patient does not in fact have the disease. Several researchers have examined this phenomenon as it relates to various diseases, including Tang et al. (2015a), Tomiyama et al. (2016) and Gosling and Saloniki

(2014). The two examples outlined in Section 1.1 give possible scenarios explaining binary variable misclassification and the use of validated data.

Much research has been done on misclassification of 2×2 contingency tables in the event of binary exposure misclassification. Using frequentist methods, several methods have been proposed; Barron (1977) developed a matrix method (using known sensitivity and specificity), Greenland (1988) uses an inverse variance weighted estimate, and Marshall (1990) developed the inverse matrix method (using the probabilities of positive and negative predicted values). All of these methods have attempted to circumnavigate the use of maximum likelihood estimation in the hopes of developing a less computationally intense estimation method that was more intuitive. The maximum likelihood estimates (MLE) are laborious and computationally intensive to find. Lyles, (2002) however, used a re-parameterization of the likelihood to show that the estimates from the inverse matrix method is the same as the MLE's for the differential misclassification case. Due to symmetry in the odds ratio, each of these methods can be generalized to the case of binary response misclassification, for which logistic regression is a typical analytic tool used for prediction.

Magder and Hughes (1997) used an expectation-maximization (EM) algorithm for MLE's that incorporates diagnostic error rates on the parameters of a logistic regression. By utilizing a validation subsample Lyles et al. (2011) developed methods to use the EM algorithm for the case when there are unknown diagnostic error rates for differential misclassification. Edwards et al. (2013) used internal validation data as well, however they used this information to develop a multiple imputation approach instead of the EM algorithm.

In order to fully account for uncertainty in parameter estimates, research has been done on the utilization of validation data instead of relying solely on assumptions about the probabilities of false positives and false negatives. Carroll et al. (2006) used likelihood-based methods to incorporate validation data in estimating regression parameters when differential misclassification exists in the response variable. Pepe (1992) used non-parametric

kernel methods, while Holcroft et al. (1997) used inverse probability weighting. Lyles et al. (2011) developed accessible maximum likelihood methods using optimization tools through SAS® software to incorporate internal validation data in the estimation of regression parameters when the response is differentially misclassified. Bayesian methods were used by Paulino et al. (2003) to incorporate prior expert opinions for the same scenario. McInturff et al. (2004) explore the use of conditional means priors while Gerlach and Stamey (2007) explore methods for variable selection using Bayesian methods with a misclassified response.

In this dissertation we use simulation studies based upon the real-world data found in Smith et al. (1997) regarding the presence of bacterial vaginosis in women. Frequentist methods related to the two examples discussed in Section 1.1 were originally discussed by Tang et al. (2015a, 2013) Their work did not establish methods under the Bayesian approach; they also did not examine how the methods perform when the validation sample size is small.

1.3 Plan of the Dissertation

In Chapter Two we look at the different types of misclassification that can affect binary logistic regression as discussed in a frequentist framework by Tang et al. (2015a) Using the HIV Epidemiology Research Study (HERS) data as a template for a simulation study, we examine the differences between the Bayesian and frequentist approaches for logistic regression with dependent differential misclassification on both the response and exposure variables (Smith et al. 1997). In Chapter Three we investigate the performance of the methods developed in Chapter Two under a variety of scenarios. We aim to understand how the results are affected by varying the assumptions from dependent to independent differential misclassification, increasing the overall sample size, and adjusting the information in the prior distributions of our parameters under the Bayesian approach.

In Chapter Four we examine the setting in which the response variable is correlated over time; i.e. we have multiple visits from the same patient and measure the response each time the patient visits. We will explore the effect of these correlated responses with partially validated data and dependent errors using both frequentist (refer also to Tang et al. (2013)) and Bayesian methods. We will perform several analyses using varying assumptions of the model at hand and use model selection criteria to ensure the "correct" model is selected under each approach.

CHAPTER TWO

Bayesian Misclassification with Partially Validated Data

2.1 Introduction to Misclassification and Literature Review

Reliably estimating a relationship between an exposure on an outcome is a primary goal for many observational studies. For epidemiological studies, this goal is typically focused on variables describing the presence of a certain health trait or not. These variables are binary in nature and are often times hindered by measurement errors of some kind (e.g.: imprecise testing methods). For example, suppose we have a readily available diagnostic screen for a certain disease but this screening tool is error-prone. The use of this screen may produce false negative and false positive results for certain disease. This error in measuring for the disease is called misclassification and is more broadly defined as the measurement error stemming from categorical variables.

Several researchers have developed methods for correcting for misclassification in either the response or exposure variable. Magder and Hughes (1997) looked at misclassification in the response variable using logistic regression; Lash and Fink (2003) used a sensitivity analysis to correct for misclassification in an exposure variable. Fox et al. (2005) extended this work to included ranges of possible values for the sensitivity and specificity. Lyles and Lin (2010) used a sensitivity analysis in terms of response misclassification; they also studied a separate sensitivity analysis approach to correct for exposure misclassification that utilized expert opinion to develop weights in fitting their models. While this research is instructive on the needs for accounting for misclassification, none of these use validation data to further refine the misclassification.

Internal validation data can provide a wealth of information for variable misclassification. In terms of differential and non-differential exposure misclassification, Marshall (1990) and Greenland (1988) explored variance estimators for the effects estimates for a

couple of methods. Lyles (2002), Greenland (2008) and Morrissey and Spiegelman (1999) all discuss the merits of validation data on finding likelihood-based estimates when misclassification exists. Tang et al. (2015a) tie in the methodology to provide practical analytic solutions under the frequentist framework for the scenario when both the response and exposure are subject to misclassification.

Considerable work from the Bayesian perspective has been done on misclassification and measurement error models. Paulino et al. (2003) and Bedrick et al. (1996) investigated methods to allow for model selection when the response variable for a binomial regression is subject to an unconstrained misclassification process. Using assumptions on the sensitivity and specificity rather than a validation subsample, Goldstein et al. (2016) use a case-control study to account for non-differential misclassification. McInturff et al. (2004) developed methods for binomial regression with response misclassification that utilize conditional means priors to aid in estimates of diagnostic sensitivity and specificity. Closely related to our research, Gerlach and Stamey (2007) explore the effects of both differential and non-differential misclassification in the context of logistic regression when the response is misclassified using internal validation data. All of this research focused solely on methods for adjusting for response or exposure misclassification, not on the methods to adjust for both.

Prior research often focuses on contingency tables when the response and exposure are both misclassified. Since regression is a much more widely used approach to developing relationships between variables we will focus our research on these methods. We not only use the validation designs and likelihood methods proposed by Tang et al. (2015a), but we expand upon their methods to the Bayesian approach. We also test each approach under varying validation sizes to understand which approach prevails after adjustments.

This chapter begins with background information regarding misclassification and the key terminology when discussing models with this type of error. We then discuss three types of misclassification focusing mainly on dependent differential misclassification. We

first focus on the frequentist approach and then delve into the Bayesian approach, complete with a discussion on priors. The comparison of the two approaches is developed from a simulation study based off of the HERS dataset. We discuss the convergence of both approaches and then compare the results of each approach.

2.2 Background: 3 Types of Misclassification

Suppose that we are interested in a binary response, Y, and we have p covariates with one exposure, X, that is of interest. For this, we assume that:

$$Y \sim Bern(P(Y = 1|X, C_1, C_2, \dots C_P)]), \tag{2.1}$$

where we often use the logit model for binary response data,

$$\text{logit}[P(Y = 1|X, C_1, C_2, \dots C_P)] = \beta_0 + \beta_1 X + \sum_{p=2}^{P+1} \beta_p C_p. \tag{2.2}$$

Note that in this case both X and Y are gold-standard measurements, meaning they are highly expensive to measure. In order to not spend inordinate amounts of time, cost, or energy these results are only available for a small sub-sample of our subjects. For every subject however, we have X^* and Y^*, the error prone analog measurements for X and Y, respectively. These measurements are readily available, cheaper, or easier to assess than their respective gold-standard measurements. A naïve analysis would simply replace X and Y, the gold-standard measurements, with X^* and Y^*, the error prone measurements, in equation 2.2 since these are available for all subjects. This gives the following model:

$$\text{logit}[P(Y^* = 1|X^*, C_1, C_2, \dots C_P)] = \beta_0^* + \beta_1^* X^* + \sum_{p=2}^{P+1} \beta_p^* C_p, \tag{2.3}$$

which then relates to the assumed distribution of the error-prone measurements; specifically, $Y^* \sim Bern(P(Y^*|X^*, C_1, C_2, \dots C_P))$. Estimates of the β_j^* under this model are biased for the parameters of interest, the β_j.

The naïve model fails to incorporate the information from those subjects with both the error prone and gold-standard measurements. This data, the validated data, offers an

important piece of information, the relationship between the error prone and gold-standard measurements. We must account for this information if it is available otherwise our estimates would be unnecessarily more biased. We can incorporate this data in our models to help develop the relationship for the remainder of our sample, the main data, that only has the error prone measurements.

In the measurement error literature (e.g. Richardson and Gilks (1993)), three models are developed to explore relationships: the response model, exposure model, and the measurement model(s). We have already seen our response model under the ideal scenario:

$$\text{logit}[P(Y = 1|X, C_1, C_2, \dots C_P)] = \beta_0 + \beta_1 X + \sum_{p=2}^{P+1} \beta_p C_p. \tag{2.4}$$

Assuming $X \sim Bern(P(X = 1|C_1, C_2, \dots C_P)$, we use another model for our exposure's relationship to the covariates, called the exposure model:

$$\text{logit}[P(X = 1|C_1, C_2, \dots C_Q)] = \gamma_0 + \sum_{q=1}^{Q} \gamma_q C_q. \tag{2.5}$$

Finally, our measurement models are used to incorporate our error prone response and exposure variables, $y^* \sim Bern(P(y^*|y))$ and $x^* \sim Bern(P(x^*|x))$, respectively. Depending on the assumptions relating to the misclassification scheme, the underlying probability models range from simple to quite complex. The model assumptions we will cover include: independent non-differential misclassification, dependent non-differential misclassification, and dependent differential misclassification.

2.2.1 Independent Non-Differential Misclassification

Independent non-differential misclassification has two main assumptions. The independence implies that the misclassification patterns of X and Y are independent. This means that the pattern by which the response is missing does not depend on the pattern for which the exposure is missing. The second assumption is that the sensitivity, probability of a true positive, and specificity, probability of a true negative, are roughly constant across all values of the variables. This is referred to as "non-differential" misclassification.

To incorporate these two assumptions, we will keep our response and exposure models as:

$$\text{logit}[P(Y = 1|X, C_1, C_2, \ldots C_P)] = \beta_0 + \beta_1 X + \sum_{p=2}^{P+1} \beta_p C_p \tag{2.6}$$

$$\text{logit}[P(X = 1|C_1, C_2, \ldots C_Q)] = \gamma_0 + \sum_{q=1}^{Q} \gamma_q C_q. \tag{2.7}$$

However, to incorporate the misclassification information, we could utilize the following relations involving the sensitivity and specificity of both X and Y to build our measurement models:

$$y^*|y \sim \text{Bern}(yS_y + (1-y)(1-C_y)) \tag{2.8}$$

$$x^*|x \sim \text{Bern}(xS_x + (1-x)(1-C_x)). \tag{2.9}$$

Equation 2.8 yields the following likelihood contribution:

$$P(y^*|y) = (yS_y + (1-y)(1-C_y))^{y^*}(1-(yS_y + (1-y)(1-C_y)))^{1-y^*}, \tag{2.10}$$

which models the sensitivity of Y when $y = 1$ and the specificity of Y when $y = 0$. The same relationship can be seen with equation 2.9 and the sensitivity and specificity of X.

All four of these models are used to calculate the joint probability, which can be broken down as:

$$\begin{aligned} &P(X^* = x^*, Y^* = y^*|X = x, Y = y, C = c) \\ &\qquad = \sum_{y=0}^{y=1} \sum_{x=0}^{x=1} P(y^*|y)P(x^*|x)P(y|x,c)P(x|c), \end{aligned} \tag{2.11}$$

where the four pieces on the right-hand side of the equation represent vital pieces of information from our data (response model, exposure model, and measurement models). The joint likelihood can thus be written as the product of the following two components:

$$L_m = \prod_{i=1}^{n_m} \left\{ \sum_{y_i=0}^{y_i=1} \sum_{x_i=0}^{x_i=1} P(y_i^*|y_i)P(x_i^*|x_i)P(y_i|x_i, c_{iy})P(x_i|c_{ix}) \right\} \tag{2.12}$$

$$L_v = \prod_{j=1}^{n_v} \left\{ P(y_j^*|y_j) P(x_j^*|x_j) P(y_j|x_j, c_{jy}) P(x_j|c_{jx}) \right\}. \tag{2.13}$$

L_m (2.12) is the likelihood component for those observations in the "main" study; notice that we sum over all possible combinations of X and Y, since they are not observed. L_v (2.13) is the likelihood component for those observations in the "validation" study, the observations where both the gold-standard and error prone measurements are known, and thus do not need to be summed over because the true statuses are known.

2.2.2 Independent Differential Misclassification

Independent differential misclassification is an extension of independent non-differential misclassification. We still assume the error prone measurements are independent but now we assume that the sensitivity and/or specificity of X (Y) depends on Y (X) and the covariates. Under this model, our likelihood components become:

$$L_m = \prod_{i=1}^{n_m} \left\{ \sum_{y_i=0}^{y_i=1} \sum_{x_i=0}^{x_i=1} P(y_i^*|y_i, x_i, c_{iy^*}) P(x_i^*|x_i, y_i, c_{ix^*}) P(y_i|x_i, c_{iy}) P(x_i|c_{ix}) \right\} \tag{2.14}$$

$$L_v = \prod_{j=1}^{n_v} \left\{ P(y_j^*|y_j, x_j, c_{jy^*}) P(x_j^*|x_j, y_j, c_{jx^*}) P(y_j|x_j, c_{jy}) P(x_j|c_{jx}) \right\}. \tag{2.15}$$

The joint likelihood for this misclassification model is the product of these two components where L_m (2.14) is the likelihood component for those observations in the "main" study while L_v (2.15) is the likelihood component for those observations in the "validation" study, the observations where both the gold-standard and error prone measurements are known.

The independent differential misclassification model now has the following four logistic regression models rather than the two logistic regressions used for independent non-differential misclassification model (equations 2.6 and 2.7):

$$\text{logit}[P(Y = 1|X, C_1, C_2, \ldots C_P)] = \beta_0 + \beta_1 X + \sum_{p=2}^{P+1} \beta_p C_p \tag{2.16}$$

$$\text{logit}[P(X = 1|C_1, C_2, \ldots C_Q)] = \gamma_0 + \sum_{q=1}^{Q} \gamma_q C_q \tag{2.17}$$

$$\text{logit}[P(Y^* = 1|Y, X, X^*, C_1, C_2, \ldots C_R)] = \theta_0 + \theta_1 X + \theta_2 Y + \sum_{r=3}^{R+2} \theta_r C_r \tag{2.18}$$

$$\text{logit}[P(X^* = 1|X, Y, C_1, C_2, \ldots C_S)] = \delta_0 + \delta_1 X + \delta_2 Y + \sum_{s=3}^{S+2} \delta_s C_s. \tag{2.19}$$

Equation 2.16 and 2.18 are our original response and exposure models, respectively, while equations 2.18 and 2.19 are our updated measurement models. These two new logistic regressions are required to allow the sensitivity and specificity to vary with the covariates. This model is now significantly more complex than the previous models because differential misclassification adds several parameters. As the number of parameters increases, their values become harder to estimate and will typically require a larger validation sample size to estimate accurately. In fact, for small validation samples, it is not unusual to for both the frequentist and Bayesian approaches to have issues with convergence.

2.2.3 Dependent Differential Misclassification

The last type of misclassification model we will consider is dependent differential misclassification. Here, the assumption of independence is now relaxed to include a possibility of dependence in the error-prone to gold-standard measurements. As with independent differential misclassification models we still have the assumption that the sensitivity and/or specificity of X (Y) depends on Y (X) and the covariates.

The dependent misclassification model is determined by the following four logistic regressions:

$$\text{logit}[P(Y = 1|X, C_1, \ldots C_P)] = \beta_0 + \beta_1 X + \sum_{p=2}^{P+1} \beta_p C_p \tag{2.20}$$

$$\text{logit}[P(X = 1|C_1, \ldots C_Q)] = \gamma_0 + \sum_{q=1}^{Q} \gamma_q C_q \tag{2.21}$$

$$\text{logit}[P(Y^* = 1|Y, X, X^*, C_1, \ldots C_R)] = \theta_0 + \theta_1 X + \theta_2 X^* + \theta_3 Y + \sum_{r=4}^{R+3} \theta_r C_r \tag{2.22}$$

$$\text{logit}[P(X^* = 1|X, Y, C_1, \ldots C_S)] = \delta_0 + \delta_1 X + \delta_2 Y + \sum_{s=3}^{S+2} \delta_s C_s \tag{2.23}$$

where the added parameter θ_2 in equation 2.22 reflects the dependence between X^* and Y^*. Again, this effect is clearly shown when comparing the measurement models between the independent differential misclassification model (2.18) and the dependent differential misclassification model (2.22).

The joint likelihood for this model is proportional to the product of the following two components:

$$L_m = \prod_{i=1}^{n_m} \left\{ \sum_{y_i=0}^{y_i=1} \sum_{x_i=0}^{x_i=1} P(y_i^*|y_i, x_i, x_i^*, c_{iy^*}) P(x_i^*|x_i, y_i, c_{ix^*}) P(y_i|x_i, c_{iy}) P(x_i|c_{ix}) \right\} \quad (2.24)$$

$$L_v = \prod_{j=1}^{n_v} \left\{ P(y_j^*|y_j, x_j, x_j^*, c_{jy^*}) P(x_j^*|x_j, y_j, c_{jx^*}) P(y_j|x_j, c_{jy}) P(x_j|c_{jx}) \right\}. \quad (2.25)$$

More specifically, our likelihood is:

$$L \propto L_m^{(1-val)} L_v^{val}, \quad (2.26)$$

where the components of the likelihood are turned “on" or “off" depending on whether the subject has been validated or not.

Chapter Three will discuss the differences between independent and dependent differential misclassification in greater detail and, more specifically, investigating the added information that θ_2 gives to the misclassification model.

2.3 Bayesian Approach to Frequentist Work

For our context, we have the following formulas to work with:

$$\text{logit}[P(Y = 1|X, C_1, \ldots C_P)] = \beta_0 + \beta_1 X + \sum_{p=2}^{P+1} \beta_p C_p \quad (2.20)$$

$$\text{logit}[P(X = 1|C_1, \ldots C_Q)] = \gamma_0 + \sum_{q=1}^{Q} \gamma_q C_q \quad (2.21)$$

$$\text{logit}[P(Y^* = 1|Y, X, X^*, C_1, \ldots C_R)] = \theta_0 + \theta_1 X + \theta_2 X^* + \theta_3 Y + \sum_{r=4}^{R+3} \theta_r C_r \quad (2.22)$$

$$\text{logit}[P(X^* = 1|X, Y, C_1, \ldots C_S)] = \delta_0 + \delta_1 X + \delta_2 Y + \sum_{s=3}^{S+2} \delta_s C_s. \quad (2.23)$$

We are dealing with logistic regressions for which the mean of the responses, when appropriately transformed, will hover between zero and one. A one-unit change in any of the coefficients for the explanatory variables would rarely result in an increase in the predictive probability from 0.01 to 0.99. Likewise, prior to seeing any data, we believe we would rarely estimate a coefficient to which the resulting predicted probability of testing positive in the response would increase from 1% to 99% simply by changing one input variable, for example increasing in age by one year. This change would, when back transformed to the scale of the coefficients, result in a coefficient with a magnitude of 10. In lieu of expert opinion, this belief in our coefficients allows us to formulate our prior distributions on the parameters. We will explore the impacts of prior information in Chapter Three.

For now, we set relatively uninformative priors on the parameters of interest so that they are centered about zero with moderately large variance. For this, we will use:

$$\beta_0 \dots \beta_{P+1} \sim Normal(0, \sigma^2 = 10) \tag{2.27}$$

$$\gamma_0 \dots \gamma_{Q+1} \sim Normal(0, \sigma^2 = 10) \tag{2.28}$$

$$\theta_0 \dots \theta_{R+1} \sim Normal(0, \sigma^2 = 10) \tag{2.29}$$

$$\delta_0 \dots \delta_{S+1} \sim Normal(0, \sigma^2 = 10) \tag{2.30}$$

This distribution is given in Figure 2.3.1 in which you can see that most of the values will hover around 0, but that we still allow for the relatively unlikely values of 10 and -10. This prior is placed on each of the parameter values, allowing us to have a very broad belief in our parameter values. This prior doesn't constrain the posterior probabilities too much but does give exceptionally large values low probability. These priors allow for the very unlikely example that we discussed before as well as the more likely and smaller percentile jumps associated with values towards zero.

2.4 Example Based Upon HERS Dataset

Our simulation study is motivated by Tang et al. (2015a) The dataset utilized was from the HIV Epidemiology Research Study based on a sample of 904 women. The study

aim was to predict the occurrence of bacterial vaginosis (BV), a sexually transmitted infection, from the presence of trichomoniasis (TRICH), another type of sexually transmitted infection, and other variables, such as age and race. In order to replicate this example, we will use the same general set-up of their data to develop a model for our simulation study. The gold-standard response is a laboratory-based method for testing of BV while the error prone measurement is a clinically based test for BV. For the exposure, the gold-standard predictor is a culture test for TRICH while the error prone measurement is called a wet mount test for TRICH. The covariates of interest are: age (with a median of 37 years), race (identified as “black" or not), HIV risk cohort (intravenous drug transmission versus sexually transmitted), and HIV status (positive or negative). The authors had validation data (where the gold-standard and error prone measurements are known) for 214 of the 904 women; we will use varying validation sizes for our study, as discussed in sections 2.5 and 2.6.

Based on model selection criteria presented by Tang et al. (2015a) and using the dependent differential misclassification model proposed by them, the following models and predictors were used to generate the data. The models for the validation data are:

$$\begin{aligned}\text{logit}[P(\text{BV}^* = 1|\ldots)] = \theta_0 &+ \theta_1\text{Trich} + \theta_2\text{Trich}^* + \theta_3\text{BV} \\ &+ \theta_4\text{RiskCohort} + \theta_5\text{HIVStatus}\end{aligned}$$

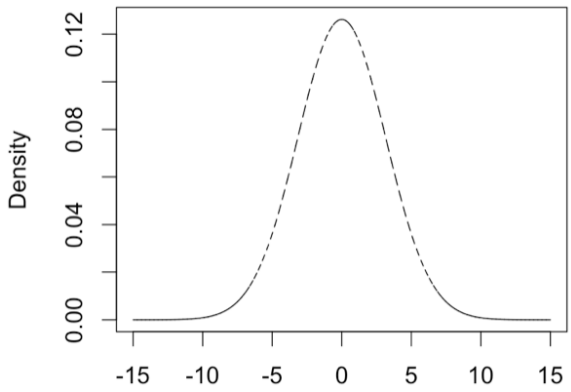

Figure 2.3.1. Prior Distribution for Bayesian Parameter Estimation using $N(\mu = 0, \sigma^2 = 10)$

$$\begin{aligned}
\text{logit}[P(\text{Trich}^* = 1|\ldots)] &= \delta_0 + \delta_1\text{Trich} + \delta_2\text{BV} + \delta_3\text{RiskCohort} \\
\text{logit}[P(\text{BV} = 1|\ldots)] &= \beta_0 + \beta_1\text{Trich} + \beta_2\text{Age} + \beta_3\text{Race} \\
&\quad + \beta_4\text{RiskCohort} + \beta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich} = 1|\ldots)] &= \gamma_0 + \gamma_1\text{Race} \qquad (2.31)
\end{aligned}$$

These will each be utilized to develop the validation data component to the likelihood by using: The joint likelihood for this model is proportional to the product of the following two components:

$$\begin{aligned}
L_v = \prod_{j=1}^{n_v} \Big\{ & P(\text{BV}^* = 1|\ldots)P(\text{Trich}^* = 1|\ldots) \\
& P(\text{BV} = 1|\ldots)P(\text{Trich} = 1|\ldots) \Big\}. \qquad (2.32)
\end{aligned}$$

Recall, however, that both BV and TRICH are the gold-standard, unobserved measurements for the main study participants thus the likelihood component relating to these individuals is as follows:

$$\begin{aligned}
L_m = \prod_{i=1}^{n_m} \Big\{ & \sum_{y_i=0}^{y_i=1} \sum_{x_i=0}^{x_i=1} P(\text{BV}^* = 1|\ldots)P(\text{Trich}^* = 1|\ldots) \\
& P(\text{BV} = 1|\ldots)P(\text{Trich} = 1|\ldots) \Big\} \qquad (2.33)
\end{aligned}$$

For this calculation of the likelihood, we sum over the multiplicative contribution of each case depending on the error prone response:

(1) Both BV and TRICH set to a value of 1:

$$\begin{aligned}
\text{logit}[P(\text{BV}^* = 1|\ldots)] &= \theta_0 + (\theta_1) + \theta_2\text{Trich}^* + (\theta_3) \\
&\quad + \theta_4\text{RiskCohort} + \theta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich}^* = 1|\ldots)] &= \delta_0 + (\delta_1) + (\delta_2) + \delta_3\text{RiskCohort} \\
\text{logit}[P(\text{BV} = 1|\ldots)] &= \beta_0 + (\beta_1) + \beta_2\text{Age} + \beta_3\text{Race} \\
&\quad + \beta_4\text{RiskCohort} + \beta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich} = 1|\ldots)] &= \gamma_0 + \gamma_1\text{Race} \qquad (2.34)
\end{aligned}$$

(2) Both BV and TRICH are being set to a value of 0:

$$\begin{aligned}
\text{logit}[P(\text{BV}^* = 1|\ldots)] &= \theta_0 + \theta_2\text{Trich}^* \\
&\quad + \theta_4\text{RiskCohort} + \theta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich}^* = 1|\ldots)] &= \delta_0 + \delta_3\text{RiskCohort} \\
\text{logit}[P(\text{BV} = 1|\ldots)] &= \beta_0 + \beta_2\text{Age} + \beta_3\text{Race} \\
&\quad + \beta_4\text{RiskCohort} + \beta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich} = 1|\ldots)] &= \gamma_0 + \gamma_1\text{Race}
\end{aligned} \tag{2.35}$$

(3) BV is set to a value of 1 and TRICH is set to a value of 0:

$$\begin{aligned}
\text{logit}[P(\text{BV}^* = 1|\ldots)] &= \theta_0 + \theta_2\text{Trich}^* + (\theta_3) \\
&\quad + \theta_4\text{RiskCohort} + \theta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich}^* = 1|\ldots)] &= \delta_0 + (\delta_2) + \delta_3\text{RiskCohort} \\
\text{logit}[P(\text{BV} = 1|\ldots)] &= \beta_0 + \beta_2\text{Age} + \beta_3\text{Race} \\
&\quad + \beta_4\text{RiskCohort} + \beta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich} = 1|\ldots)] &= \gamma_0 + \gamma_1\text{Race}
\end{aligned} \tag{2.36}$$

(4) BV is set to a value of 0 and TRICH is set to a value of 1:

$$\begin{aligned}
\text{logit}[P(\text{BV}^* = 1|\ldots)] &= \theta_0 + (\theta_1) + \theta_2\text{Trich}^* \\
&\quad + \theta_4\text{RiskCohort} + \theta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich}^* = 1|\ldots)] &= \delta_0 + (\delta_1) + \delta_3\text{RiskCohort} \\
\text{logit}[P(\text{BV} = 1|\ldots)] &= \beta_0 + (\beta_1) + \beta_2\text{Age} + \beta_3\text{Race} \\
&\quad + \beta_4\text{RiskCohort} + \beta_5\text{HIVStatus} \\
\text{logit}[P(\text{Trich} = 1|\ldots)] &= \gamma_0 + \gamma_1\text{Race}
\end{aligned} \tag{2.37}$$

These logistic regression models are incorporated into the likelihood models as seen in equation 2.33. For a more in depth look at how this is accomplished, refer to Appendix B.1.1 and B.1.2.

2.5 Data Generation/Simulation, Convergence, and Methods of Estimation

2.5.1 Data Generation/Simulation

In order to compare the Bayesian versus frequentist models, a simulation study was performed. The overall process of simulation, based on the real-world example using the HIV Epidemiology Research Study as presented by Tang et al. (2015a), is:

(1) Set the true estimates of the parameters based on prior work. In this case we assume:

   A. A total of 904 women with complete data on BV, TRICH, and other risk factors at the fourth visit were considered.

   B. Among them, 61.7% were Black

   C. 67.4% were HIV positive

   D. 52% were intravenous drug users

   E. The median age at enrollment was 37 years.

(2) For the Bayesian approach, each of the coefficients are given Normal$(\mu = 0, \sigma^2 = 10)$ prior distributions.

(3) Generate a full dataset from these parameters using the assumptions and formulas described for dependent differential misclassification (equations 2.20 - 2.25).

(4) Estimate the parameters using each of the approaches for the current dataset that is generated.

(5) Compare the estimated parameters to the “true estimates" used to generate the data.

(6) Repeat these steps 500 times and report the appropriate summaries for each paradigm.

This process is replicated for a variety of settings to find optimal validation sizes, as well as to see the merit of the Bayesian approach. We run both paradigms for the ideal model (eq. 2.2), naive model (eq. 2.3), and various validation sizes under the complete analysis that includes misclassification. The validation sizes proposed by Tang, et al. (2015a), were 25%, 15%, 10%, and 5%. For further investigation, a validation size of 2.5% was added to the analysis plan they chose.

2.5.2 Convergence

Convergence for the frequentist results decreased as the percentage of validation data decreased: 25% validation size had 498/500 datasets that converged; 15% had 492/500 converged datasets; 10% had 480/500 converged datasets; 5% had 429/500 converged datasets; and 2.5% had 414/500 converged datasets. Datasets that did not converge were removed from the tabulation of our results.

Convergence was assessed for the Bayesian versions as well; here, all 500 iterations were used for each setting. Refer to appendix A.1 for a full display of each of the convergence diagnostic checks for each validation size. For the sake of brevity, we will discuss the convergence diagnostics for the largest and smallest validation sizes, 25% and 2.5%, both

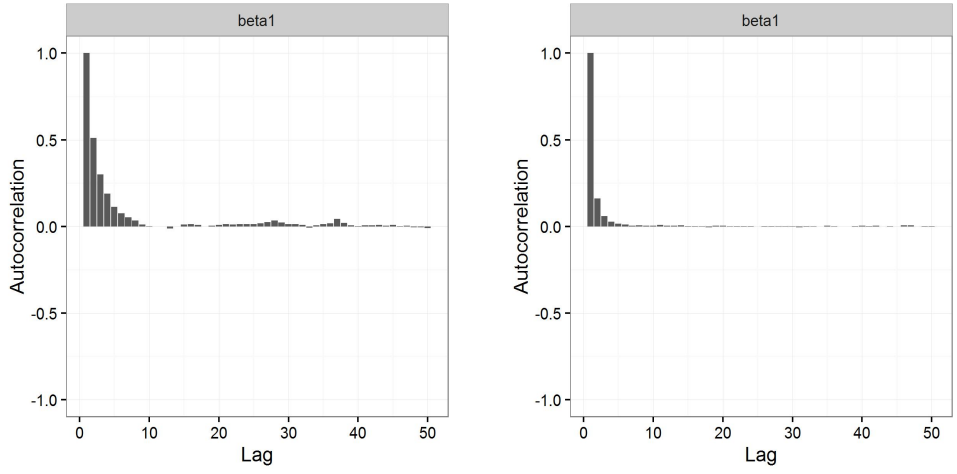

Figure 2.5.1. Auto-correlation plots for both 25% and 2.5% validation sizes

of which are satisfied. For a validation size of 25% we used an iteration size of 14,000, a burn-in of 4,000, and a thin of 1; for a validation size of 2.5% we used an iteration size of 225,000, a burn-in of 105,000, and a thin of 12. Figure 2.5.1 displays the autocorrelation plots for validation sizes; typically, you should see that as the lag increases, the autocorrelation falls to zero and remains close to zero.

Figure 2.5.2 provides the trace plots for both validation sizes; typically, you should see that as the iterations increase, there is no discernible pattern, and the values are centered

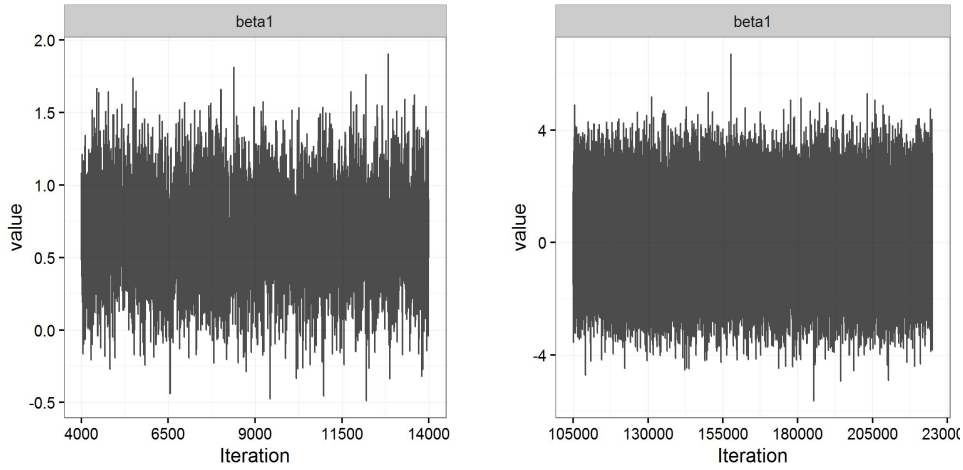

Figure 2.5.2. Trace plots for both 25% and 2.5% validation sizes

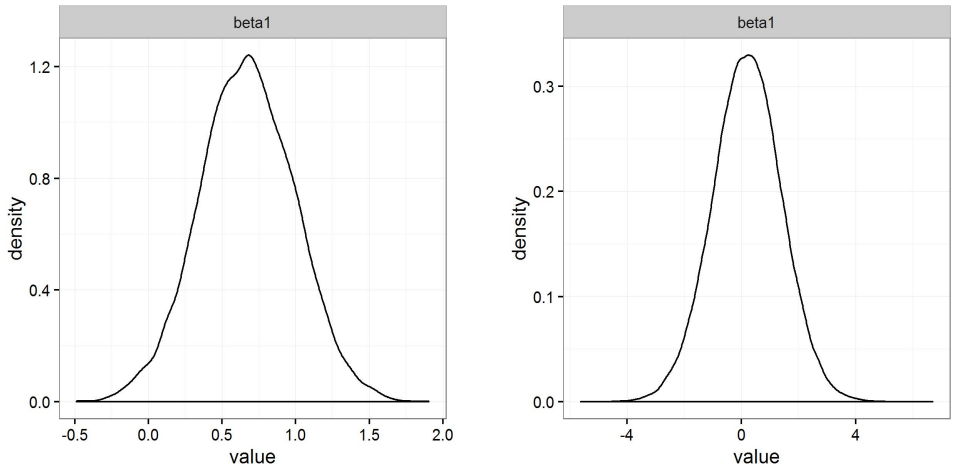

Figure 2.5.3. Density plots for both 25% and 2.5% validation sizes

appropriately. In this case, we should see that the estimates hover near $\beta_1 = 0.64$. As you can see between the two trace plots, the values show no discernible pattern, however there is much more variability in the values for the validation size of 2.5%.

Figure 2.5.3 shows the density plots for both validation sizes; we should see a smooth curve with the height of the curve at the true value of our parameter estimate. For these two validation sizes for β_1 we see that both seem to have heights near $\beta_1 = 0.64$, however there is much more variability with the smaller validation size of 2.5%.

2.5.3 Methods of Estimation

Recall that our likelihood can be specified and broken down as follows:

$$L \propto L_m^{(1-val)} L_v^{val} \tag{2.38}$$

$$L_v = \prod_{j=1}^{n_v} \left\{ P(y_j^*|y_j, x_j, x_j^*, c_{jy^*}) P(x_j^*|x_j, y_j, c_{jx^*}) P(y_j|x_j, c_{jy}) P(x_j|c_{jx}) \right\} \tag{2.39}$$

$$\begin{aligned} L_m = \prod_{i=1}^{n_m} \big\{ & P(y_i^*|y_i = 1, x_i = 1, x_i^*, c_{iy^*}) P(x_i^*|x_i = 1, y_i = 1, c_{ix^*}) \\ & P(y_i = 1|x_i = 1, c_{iy}) P(x_i = 1|c_{ix}) \\ & + P(y_i^*|y_i = 1, x_i = 0, x_i^*, c_{iy^*}) P(x_i^*|x_i = 0, y_i = 1, c_{ix^*}) \\ & P(y_i = 1|x_i = 0, c_{iy}) P(x_i = 0|c_{ix}) \\ & + P(y_i^*|y_i = 0, x_i = 1, x_i^*, c_{iy^*}) P(x_i^*|x_i = 1, y_i = 0, c_{ix^*}) \\ & P(y_i = 0|x_i = 1, c_{iy}) P(x_i = 1|c_{ix}) \\ & + P(y_i^*|y_i = 0, x_i = 0, x_i^*, c_{iy^*}) P(x_i^*|x_i = 0, y_i = 0, c_{ix^*}) \\ & P(y_i = 0|x_i = 0, c_{iy}) P(x_i = 0|c_{ix}) \big\}. \end{aligned} \tag{2.40}$$

The component for the main study data, L_m, is entirely dependent upon the values of the error prone measurements, while the validated data contributions, L_v are also dependent upon the observed gold-standard measurements. In either case, the probabilities are attained through the logistic regressions described previously (refer to Section 2.4), which makes the likelihood very complex in nature.

Using the frequentist approach, we utilize maximum likelihood optimization techniques within SAS software, specifically PROC NLMIXED to estimate the parameters. Quasi-Newtonian optimization algorithms are used as the default optimization algorithm in PROC NLMIXED. “QUANEW" is a first-derivative method in which only the gradient is computed, and second-order derivatives are approximated. The SAS code used for this simulation study is found in Appendix B.1.1.

For the Bayesian approach, the posterior distributions are found by multiplying the likelihood by the priors (Casella and Berger, 2002). The resulting marginal posteriors for all parameters of interest are not available in closed form. We use MCMC techniques to estimate the parameters with a set total number of iterations, burn-in, and thinning that has been shown to converge appropriately; packages such as OpenBUGS and JAGS can be used to fit the models. Christensen et al. (2011) explain that the idea behind MCMC techniques is to generate random sequences of vectors. This sequence can, over time, converge to a distribution called the posterior. A burn-in allows for assurance that the more volatile beginning iterations in the sequence are not skewing the resulting end of the sequence, while thinning ensures that the sequence is independent from one vector to the next. Refer to Appendix B.1.2 for the models used in this chapter for the Bayesian approach.

2.6 Comparison of Results

We now discuss the results of simulation experiments illustrating how the two procedures perform for different levels of validation data. The results of these two procedures are shown in Table 2.1.

We first consider the “naïve” model, that is, the model that uses only the error prone laboratory results of Trichomoniasis and bacterial vaginosis as stand-ins for the “gold-standard” clinical results. This model ignores all of the “gold-standard” clinical results for the validated data and yields extremely low coverage probability for both paradigms across all parameters. Considering this is an incorrect model, we would expect the coverage

Table 2.1. Results Comparison for Bayesian and Frequentist Techniques

	Frequentist			Bayesian		
Variable(β)	$\hat{\beta}$	(SE)	Coverage	$\hat{\beta}$	(SD)	Coverage
Naive Analysis						
Trichomoniasis ($\beta_1 = 0.64$)	1.546	(0.276)	8.8	1.526	(0.233)	3.8
Age ($\beta_2 = -0.05$)	-0.028	(0.010)	45.4	-0.029	(0.010)	48.8
Race ($\beta_3 = 0.79$)	0.566	(0.182)	76.8	0.612	(0.180)	79.8
HIV Risk Chrt ($\beta_4 = 0.28$)	0.826	(0.173)	9.8	0.777	(0.171)	18.2
HIV Status ($\beta_5 = 0.23$)	-0.314	(0.174)	13.2	-0.244	(0.174)	24.2
Ideal Analysis						
Trichomoniasis ($\beta_1 = 0.64$)	0.661	(0.190)	95.8	0.639	(0.169)	95.8
Age ($\beta_2 = -0.05$)	-0.050	(0.009)	95.0	-0.051	(0.009)	94.6
Race ($\beta_3 = 0.79$)	0.789	(0.156)	95.6	0.813	(0.161)	94.4
HIV Risk Chrt ($\beta_4 = 0.28$)	0.293	(0.143)	95.8	0.281	(0.146)	95.2
HIV Status ($\beta_5 = 0.23$)	0.232	(0.153)	95.0	0.235	(0.155)	97.2
Complete Analysis (25%)						
Trichomoniasis ($\beta_1 = 0.64$)	0.662	(0.378)	94.6	0.680	(0.338)	94.2
Age ($\beta_2 = -0.05$)	-0.052	(0.017)	94.4	-0.050	(0.017)	95.8
Race ($\beta_3 = 0.79$)	0.794	(0.289)	93.6	0.838	(0.304)	94.0
HIV Risk Chrt ($\beta_4 = 0.28$)	0.309	(0.262)	94.0	0.309	(0.265)	93.0
HIV Status ($\beta_5 = 0.23$)	0.238	(0.277)	95.2	0.263	(0.285)	94.0
Complete Analysis (15%)						
Trichomoniasis ($\beta_1 = 0.64$)	0.662	(0.498)	94.4	0.735	(0.441)	95.0
Age ($\beta_2 = -0.05$)	-0.053	(0.022)	93.8	-0.049	(0.022)	95.0
Race ($\beta_3 = 0.79$)	0.808	(0.374)	94.0	0.853	(0.391)	96.8
HIV Risk Chrt ($\beta_4 = 0.28$)	0.301	(0.340)	93.8	0.316	(0.342)	96.2
HIV Status ($\beta_5 = 0.23$)	0.233	(0.360)	94.2	0.257	(0.368)	92.4
Complete Analysis (10%)						
Trichomoniasis ($\beta_1 = 0.64$)	0.711	(0.613)	92.6	0.721	(0.555)	93.2
Age ($\beta_2 = -0.05$)	-0.055	(0.027)	90.2	-0.048	(0.026)	95.2
Race ($\beta_3 = 0.79$)	0.831	(0.456)	93.8	0.916	(0.488)	93.4
HIV Risk Chrt ($\beta_4 = 0.28$)	0.257	(0.410)	92.8	0.346	(0.425)	92.6
HIV Status ($\beta_5 = 0.23$)	0.238	(0.435)	93.8	0.299	(0.459)	93.2
Complete Analysis (5%)						
Trichomoniasis ($\beta_1 = 0.64$)	0.926	(1.040)	83.0	0.779	(0.832)	95.2
Age ($\beta_2 = -0.05$)	-0.056	(0.036)	76.2	-0.046	(0.037)	93.4
Race ($\beta_3 = 0.79$)	0.951	(0.651)	82.4	0.995	(0.721)	93.4
HIV Risk Chrt ($\beta_4 = 0.28$)	0.361	(0.574)	78.2	0.376	(0.626)	93.2
HIV Status ($\beta_5 = 0.23$)	0.204	(0.606)	78.8	0.337	(0.675)	91.8
Complete Analysis (2.5%)						
Trichomoniasis ($\beta_1 = 0.64$)	1.508	(2.221)	73.6	1.068	(1.310)	91.8
Age ($\beta_2 = -0.05$)	-0.071	(0.048)	66.4	-0.043	(0.052)	95.0
Race ($\beta_3 = 0.79$)	1.282	(1.083)	73.0	1.069	(1.101)	92.4
HIV Risk Chrt ($\beta_4 = 0.28$)	0.367	(0.855)	72.2	0.357	(0.974)	93.2
HIV Status ($\beta_5 = 0.23$)	0.513	(2.118)	68.4	0.364	(1.045)	95.6

probabilities to be small; for the primary parameter of interest, β_1, the coefficient for the predictor of X on Y, we get coverage probabilities of 8.8% for the frequentist paradigm and 3.8% for the Bayesian paradigm. These probabilities are in large part due to the high degree of bias in the estimates. For β_1, the true parameter is 0.64 while the estimates are about 1.5 for both paradigms. It is well known that non-differential misclassification leads to estimators biased towards the null; since this is differential misclassification, the bias could be in either direction, and in this case is biased high. This, coupled with the small standard errors estimated under each paradigm, results in the small coverage that can be seen for each of the five estimated parameters.

In stark contrast to the "naïve" model is the "ideal" model; here we assume that the gold-standard measurements are available for all subjects. Based upon the model assumption, which is that the data are error free, we would expect to see coverage probabilities close to the nominal level of 95%. As shown in Table 2.1, the coverage is near 95% for each parameter, with estimates that are centered close to the truth and with small standard errors. These results are a "best case" scenario that we can use to compare to the results from the various levels of validation data.

We next consider the 25% validation sample fraction case for the misclassification model. Here we see that both the frequentist and Bayesian approach are very close to attaining 95% coverage probability. Accounting for misclassification is known to increase variance in estimation, so it is not surprising that we see some increase in the standard errors from the ideal method with each standard error almost doubling; similar results are seen for the posterior standard deviations. Though the variability is increased, the coverage probabilities are close to nominal and the bias is relatively small, thus 25% validation appears effective for both the frequentist and Bayesian approach.

We next consider the case in which only 15% of our data is validated by gold-standard measurements. As with the 25% validation case, we see that both the frequentist and Bayesian intervals are close to attaining 95% coverage probability. Though the cov-

erages are similar, there is an increase in the variability of the estimates from the 25% validation case. We now start to see a difference in the two approaches; here, the Bayesian approach attains a 95% coverage probability for all but one estimate, while the frequentist approach does not attain 95% coverage probability for any of the estimates. Both approaches seem to produce estimates that center around the truth and have relatively similar standard errors at this validation level.

Decreasing the validation data to 10% yields more differences between the two procedures. However, the more drastic jump is at a validation level of 5%. The Bayesian approach has approximate 95% coverage probability for all of the parameters. The frequentist approach, on the other hand, has an 83% coverage probability or less for all of the parameters.

For a validation size of 2.5% the frequentist paradigm has issues for both bias and coverage. As can be seen in Table 2.1, the coverage probabilities are different from the 95% nominal value. In addition, there is considerable bias and the standard errors are considerably larger. It is interesting to note that even with the large degree of variability in the estimates, the coverage probabilities are small. Comparing these findings with the results from the Bayesian approach, we see that, even at the 2.5% validation size, the results are behaving quite well. Though there is more bias than the cases with higher validation sizes, estimates are still much closer than the frequentist approach. The standard deviations are also the same or much smaller than the frequentist approach. Combined with the fact that this approach actually captured the truth relatively well (the coverage probabilities are all above 90%) we see that the Bayesian approach performs much better than the frequentist approach for this small validation case.

The results indicate that the priors, $N(0, 1)$, are actually moderately informative. With some datasets the frequentist estimates "explode" and result in very large values for the parameters. The priors keep the posterior from large values, thus keeping estimates

Table 2.2. Results Comparison for Two Datasets Exhibiting Extreme Frequentist β_1 Estimates (True Value: 0.640)

Frequentist				Bayesian			
$\hat{\beta}_1$	(SE)	2.5^{th} %ile	97.5^{th} %ile	$\hat{\beta}_1$	(SD)	2.5^{th} %ile	97.5^{th} %ile
10.395	(8.337)	-5.945	26.735	4.601	(1.968)	1.060	8.767
10.100	(NA)	NA	NA	4.478	(1.732)	1.547	8.422

more reasonable. Of course, these priors should be justified but, in this case, large odds ratios would not be expected.

2.6.1 Investigation into Data Sets with Divergent Results

As shown in Table 2.1, the largest disparity between the Bayesian and frequentist approaches occurs at a validation size of 2.5%. In order to better understand the incongruence between the approaches we examine specific datasets for which we see "extreme" parameter estimates. Namely, we focus on datasets for which β_1, our parameter of interest, was estimated under the frequentist approach to be great than 10. We used this as our cutoff because of the discussion on Bayesian priors in Section 2.3. We discuss the results of two different datasets that had this result.

The results from these datasets are shown in Table 2.2. Clearly, the estimates of β_1 using the frequentist approach were much more biased than the Bayesian estimates. The frequentist standard error for the first dataset is more than four times larger than the Bayesian standard deviation. This gives frequentist interval estimates that span on both sides of zero compared to the Bayesian interval estimates that remain on the positive side of zero, where the truth lies. This shows that the conservative variance on the priors for the Bayesian approach are reigning in the estimates enough to make them not "explode". For the second dataset, although we were able to get an estimate of β_1 the analysis did not converge and thus the standard error was unable to be calculated. A data set with these issues under the frequentist approach would have been removed from our frequentist analysis in the simulation study. Here we see that the Bayesian approach was not only able to converge

upon an estimate and a standard deviation, but we also see that the estimate is far less biased than the potential estimate from the frequentist approach. Again, this shows that our prior information, although diffuse, is providing enough information to garner estimates from the Bayesian approach where the frequentist would have failed to.

Interestingly, both datasets did have a common quality in the validation sub-sample. Neither data set observed any patients with false negative exposure measurements. This could mean that we had a relatively flat likelihood as a result; the frequentist approach struggles to handle this problem, while the $N(0, 10)$ prior used Bayesian approach protects against extreme values in flat likelihoods, especially as the validation size decreases. This prior "information" results in fairly unbiased and precise estimates of the parameters compared to the frequentist approach.

We can see from the posterior densities of β_1 in Figure 2.6.1 the densities are smooth and suitably wide given the lack of information in the validation data. Thus, we have evidence of convergence and we are able to see reasonable inferences for both data sets.

2.6.2 Investigation into Likelihood Contributions

Finally, we wish to compare the overall ability of each approaches likelihood calculations. For this, we compare the estimated likelihood contribution for a specific individual to the true likelihood contribution for that individual using the true parameters. We will do this for an individual in the validation subsample and the main study for each of the

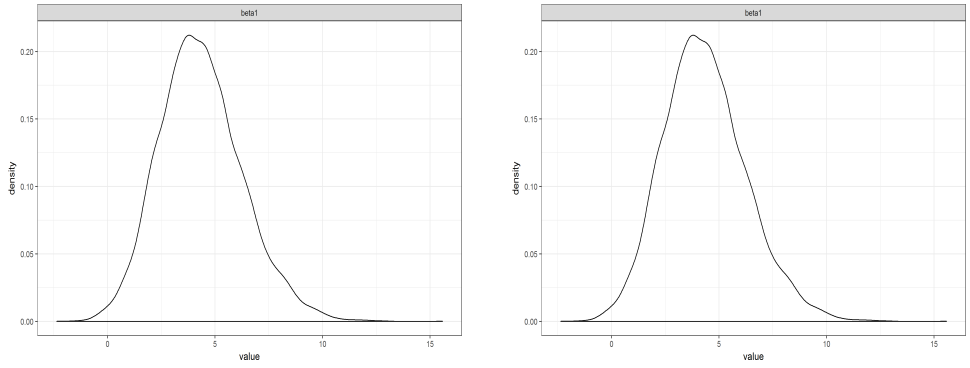

Figure 2.6.1. Posterior Densities of β_1 for Datasets with Extreme Frequentist Estimates

Table 2.3. Likelihood Contributions Comparisons between the Truth, Bayesian Estimates, and Frequentist Estimates

Subsample	True Parameters	Validation Size(%)	Frequentist Estimates	Bayesian Estimates
Validation	0.056912	25	0.054626	0.054526
		15	0.052251	0.052908
		10	0.050135	0.048159
		5	0.034381	0.043394
		2.5	0.019848	0.048117
Main	0.739109	25	0.746559	0.742698
		15	0.746883	0.743737
		10	0.743093	0.748608
		5	0.732732	0.753509
		2.5	0.719765	0.776263

validation sizes. Suppose the individual of interest in both cases has an error prone response measurement of $BV^* = 0$ but a gold standard response measurement of $BV = 1$ once validated; further, suppose the individual has an error prone exposure measurement of $Trich^* = 0$ with a gold standard exposure measurement of $Trich = 1$ once validated. Also, the individual is Black ($Race = 1$), not HIV Positive ($HIVStatus = 0$), not an intravenous drug user ($RiskCohort = 0$), and 37 years old ($Age = 37$). The results for this individual are shown in Table 2.3; there, you can see that there is a considerable difference in the likelihood contributions depending on whether the data had been validated, or not.

The results show that the likelihood contribution for the validation data is much smaller than the contribution for the main study data. Also, for this individual, as the validation size decreases, both approaches provide estimated likelihood contributions that deviate from the true likelihood contribution. Interestingly, it appears that the Bayesian approach estimates the likelihood contribution similarly for validated data and main study data; each type of data biases slightly as the validation size decreases. Conversely, for the frequentist approach, the estimate of the likelihood contribution for validated data is extremely biased for small validation sizes, while the main study data slightly biases as the validation size decreases.

CHAPTER THREE

Finding the Breaking Points

3.1 Baseline Comparison Model

In Chapter Two we discussed the methods for a binary regression with a misclassified response and exposure. We demonstrated that results were better for larger validation fractions. In this chapter we further investigate the various points at which these methods can break down either with lack of convergence or poor performance. We will examine these so-called "breaking points" via both the Bayesian and frequentist approach and, where appropriate, determine which approach performs better.

We use similar models presented in section 2.2.3 that are adjusted to illustrate important points. We will use only a single covariate in our models. This means, the logistic regressions we use are as follows:

$$\text{logit}[Pr(X = 1|C_1)] = \gamma_0 + \gamma_1 C_1 \tag{3.1}$$

$$\text{logit}[Pr(Y = 1|X, C_1)] = \beta_0 + \beta_1 X + \beta_2 C_1 \tag{3.2}$$

$$\text{logit}[Pr(X^* = 1|X, Y, C_1)] = \delta_0 + \delta_1 X + \delta_2 Y + \delta_3 C_1 \tag{3.3}$$

$$\text{logit}[Pr(Y^* = 1|Y, X, X^*, C_1)] = \theta_0 + \theta_1 X + \theta_2 Y + \theta_3 X^* + \theta_4 C_1. \tag{3.4}$$

The logistic regression models are then utilized in the likelihood components for the main study data and validated data as:

$$L_m = \prod_{i=1}^{n_m} \Bigg\{ \sum_{y_i=0}^{y_i=1} \sum_{x_i=0}^{x_i=1} Pr(Y_i^*|Y_i, X_i, X_i^*, C_{1i}) Pr(X_i^*|X_i, Y_i, C_{1i}) \\ Pr(Y_i|X_i, C_{1i}) Pr(X_i|C_{1i}) \Bigg\} \tag{3.5}$$

$$L_v = \prod_{j=1}^{n_v} \{ Pr(Y_j^*|Y_j, X_j, X_j^*, C_{1j}) Pr(X_j^*|X_j, Y_j, C_{1j}) \\ Pr(Y_j|X_j, C_{1j}) Pr(X_j|C_{1j}) \}. \tag{3.6}$$

Recall, the product of these components is proportional to the full likelihood.

In order to create a comparison between the two approaches, we develop the following data generation and simulation process:

(1) Set the true values of the parameters. In this case we use:

   (a) $\beta_0 = 0.5$, $\beta_1 = 2$, $\beta_2 = -1$,

   (b) $\delta_0 = -3$, $\delta_1 = 1.5$, $\delta_2 = 1.5$, $\delta_3 = 1$,

   (c) $\gamma_0 = -2$, $\gamma_1 = 0.75$,

   (d) $\theta_0 = -2$, $\theta_1 = 1$, $\theta_2 = 3$, $\theta_3 = 2$, and $\theta_4 = -2$.

(2) For the Bayesian approach, each of the coefficients are given Normal prior distributions with mean 0 and variance $\sigma^2 = 10$, as discussed in Chapter Two (Section 2.3).

(3) Generate a full dataset from these parameters using the assumptions and formulas described for dependent differential misclassification (equations 3.1 to 3.6).

(4) Estimate the parameters assuming dependent differential misclassification; specifically utilizing equation 3.10.

   (a) Using the frequentist approach, we utilize maximum likelihood optimization techniques to estimate the parameters.

   (b) Under the Bayesian approach, we will use MCMC techniques to estimate the parameters after verifying convergence.

(5) Compare the estimated parameters to the “true values" used to generate the data.

(6) Repeat these steps 500 times and report the appropriate summaries for each method.

As stated in steps 4a - 4b, we will check the convergence of each method. As can be seen in table 3.1, dataset convergence for the Frequentist results decreased as the percentage of validation data decreased beyond 15%; roughly three-fifths of the datasets converged once the validation size decreased to 2.5%. Convergence was assessed for the Bayesian

Table 3.1. Convergence for Frequentist Results for the Baseline Model

Validation Size	Converged Datasets/Total Datasets
25%	$500/500$
15%	$500/500$
10%	$495/500$
5%	$439/500$
2.5%	$302/500$

Table 3.2. Settings for Bayesian Results for the Baseline Model

Validation Size	Iterations	Burn-in	Thin
25%	$25,000$	$15,000$	1
15%	$40,000$	$20,000$	2
10%	$45,000$	$25,000$	2
5%	$70,000$	$40,000$	3
2.5%	$110,000$	$60,000$	5

versions; we used the settings for these runs as shown in table 3.2. Refer to appendix A.2 for a full display of each of the convergence diagnostic checks for each validation size.

For the sake of brevity, we will discuss the convergence diagnostics for the smallest validation size of 2.5%. We discuss these plots since it has been shown that for smaller validation sizes convergence may be more difficult to attain. For a validation size of 2.5% we used an iteration size of $110,000$, a burn-in of $60,000$, and used every 5^{th} iteration for inference. Figure 3.1.1 displays the autocorrelation plot and trace plot for this validation size. For the autocorrelation plot you should typically see that as the lag increases, the autocorrelation dampens to zero and remains close to zero. Trace plots should show no discernible pattern as they illustrate the chains exploring the parameter space. As you can see between the two plots, the autocorrelation does tend to zero and the trace plot appears to appropriately explore the parameter space.

Figure 3.1.2 shows the density plot and running means plot for the 2.5% validation size. The density plot, if we have attained convergence, should show a smooth curve with the height of the curve at the value of our parameter estimate. The running mean plot, again, if convergence is attained, should show a chain that converges to our parameter estimate

over the length of the chain. For this validation size we see that the density is in fact smooth. We can also see that the running mean plot does converge upon an estimate, even if it is skewed high compared to our known true value of $\beta_1 = 2$. These estimates could be skewed due to random error or simply because we are examining a rather extreme case in which $n_v = 25$. Using this same criterion and the settings in Table 3.2, we have shown evidence of chain convergence for this validation size and for each of the other validation sizes as provided in Appendix A.2.

Now that convergence has been checked, we can compare the results of the methods for this baseline model. Table 3.3 outlines the results from this process; we can see a similar result here as we saw in Chapter Two. The Bayesian approach estimates the parameter of interest well for each validation size, with coverage probabilities close to 95%. The frequentist version does well for large validation sizes, but as the validated sample size is reduced the results diverge from the 95% nominal coverage probabilities. The frequentist estimates are also tending to be positively biased with extremely large standard errors,

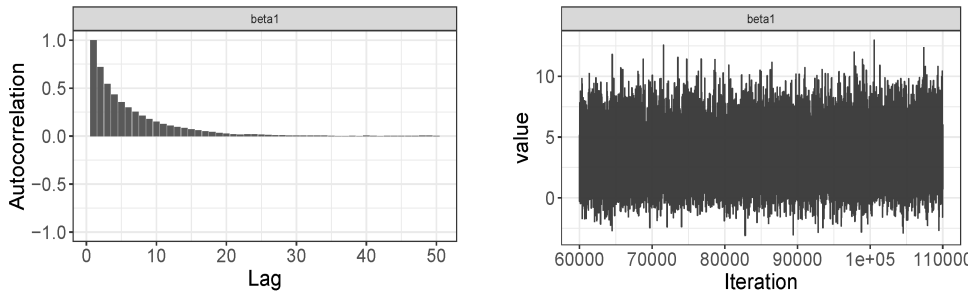

Figure 3.1.1. Autocorrelation and trace plots for 2.5% validation size for the baseline model

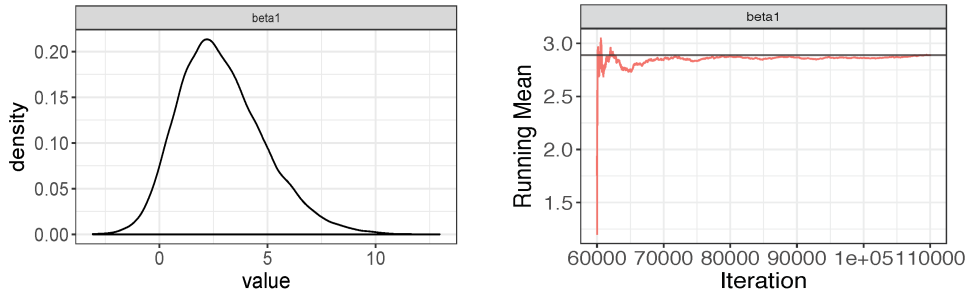

Figure 3.1.2. Density and running mean plot for the 2.5% validation size for the baseline model

Table 3.3. Results Comparison for Bayesian and Frequentist Techniques Assuming Dependent Differential Misclassification

	Frequentist			Bayesian		
Variable, β	$\hat{\beta}$	(SE)	Coverage	$\hat{\beta}$	(SD)	Coverage
25% Validated						
$\beta_1 = 2.00$	2.188	(1.197)	97.4	2.215	(0.625)	93.4
$\beta_2 = -1.00$	-1.015	(0.303)	95.8	-1.007	(0.305)	93.4
15% Validated						
$\beta_1 = 2.00$	2.641	(3.607)	95.6	2.296	(0.826)	94.6
$\beta_2 = -1.00$	-1.015	(0.392)	95.0	-1.021	(0.393)	96.4
10% Validated						
$\beta_1 = 2.00$	3.225	(20.395)	95.0	2.365	(1.023)	96.2
$\beta_2 = -1.00$	-1.018	(0.485)	95.4	-1.017	(0.482)	94.8
5% Validated						
$\beta_1 = 2.00$	5.517	(182.732)	84.6	2.505	(1.430)	96.2
$\beta_2 = -1.00$	-1.243	(3.401)	83.2	-1.043	(0.724)	93.0
2.5% Validated						
$\beta_1 = 2.00$	8.139	(412.282)	58.6	2.346	(1.840)	97.8
$\beta_2 = -1.00$	-1.685	(48.468)	57.6	-1.066	(1.093)	96.2

especially for the parameter β_1. As we saw in Chapter Two, our priors are only mildly informative, but they are useful in cases that lead to flat likelihoods. Refer to Section 2.6.1 for an example and discussion of this phenomenon.

The results of this section (Table 3.3) will be used throughout the remainder of this chapter to compare to all of the "breaking point" scenarios. The scenarios we will examine are the effects of changing the assumption of dependent differential misclassification to independent differential misclassification, increasing the overall sample size by a factor of 10, and, lastly, adding more information to the prior distributions.

3.2 Independent vs. Dependent Misclassification

Tang et al. (2015a) proposed a dependent differential misclassification model. In their model the error prone exposure variable, X^*, is related to the error prone response variable, Y^*. The relationship is modeled via the following logistic regressions for our likelihoods:

$$\text{logit}[Pr(X = 1|C_1, \ldots C_P)] = \gamma_0 + \sum_{p=1}^{Q} \gamma_p C_p \tag{3.7}$$

$$\text{logit}[Pr(Y = 1|X, C_1, \ldots C_Q)] = \beta_0 + \sum_{q=1}^{Q} \beta_q C_q + \beta_{Q+1} X \tag{3.8}$$

$$\text{logit}[Pr(X^* = 1|X, Y, C_1, \ldots C_R)] = \delta_0 + \sum_{r=1}^{R} \delta_r C_r + \delta_{R+1} X + \delta_{R+2} Y \tag{3.9}$$

$$\begin{aligned}\text{logit}[Pr(Y^* = 1|Y, X, X^*, C_1, \ldots C_S)] = \theta_0 &+ \sum_{s=1}^{S} \theta_s C_s \\ &+ \theta_{S+1} X + \theta_{S+2} Y + \theta_{S+3} X^*\end{aligned} \tag{3.10}$$

Assuming that the two error prone measurements are not only dependent on the gold-standard measurements but are themselves dependent (equation 3.10) may seem counter intuitive. One might think that the gold-standard measurements would render the error prone measurements non-informative without the added complexity of another parameter to estimate for the dependence between the two error prone measurements. In this section we explore the assumption of dependence among the error prone measurements. To relax this assumption, we will use the following model to run our analyses:

$$\text{logit}\,[Pr(Y^* = 1|Y, X, C_1, C_2, \ldots C_S)] = \theta_0 + \sum_{s=1}^{S} \theta_s C_s + \theta_{S+1} X + \theta_{S+2} Y \tag{3.11}$$

This model is covered in more detail in Section 2.2.2, but of note is the difference in model definition 3.11 (also model 2.18) and our previous dependent model definition, model 3.10 (also model 2.23). In the case of dependent differential misclassification, there is an added parameter to estimate, the parameter associated with the error prone exposure variable. For this section, we will examine whether the error prone measurement on the exposure truly does help to predict the error prone response measurement, even with the presence of both gold-standard measurements.

To accomplish this, we will follow a similar data generation and simulation process as in Section 3.1:

(1) Set the true estimates of the parameters based on a general example outlined in more detail in section 3.1.

Table 3.4. Convergence for Frequentist Results for the Independent Model

Validation Size	Converged Datasets/Total Datasets
25%	$500/500$
15%	$500/500$
10%	$492/500$
5%	$421/500$
2.5%	$327/500$

(2) Generate a full dataset from these parameters using the assumptions and formulas described for dependent differential misclassification (eqs. 3.7-3.10).

(3) For the Bayesian approach, each of the coefficients are given Normal$(\mu = 0, \sigma^2 = 10)$ prior distributions.

(4) Estimate the parameters using both of the approaches assuming independent differential misclassification; specifically utilizing equation 3.11 instead of our previous equation 3.10.

   (a) Using the frequentist approach, we utilize maximum likelihood optimization techniques to estimate the parameters.

   (b) Under the Bayesian approach, we will use MCMC techniques to estimate the parameters.

(5) Compare the estimated parameters under the assumption of independent differential misclassification to the “true estimates" used to generate the data under our assumption of dependent differential misclassification.

(6) Repeat these steps 500 times and report the appropriate summaries for each method.

As stated in steps 4a - 4b, we will check the convergence of each method. As can be seen in table 3.4, convergence for the frequentist results decreased as the percentage of validation data decreased below 15%; less than two-thirds of the datasets converged once the validation size decreased to 2.5%. Convergence was assessed for the Bayesian versions; we used the same settings for these runs as we used for the “Baseline” models.

Refer to appendix A.3 for a full display of each of the convergence diagnostic checks for each validation size.

For the sake of brevity, we will discuss the convergence diagnostics for the smallest validation size of 2.5%. For a sample size of $1,000$, a validation less than 2.5% with $N(0.0.1)$ priors we could have considerable convergence problems. Here we used an iteration size of $110,000$, a burn-in of $60,000$, and a thin of 5. Figure 3.2.1 displays the autocorrelation plot and trace plot for this validation size. For the autocorrelation plot you should typically see that as the lag increases, the autocorrelation lowers toward zero and remains close to zero. For the trace plot you should typically see that as the iteration increases, there is no discernible pattern to the estimates, and the estimates seem to hover evenly over the estimate. As you can see between the two plots, the autocorrelation does tend to zero, although slowly, and the trace plot has no pattern. Both plots show no obvious signs that convergence has been violated. Figure 3.2.2 shows the density plot and running means plot for the 2.5% validation size. The density plot, if we have attained convergence, should show a smooth curve over the mode of the distribution; it can be asymmetric. The running mean plot, again, if convergence is attained, should show a dampening of values toward the mode of the distribution. For this validation size we see that the density is in fact smooth although slightly asymmetric. We can also see that the running mean plot does converge upon an estimate.

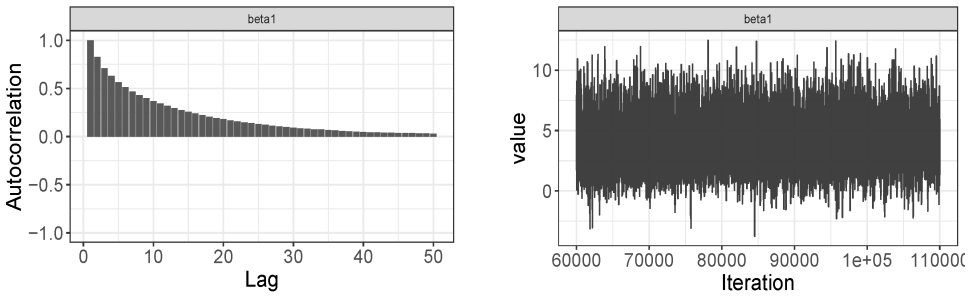

Figure 3.2.1. Autocorrelation and trace plots for 2.5% validation size under the Independent model

The results of the frequentist methods for both the analysis assuming dependent differential misclassification (baseline model) and independent differential misclassification are shown in table 3.5. Here we can see that for a validation size of 25% the assumption change has an effect on the estimate of β_1, our parameter of interest, which relates the gold-standard exposure variable to the gold-standard response. Our estimate of $\beta_1 = 2.00$ increases from 2.188 to 2.946, and although the standard errors seem to be relatively unaffected, this change in estimate results in coverage decreasing from 97.4% to 84.0%.

Under both analyses as the validation size decreases we see more bias in our estimates and higher standard errors for both parameters. Recall, that the correct analysis for the data is dependent differential misclassification, however by decreasing the amount of validated data to 2.5% the difference between the two analyses is difficult to discern. The estimates for both β_1 and β_2 are now both biased (more so for the analysis using independent differential misclassification) and the standard errors are now unrealistically large for both analyses. The standard errors are actually larger for the dependent differential misclassification analysis due, in part, to a more complicated model. In either case, we can see that the frequentist methods do not adequately predict for the gold-standard response when the validation size is small, or in this case when $n_v = 25$.

The results of the Bayesian approach for both the analysis assuming dependent differential misclassification (baseline model) and independent differential misclassification are shown in table 3.6. Here we can see that for a validation size of 25% the additional

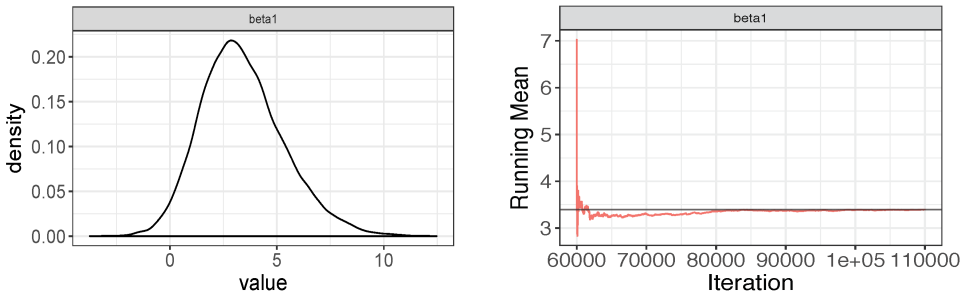

Figure 3.2.2. Density plot and running mean plot for the 2.5% validation size under the Independent model

Table 3.5. Results Comparison for Frequentist Techniques Assuming Dependent Differential Misclassification and Independent Differential Misclassification

	Dependent			Independent		
Variable, β	$\hat{\beta}$	(SE)	Coverage	$\hat{\beta}$	(SE)	Coverage
25% Validated						
$\beta_1 = 2.00$	2.188	(1.197)	97.4	2.946	(1.137)	84.0
$\beta_2 = -1.00$	-1.015	(0.303)	95.8	-1.084	(0.283)	93.6
15% Validated						
$\beta_1 = 2.00$	2.641	(3.607)	95.6	3.654	(6.311)	90.0
$\beta_2 = -1.00$	-1.015	(0.392)	95.0	-1.097	(0.355)	95.0
10% Validated						
$\beta_1 = 2.00$	3.225	(20.395)	95.0	4.185	(28.842)	93.2
$\beta_2 = -1.00$	-1.018	(0.485)	95.4	-1.122	(0.434)	92.6
5% Validated						
$\beta_1 = 2.00$	5.517	(182.732)	84.6	6.555	(140.911)	83.2
$\beta_2 = -1.00$	-1.243	(3.401)	83.2	-1.208	(3.923)	78.2
2.5% Validated						
$\beta_1 = 2.00$	8.139	(412.282)	58.6	9.086	(310.449)	64.6
$\beta_2 = -1.00$	-1.685	(48.468)	57.6	-2.370	(43.615)	60.8

parameter has a substantial effect on the estimate of β_1, our parameter of interest, which relates the gold-standard exposure variable to the gold-standard response. Our estimate of $\beta_1 = 2.00$ increases from 2.215 to 2.889, and although the standard errors seem to be relatively unaffected, this change in estimate results in the coverage decreasing from 93.4% to 67.6%.

The Bayesian methods mimic the trends in results that we saw from the frequentist methods. We see here that under both assumptions as the validation size decreases we see more bias and less precision in our estimates. Interestingly, the independent differential misclassification analysis appears to have smaller standard error than the dependent differential misclassification analysis by validation size, although the estimates themselves are much more biased.

Of interest is the difference between the frequentist and Bayesian methods when the dependent differential misclassification model is correct, but the independent differential misclassification model is applied. As we saw in Table 3.5, the frequentist results became

Table 3.6. Results Comparison for Bayesian Techniques Assuming Dependent Differential Misclassification and Independent Differential Misclassification

	Dependent			Independent		
Variable, β	$\hat{\beta}$	(SD)	Coverage	$\hat{\beta}$	(SD)	Coverage
25% Validated						
$\beta_1 = 2.00$	2.215	(0.625)	93.4	2.889	(0.595)	67.6
$\beta_2 = -1.00$	-1.007	(0.305)	93.4	-1.079	(0.284)	93.6
15% Validated						
$\beta_1 = 2.00$	2.296	(0.826)	94.6	3.308	(0.817)	59.6
$\beta_2 = -1.00$	-1.021	(0.393)	96.4	-1.107	(0.359)	94.0
10% Validated						
$\beta_1 = 2.00$	2.365	(1.023)	96.2	3.329	(0.956)	70.0
$\beta_2 = -1.00$	-1.017	(0.482)	94.8	-1.065	(0.433)	95.0
5% Validated						
$\beta_1 = 2.00$	2.505	(1.430)	96.2	3.376	(1.325)	86.8
$\beta_2 = -1.00$	-1.043	(0.724)	93.0	-1.003	(0.609)	93.4
2.5% Validated						
$\beta_1 = 2.00$	2.346	(1.840)	97.8	2.920	(1.753)	97.2
$\beta_2 = -1.00$	-1.066	(1.093)	96.2	-0.991	(0.923)	92.4

increasingly more biased as the validation size decreased; this was true for both parameters we examined. However, the Bayesian method, as shown in Table 3.6, only had biased results for the parameter of interest, β_1, and the severity of the bias seemed to be consistent as the validation size decreased. The difference between the two methods is even further highlighted when you examine the difference in frequentist standard errors and Bayesian standard deviations. These two measures should be relatively similar in terms of magnitude, but as we can see between the two results tables, these quantities vary greatly for the estimates of β_1 for each validation case. In fact, the "best" that the frequentist method produces is still double the size of the Bayesian quantity (1.137 vs. 0.595 for the 25% validation size). Furthermore, for the 5% and 2.5% validation sizes these quantities vary greatly for the estimate of β_2 as well. Again, we see that the frequentist quantities are huge in comparison to the Bayesian counterparts.

3.3 Sample Size Determination

For the original study by Tang et al. (2015a) the authors used a data set that had 904 women in the study to investigate the models proposed. We follow up their work with a simulation study that uses a sample size of $1,000$ to model our "baseline" model in Section 3.1. Generally, larger sample sizes lead to more precise inferences. In this section, we explore the effects of increasing our sample size ten-fold for both the frequentist and Bayesian approaches.

The data generation and simulation steps are exactly the same as in Section 3.1, however we now use an overall sample size of $10,000$. Again, we must verify that we have attained convergence for both approaches. Dataset convergence for the frequentist results achieved 100% success for all validation sizes. Unlike the baseline models, which saw decreasing convergence as the validation size decreased, here we see that the increase in overall sample size has allowed the models to converge in each validation size scenario. Convergence was examined for the Bayesian versions as well; we used the settings for these runs as shown in table 3.7. Refer to Appendix A.4 for a full display of each of the convergence diagnostic checks for each validation size.

To be concise, we will discuss the convergence diagnostics for the smallest validation size of 2.5%; the discussions are similar for the remainder of validation sizes. For the validation size of 2.5% we used an iteration size of $75,000$, a burn-in of $55,000$, and kept every other iteration. Figure 3.3.1 displays the autocorrelation plot and trace plot for this validation size. To assess convergence with an autocorrelation plot you should see that as

Table 3.7. Settings for Bayesian Results for the Increased Sample Size Analysis

Validation Size	Iterations	Burn-in	Thin
25%	$16,000$	$6,000$	1
15%	$18,000$	$8,000$	1
10%	$35,000$	$15,000$	2
5%	$40,000$	$20,000$	2
2.5%	$75,000$	$55,000$	2

the lag increases, the autocorrelation falls to zero and remains close to zero. For the trace plot you should typically see that as the iteration increases, there is no discernible pattern to the estimates and a fairly even spread above and below the mode of the distribution. As you can see between the two plots, the autocorrelation does tend to zero and the trace plot has no discernible pattern.

Figure 3.3.2 shows the density plot and running means plot for the 2.5% validation size. The density plot, if we have attained convergence, should show a smooth, possibly asymmetric curve. The running mean plot, again, if convergence is attained, should show a chain that converges to the mode of the distribution. For this validation size we see that the density is in fact smooth and the running mean plot does converge upon a mode.

Table 3.8 provides the frequentist results comparing the baseline sample size of $1,000$ to the increased sample size of $10,000$. The frequentist approach benefits greatly from an increased sample size. As the validation size decreased in the Baseline results we

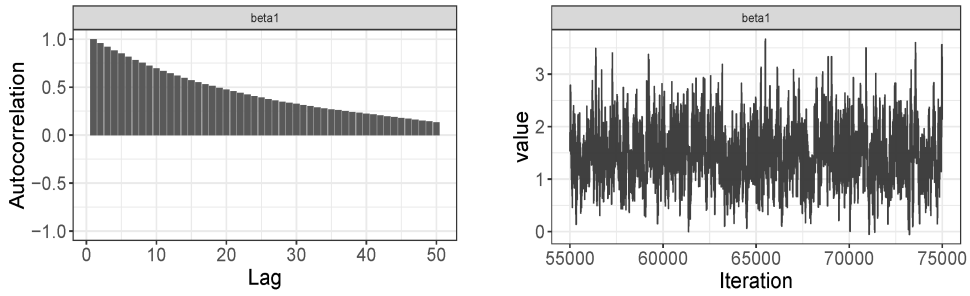

Figure 3.3.1. Autocorrelation and trace plots for 2.5% validation size with an overall sample size of $10,000$

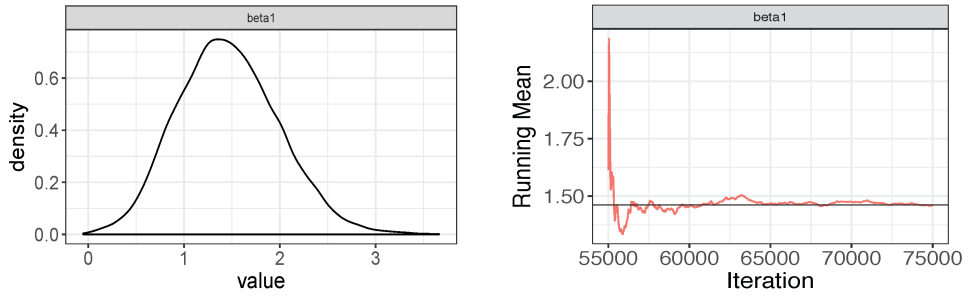

Figure 3.3.2. Density plot and running mean plot for the 2.5% validation size with an overall sample size of $10,000$

Table 3.8. Results Comparison for Frequentist Techniques with Samples Sizes Increased Ten-Fold

	$n = 1,000$			$n = 10,000$		
Variable, β	$\hat{\beta}$	(SE)	Coverage	$\hat{\beta}$	(SE)	Coverage
25% Validated						
$\beta_1 = 2.00$	2.188	(1.197)	97.4	2.004	(0.176)	92.8
$\beta_2 = -1.00$	-1.015	(0.303)	95.8	-1.009	(0.095)	93.8
15% Validated						
$\beta_1 = 2.00$	2.641	(3.607)	95.6	2.006	(0.228)	96.0
$\beta_2 = -1.00$	-1.015	(0.392)	95.0	-0.996	(0.121)	93.8
10% Validated						
$\beta_1 = 2.00$	3.225	(20.395)	95.0	2.039	(0.282)	94.8
$\beta_2 = -1.00$	-1.018	(0.485)	95.4	-1.003	(0.148)	94.2
5% Validated						
$\beta_1 = 2.00$	5.517	(182.732)	84.6	2.091	(0.412)	96.4
$\beta_2 = -1.00$	-1.243	(3.401)	83.2	-1.014	(0.208)	94.2
2.5% Validated						
$\beta_1 = 2.00$	8.139	(412.282)	58.6	2.235	(1.609)	98.0
$\beta_2 = -1.00$	-1.685	(48.468)	57.6	-1.053	(0.296)	94.8

saw an increase in the bias of our estimates as well as a huge increase in the precision of our estimates (the standard errors balloon up as the validation size decreases). With the increase in overall sample sizes, the amount of validated data has increased as well. With this increase, we can see that the results are now much less biased and do not suffer from the same drastic increase in standard errors as we saw in the baseline model.

At the largest validation size, 25%, we can see that the increased sample size results in more accurate and precise estimates; our estimates for both β_1 and β_2 are closer to the truth with smaller standard errors for the overall sample size of $10,000$ compared to our estimates and standard errors from an overall sample size of $1,000$. More notably, at the 2.5% validation size we can see that our estimate of β_1 decreased appropriately from 8.139 to 2.235, which is much closer to the true value of $\beta_1 = 2.00$. The estimate of β_2 also decreased in magnitude at this validation size, however the difference between the two is not as striking. The much more interesting change occurred in the standard error estimates between the two sample size analyses. With an overall sample size of $n = 1,000$ we see that our standard error is unreasonably large (412.282), however with an overall sample size

Table 3.9. Results Comparison for Bayesian Techniques with Samples Sizes Increased Ten-Fold

	$n = 1,000$			$n = 10,000$		
Variable, β	$\hat{\beta}$	(SD)	Coverage	$\hat{\beta}$	(SD)	Coverage
25% Validated						
$\beta_1 = 2.00$	2.215	(0.625)	93.4	2.013	(0.176)	94.6
$\beta_2 = -1.00$	-1.007	(0.305)	93.4	-1.000	(0.095)	95.8
15% Validated						
$\beta_1 = 2.00$	2.296	(0.826)	94.6	2.017	(0.228)	95.0
$\beta_2 = -1.00$	-1.021	(0.393)	96.4	-0.992	(0.121)	94.8
10% Validated						
$\beta_1 = 2.00$	2.365	(1.023)	96.2	2.036	(0.282)	95.6
$\beta_2 = -1.00$	-1.017	(0.482)	94.8	-0.991	(0.147)	95.4
5% Validated						
$\beta_1 = 2.00$	2.505	(1.430)	96.2	2.074	(0.406)	94.6
$\beta_2 = -1.00$	-1.043	(0.724)	93.0	-1.006	(0.208)	96.6
2.5% Validated						
$\beta_1 = 2.00$	2.346	(1.840)	97.8	2.210	(0.610)	93.0
$\beta_2 = -1.00$	-1.066	(1.093)	96.2	-1.012	(0.294)	93.4

of $n = 10,000$ our standard error becomes much more reasonable for a real-world scenario at a value of 1.609. Again, a similar decrease occurs for the standard error estimate for β_2 with a decrease from 48.468 to 0.296.

The results from the frequentist methods improved drastically, and while the Bayesian methods also improved (results are shown in Table 3.9), the results are not as drastic. This is due in large part to the fact that the baseline Bayesian analysis performed better than the baseline frequentist analysis. Even so, the increase in overall sample size did have a positive impact on lowering the bias in the estimates, especially as the validation size decreased. This was true for both parameters examined in Table 3.9. We can also see that the estimate of the standard deviations decreased by two-thirds in almost every case, if not more.

Interestingly, the difference between the frequentist and Bayesian methods seems to favor the frequentist methods for larger validation sizes (25% and 15%). The estimates for the frequentist method under these validation sizes seem to be slightly less biased than for the Bayesian method. The standard errors and standard deviations are very similar for both methods; thus, we use the decreased bias of the frequentist results to say that for larger

validation sizes we would tend to choose these methods for analysis. For mid-range validation sizes (i.e., 10% and 5%) we see no clear "winner" in the analysis; the frequentist and Bayesian approaches both offer fairly accurate and precise estimates of our parameters. For smaller validation sizes (i.e., 2.5%) the Bayesian approach seems to be more appropriate to use. Here we see that the Bayesian methods produced less biased estimates and that the standard deviation for the estimate of β_1 is less than half of the standard error estimate under the frequentist methods. Interestingly, the standard error and the standard deviation estimates under the appropriate methods for the estimate of β_2 are very close for all validation sizes, so it is clear that the increased sample size effects the results for our parameter of interest, not necessarily the remaining parameters.

3.4 Informative Priors vs. Non-Informative Priors

A main advantage of using the Bayesian approach in many areas is that expert opinion can be incorporated via informative prior distributions. For the sake of our comparison, so far, we have assumed that we have little knowledge from previous studies or experts on our parameters, though we have assumed extreme values are unlikely. For this reason, we have used relatively diffuse priors that follow a normal distribution with a mean of 0 and a standard deviation of about 3.2 (variance of $\sigma^2 = 10$, standard deviation is thus $\sqrt{10} = 3.162278$). As previously discussed in Section 2.3, this is a rather broad distribution that allows for a multitude of values, even some that would almost never occur naturally. For comparison, this section will explore the effects of having a slightly informative prior distribution on the parameters. We will look at three sets of prior distributions to assess the effect on the Bayesian approach: a prior that centers the distribution on the appropriate side of zero for the parameter at hand, a prior that centers the distribution with a smaller standard deviation, and a prior that centers the distribution and greatly narrows the spread of the distribution.

For each of the scenarios, we will follow similar steps as previous sections for data generation and analysis:

(1) Set the true estimates of the parameters based on a general example outlined in more detail in section 3.1.

(2) Generate a full dataset from these parameters using the assumptions and formulas described for dependent differential misclassification (eqs. 3.7-3.10).

(3) For the Bayesian approach, each of the coefficients are given a prior distribution unique to the section we are discussing.

(4) Estimate the parameters using Bayesian MCMC techniques assuming dependent differential misclassification; specifically utilizing equation 3.10.

(5) Compare the estimated parameters under each new prior distribution change.

(6) Repeat these steps 500 times and report the appropriate summaries.

Since we are obviously not affecting the frequentist methods when we change our prior distributions, for this section we will only focus on the results from the Bayesian approach.

3.4.1 Priors that are Centered on the Appropriate Side of Zero

For the sake of comparison, we will use our baseline model to compare to the new priors developed in this section. The baseline model we used in Section 3.1 utilized prior distributions that were the same for every parameter; we used: $Normal(\mu = 0, \sigma^2 = 10)$. As discussed in Chapter 2 (Section 2.3), this distribution allows for the scenario in which a one-unit change in the explanatory variable results in a predicted probability increase from 1% to 99% to be possible although highly improbable. Likewise, our prior choice allows for the improbable scenario that a one-unit change in the explanatory variable results in a predictive probability decrease from 99% to 1%. Each of these changes, when back transformed to the scale of our coefficients, means that coefficients on our parameters would

Table 3.10. Prior Means for the Priors that are Centered on the Appropriate Side of Zero

Variable	True Value	Prior Mean	Prior Variance
γ_0	-2.00	-1	10
γ_1	0.75	1	10
β_0	0.50	1	10
β_1	2.00	1	10
β_2	-1.00	-1	10
δ_0	-3.00	-2	10
δ_1	1.50	1	10
δ_2	1.50	1	10
δ_3	1.00	1	10
θ_0	-2.00	-1	10
θ_1	1.00	1	10
θ_2	3.00	2	10
θ_3	2.00	1	10
θ_4	-2.00	-1	10

very rarely be as large as 10 or -10. This prior allows for both of these situations, but gives relatively low weight to their probability of occurrence.

This prior is reasonable if the goal is to minimize influence, however in a real-world scenario you would most likely have a bit of intuition on whether or not your parameter is, at the very least, positive or negative. This leads us to choosing priors for our parameters that are somewhat reflective of which side of zero we believe they may tend to; table 3.10 shows the prior means we have chosen to utilize under this new Bayesian analysis.

As with the baseline model, we must find the optimal settings to achieve convergence without sacrificing unrealistically long run times. When we run our simulations for this model (with centered priors), we utilize the settings as displayed in Table 3.11. These settings allowed for convergence criteria to be met at all validation sizes examined; to show this we will discuss the convergence diagnostics for the 2.5% validation size. We examine the convergence of the smallest validation size since this is typically where we will face more convergence problems. We present all of the diagnostic checks for all of the validation sizes in Appendix A.5.

Table 3.11. Settings for Bayesian Results for the Priors that are Centered on the Appropriate Side of Zero

Validation Size	Iterations	Burn-in	Thin
25%	$15,000$	$5,000$	1
15%	$35,000$	$25,000$	1
10%	$50,000$	$40,000$	1
5%	$80,000$	$60,000$	2
2.5%	$105,000$	$85,000$	2

We present four diagnostic checks for each validation size: autocorrelation plot, density plot, running mean plot, and a trace plot. The autocorrelation plot should show that as the lag increases, the autocorrelation decreases. Ideally this would decrease to zero, however for this small of a validation size we see that the autocorrelation tends towards about 0.2. The density plot should show a smooth curve, and in fact we do see this for the 2.5% validation size. In the trace plot we should see that the lines have no discernible pattern and that the bouncing back and forth hovers over our true value; here we see that this is the case, even though we are seeing a wide variation in our estimates from negative to positive values. Lastly, the running means plot should show a convergence to a value over the iterations. We do see that our running mean tends towards about 0.60; we would prefer that this be closer to the truth ($\beta_1 = 2.00$), however we do still see convergence.

Since we have confirmed that convergence of our chains has been met, we can discuss the results of this process which are summarized in Table 3.12. Interestingly, the results seem to show that centering our prior distributions does not actually allow for more accurate

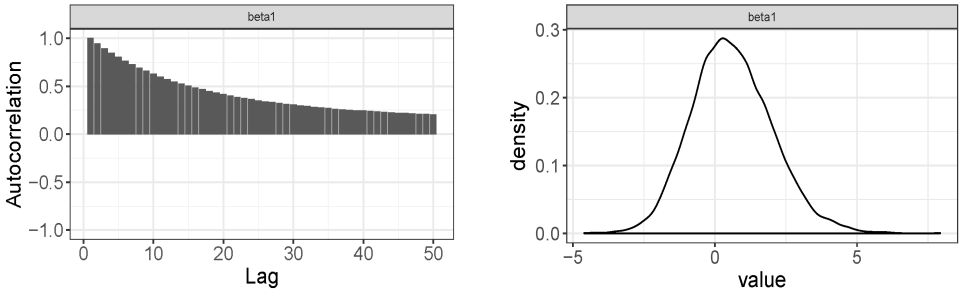

Figure 3.4.1. Auto-correlation and density plot for 2.5% validation sizes

Table 3.12. Results Comparison for Bayesian Techniques with Prior Distributions Centered on the Appropriate Side of Zero ($N(\mu, \sigma^2 = 10)$)

	Baseline $(N(0, \sigma^2 = 10))$			Centered $(N(\mu, \sigma^2 = 10))$		
Variable, β	$\hat{\beta}$	(SD)	Coverage	$\hat{\beta}$	(SD)	Coverage
25% Validated						
$\beta_1 = 2.00$	2.215	(0.625)	93.4	2.248	(0.624)	94.8
$\beta_2 = -1.00$	-1.007	(0.305)	93.4	-1.013	(0.304)	95.2
15% Validated						
$\beta_1 = 2.00$	2.296	(0.826)	94.6	2.426	(0.857)	93.6
$\beta_2 = -1.00$	-1.021	(0.393)	96.4	-1.055	(0.394)	96.2
10% Validated						
$\beta_1 = 2.00$	2.365	(1.023)	96.2	2.589	(1.091)	94.0
$\beta_2 = -1.00$	-1.017	(0.482)	94.8	-1.025	(0.483)	94.4
5% Validated						
$\beta_1 = 2.00$	2.505	(1.430)	96.2	2.726	(1.462)	97.4
$\beta_2 = -1.00$	-1.043	(0.724)	93.0	-1.089	(0.716)	93.6
2.5% Validated						
$\beta_1 = 2.00$	2.346	(1.840)	97.8	2.747	(1.914)	98.8
$\beta_2 = -1.00$	-1.066	(1.093)	96.2	-1.102	(1.093)	96.4

or precise estimates. For example, at the 25% validation size we see that our baseline model gave a slightly biased estimate of 2.215; compare this to our estimate from the analysis which centered our priors and we see that we now produce and even higher estimate at 2.248. The estimates for the standard error and standard deviation are almost the same, so we did not seem to gain higher precision either.

One possible explanation for the change in accuracy of our estimates is that we are now allowing for our distribution to spread to more extreme values. For instance, our prior

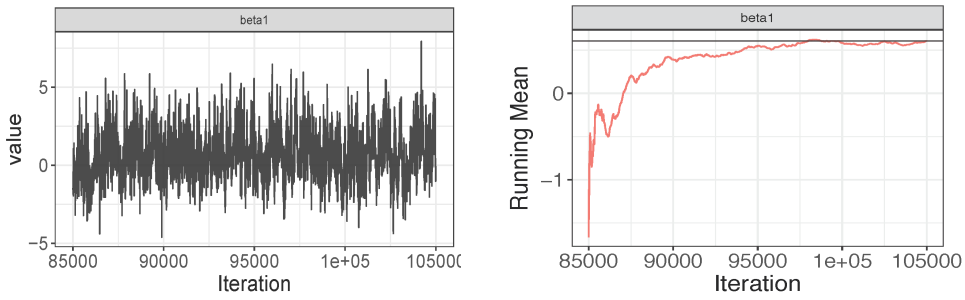

Figure 3.4.2. Trace and running means plot for 2.5% validation sizes

distribution for β_1 changed from being centered at 0, with probable values between about -7 to 7, to being centered at 1, with probable values now between -6 and 8. We are now allowing for even more extreme coefficient values to be probable, and thus the centering is actually skewing our results to become slightly more extreme than in our baseline model. For this reason, we need to examine how the analysis would be affected by a change in the prior distribution to give less probability to extreme values.

3.4.2 Priors that are Centered on the Appropriate Side of Zero with Moderate Prior Standard Deviations

We will examine the effect of narrowing our prior distributions slightly, in addition to centering the distributions on the appropriate side of zero. We will define our priors now by the same "centered" means as denoted in Table 3.10, but we will change the variance components from $\sigma^2 = 10$ to $\sigma^2 = 4$ (this gives a precision of $\tau = \frac{1}{\sigma^2} = 0.25$ and a standard deviation of $\sigma = 2$). These priors are described in Table 3.13. We run these results under the same settings as the settings used for the priors centered on the appropriate side of zero (refer to Table 3.11). Again, we must check the convergence of these settings before moving forward with our analysis process.

We will discuss the convergence for the smallest validation size (2.5%) since this size would be most likely to cause convergence errors. The remaining validation sizes did reach convergence; all diagnostics are shown in Appendix A.6. As we can see, we have attained convergence in each diagnostic. The autocorrelation plot shows that as the lag increases, our autocorrelation decreases; the density plot is smooth; the trace plot shows no discernible pattern and hovers evenly over an estimate of β_1; and the running mean plot shows that we converge upon an estimate as the iterations increase.

Since convergence criteria of our chains has been adequately checked, we can discuss the results of this process which are summarized in Table 3.14. Now that we have centered and slightly narrowed our prior distributions, we can see that our estimates have become less biased and more precise for every validation size. The bias is interestingly

Table 3.13. Prior Means for the Priors that are Centered on the Appropriate Side of Zero and Slightly Narrowed

Variable	True Value	Prior Mean	Prior Variance
γ_0	-2.00	-1	4
γ_1	0.75	1	4
β_0	0.50	1	4
β_1	2.00	1	4
β_2	-1.00	-1	4
δ_0	-3.00	-2	4
δ_1	1.50	1	4
δ_2	1.50	1	4
δ_3	1.00	1	4
θ_0	-2.00	-1	4
θ_1	1.00	1	4
θ_2	3.00	2	4
θ_3	2.00	1	4
θ_4	-2.00	-1	4

to the positive side of our true value of $\beta_1 = 2.00$ for all but one validation size. At the 2.5% validation size we see that we are now biased downward, or below $\beta_1 = 2.00$. This is

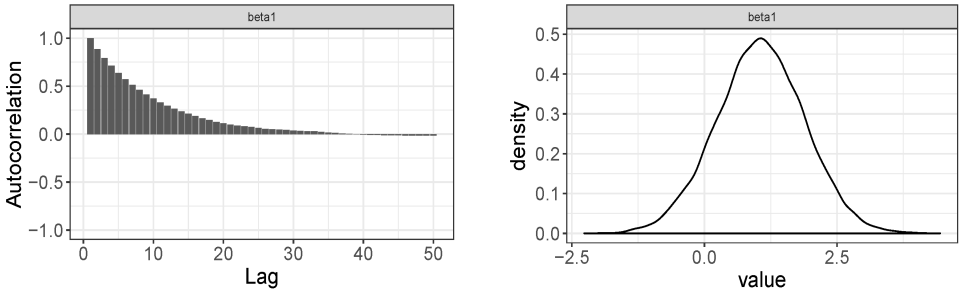

Figure 3.4.3. Auto-correlation and density plot for 2.5% validation sizes

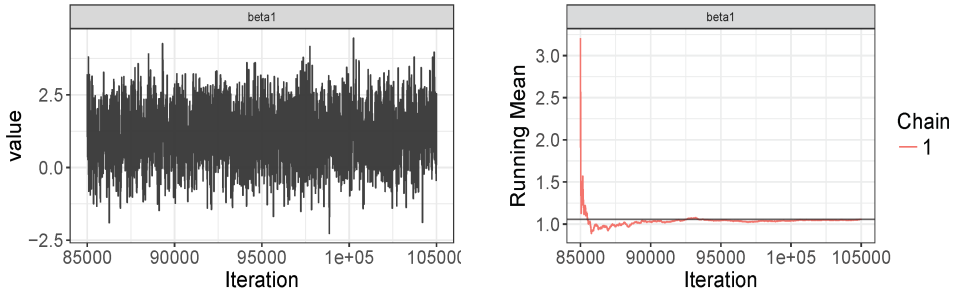

Figure 3.4.4. Trace and running means plot for 2.5% validation sizes

Table 3.14. Results Comparison for Bayesian Techniques with Prior Distributions Centered on the Appropriate Side of Zero with Slightly Smaller Spread ($N(\mu, \sigma^2 = 4)$)

	Baseline ($N(0, \sigma^2 = 10)$)			Centered and Slight Narrow ($N(\mu, \sigma^2 = 4)$)		
Variable, β	$\hat{\beta}$	(SD)	Coverage	$\hat{\beta}$	(SD)	Coverage
25% Validated						
$\beta_1 = 2.00$	2.215	(0.625)	93.4	2.161	(0.583)	96.0
$\beta_2 = -1.00$	-1.007	(0.305)	93.4	-0.997	(0.302)	94.2
15% Validated						
$\beta_1 = 2.00$	2.296	(0.826)	94.6	2.256	(0.760)	96.4
$\beta_2 = -1.00$	-1.021	(0.393)	96.4	-1.036	(0.387)	95.8
10% Validated						
$\beta_1 = 2.00$	2.365	(1.023)	96.2	2.251	(0.892)	96.8
$\beta_2 = -1.00$	-1.017	(0.482)	94.8	-1.024	(0.475)	94.8
5% Validated						
$\beta_1 = 2.00$	2.505	(1.430)	96.2	2.160	(1.142)	98.2
$\beta_2 = -1.00$	-1.043	(0.724)	93.0	-1.012	(0.668)	96.0
2.5% Validated						
$\beta_1 = 2.00$	2.346	(1.840)	97.8	1.990	(1.386)	98.6
$\beta_2 = -1.00$	-1.066	(1.093)	96.2	-1.098	(0.971)	96.6

an indication that at this level of the validation size the prior distribution, centered at 1, is having a larger effect on the posterior than is seen for the larger validation sizes.

3.4.3 Priors that are Centered on the Appropriate Side of Zero and Drastically Narrowed Distributions

Now we will examine the effect of narrowing our prior distributions even further. We will define our priors by the same "centered" means as in previous sections, but we will change the variance components from $\sigma^2 = 10$ and $\sigma^2 = 4$ to $\sigma^2 = 1$ (this gives a precision and standard deviation that are both equal to 1); refer to Table 3.15 for the specifics of our prior distributions for this section. We will be running our Bayesian approach under the same settings as the settings used for the Bayesian approach when we centered the priors and maintained a variance of $\sigma^2 = 10$ (refer to Table 3.11).

We will discuss the convergence for the smallest validation size (2.5%) since this size would be most likely to cause convergence errors. All diagnostics for all validation

Table 3.15. Prior Means for the Priors that are Centered on the Appropriate Side of Zero and Narrowed

Variable	True Value	Prior Mean	Prior Variance
γ_0	-2.00	-1	1
γ_1	0.75	1	1
β_0	0.50	1	1
β_1	2.00	1	1
β_2	-1.00	-1	1
δ_0	-3.00	-2	1
δ_1	1.50	1	1
δ_2	1.50	1	1
δ_3	1.00	1	1
θ_0	-2.00	-1	1
θ_1	1.00	1	1
θ_2	3.00	2	1
θ_3	2.00	1	1
θ_4	-2.00	-1	1

sizes are shown in Appendix A.7. As we can see in Figures 3.4.5 and 3.4.6, we attained convergence in each diagnostic. The autocorrelation plot shows that as the lag increases, our autocorrelation decreases; the density plot is smooth; the trace plot shows no discernible pattern and hovers evenly over an estimate β_1; and the running mean plot shows that we converge upon an estimate as the iterations increase. The remaining validation sizes did reach convergence with similar properties shown in their diagnostic plots.

As can be seen in Table 3.16, when we dramatically decrease the variability in our prior distributions we are effectively driving our results to be very nearly close to our prior distributions. This table gives the results for two parameters for which we see that the prior

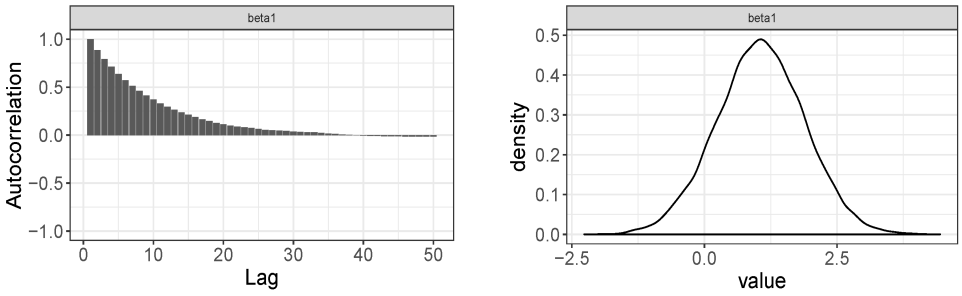

Figure 3.4.5. Auto-correlation and density plot for 2.5% validation sizes

Table 3.16. Results Comparison for Bayesian Techniques with Prior Distributions Centered on the Appropriate Side of Zero with Smaller Spread ($N(\mu, \sigma^2 = 1)$)

	Centered and Slight Narrow ($N(\mu, \sigma^2 = 4)$)			Centered and Narrowed ($N(\mu, \sigma^2 = 1)$)		
Variable, β	$\hat{\beta}$	(SD)	Coverage	$\hat{\beta}$	(SD)	Coverage
25% Validated						
$\beta_1 = 2.00$	2.161	(0.583)	96.0	1.878	(0.479)	97.4
$\beta_2 = -1.00$	-0.997	(0.302)	94.2	-0.972	(0.288)	95.8
15% Validated						
$\beta_1 = 2.00$	2.256	(0.760)	96.4	1.810	(0.568)	95.6
$\beta_2 = -1.00$	-1.036	(0.387)	95.8	-0.995	(0.362)	97.2
10% Validated						
$\beta_1 = 2.00$	2.251	(0.892)	96.8	1.734	(0.641)	97.4
$\beta_2 = -1.00$	-1.024	(0.475)	94.8	-0.949	(0.428)	97.8
5% Validated						
$\beta_1 = 2.00$	2.160	(1.142)	98.2	1.565	(0.754)	97.8
$\beta_2 = -1.00$	-1.012	(0.668)	96.0	-0.986	(0.560)	99.0
2.5% Validated						
$\beta_1 = 2.00$	1.990	(1.386)	98.6	1.354	(0.846)	98.4
$\beta_2 = -1.00$	-1.098	(0.971)	96.6	-1.000	(0.691)	99.8

is centered very near our truth for β_1 and actually on our truth for β_2. The prior for β_1 is shown now to drive our results. For example, if we focus on the most dramatic scenario, a validation size of 2.5%, we see that our prior on β_1 being centered at 1 is now guiding that estimate ($\hat{\beta}_1$ = 1.354) down towards 1 when it has a prior variance of 1. Our precision in our estimates has decreased, but with such biased estimates, this is no longer a desirable quality even if it does give us coverage probabilities near the nominal level of 95%.

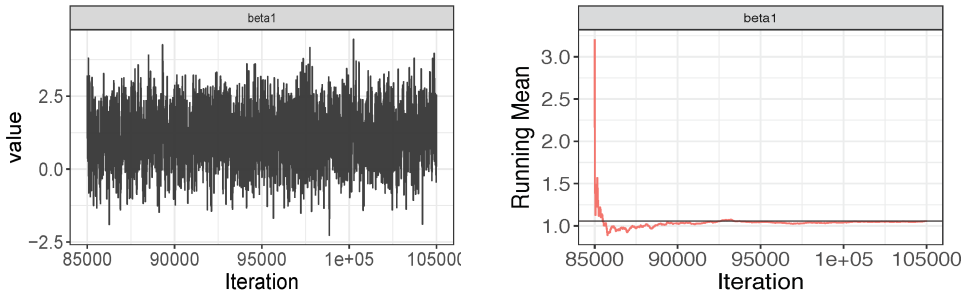

Figure 3.4.6. Trace and running means plot for 2.5% validation sizes

If we examine the results as they relate to β_2, we see a slightly different result in that our estimates are very close to the true values. We also see that our standard deviations are much smaller than the results from the slightly narrowed prior results. For example, our estimate of the standard deviation at the 2.5% validation size for β_2 was 1.386 when we used slightly narrowed priors and is now 0.846 which used priors with a variance of 1. We see that this relates to a coverage going from 98.6 to 98.4. Essentially, we are seeing very accurate and precise results for β_2 since we centered the results on the truth.

The difference in estimates for β_1 and β_2 tells us that unless we are very certain of our prior information (as shown in the case of β_2) we should refrain from using such a small variance, even if the coverage is close to our nominal level of 95%. We should utilize prior distributions with a sensible level of uncertainty (as shown in the case of β_1) to allow the data to guide our estimates, while still providing adequate prior information to improve our estimates in both bias and precision.

3.4.4 Estimating the Information in the Priors: Effective Sample Size Comparison

In Sections 3.1, 3.4.1, 3.4.2, and 3.4.3 we showed/discussed the results of using increasingly more informative priors on our results for dependent differential misclassification. A common argument against Bayesian analyses is that the diffuse (or "non-informative") priors are actually informative, and sometimes much too informative, to allow the data to drive our estimations. In this section, we explore one simple measure of the amount of information our priors provide, the prior effective sample size (ESS).

The prior effective sample size was notably discussed by Morita et al. (2008) in which they not only define the term, but they provide practical solutions to finding this measure for a variety of scenarios. ESS is, in a nutshell, the amount of information in the posterior that results from updating our prior distribution with the likelihood using Bayes Theorem. In general, if you have a diffuse prior this measure should be low, which corresponds to a higher amount of information coming from the likelihood rather than the

Table 3.17. ESS for Priors of Dependent Differential Misclassification

Priors	Summed ESS
Baseline: $N(0, \sigma^2 = 10)$	2.990698
Centered: $N(\mu, \sigma^2 = 10)$	5.923395
Centered and Moderate SD: $N(\mu, \sigma^2 = 4)$	15.03335
Centered and Drastic SD: $N(\mu, \sigma^2 = 1)$	60.00797

prior. Morita et al. (2008) developed a straight-forward R function to estimate the ESS for a logistic regression.

We used their formula to help ascertain the information in our priors by assuming that, at a maximum, the four logistic regressions will provide an additive ESS overall. Simply put, we calculate the ESS for each of the models (eqs. 3.7-3.10) and sum their results to describe the ESS for each of the four prior scenarios we used for this chapter. Refer to Appendix C.1 for a detailed description of the formula and settings we used for this work. The results of this process are shown in Table 3.17.

As you can see, the baseline prior showed the smallest ESS value of about 3 and the priors that were centered on the appropriate side of zero with drastically narrowed standard deviations had the largest ESS value of about 60. This confirms our results from Section 3.4.3, in that we saw estimates and spreads that very closely resembled the priors used on the parameters; the large ESS for this prior specification gives further verification that our prior is driving our results rather than the data. Also, for the Centered priors (in Section 3.4.1), we see a relatively small ESS value (about 6) however, the results showed that the priors had an adverse effect on our estimates in terms of bias and precision. This new information tells us that the prior is driving our results more-so than in the case of the baseline model, however it's magnitude may not be fully representative of the problems we saw in the results.

CHAPTER FOUR

Bayesian Misclassification with Differentially Correlated Binary Outcomes

4.1 Introduction

Considerable research has been done on the impact of misclassification on inference in binomial regression models. Misclassification can lead to statistical inefficiency and biased results, as discussed by Barron (1977), Copeland et al. (1977), Neuhaus (1999), and Carroll et al. (2006) Many researchers have proposed methods to correct for response misclassification. For example, in the study of ordinary logistic regression using validation data or the assumption of known misclassification probabilities (refer to Magder and Hughes (1997), Morrissey and Spiegelman (1999), Carroll et al. (2006), Green (1983), Greenland (1988), Marshall (1990), Brenner and Gefeller (1993)).

Ample work has been done using a frequentist framework to estimate regression coefficients that incorporates validation data for a differentially misclassified response variable (Carroll et al. (2006) and Holcroft et al. (1997)). Pepe (1992) use non-parametric kernel methods while Lyles et al. (2011) develop computationally straightforward methods that implements a maximum likelihood approach. The Bayesian approach has also been considered; Paulino et al. (2003) extend work by Bedrick et al. (1996) to allow for model selection in binomial regression when the response variable is subject to an unconstrained (in terms of assumed sensitivity and specificity values) misclassification process. Goldstein et al. (2016) use a case-control study to account for non-differential misclassification but used fixed values of the sensitivity and specificity provided by "expert-opinion" to drive their research. McInturff et al. (2004) utilize conditional means priors to allow for prior information and expert opinion to help provide estimates of diagnostic sensitivity and specificity from a binomial regression suffering from response misclassification. Gerlach and Stamey (2007) explore the effects of both differential and non-differential misclassification

in the context of logistic regression when the response is misclassified using internal validation data. All of this research has focused on the case in which the response measurement is measured only once.

Repeated response measurement is commonly used in medical data; patients are followed for extended periods of time and measured more than once over the study. Neuhaus et al. (2002) use frequentist methods when misclassification probabilities are either known or unknown for population averaged generalized estimating equations and cluster-specific generalized linear mixed models. Their work centers on probabilities that were fixed and independent of covariates because although maximum likelihood estimates could be obtained in theory, practically there exist identifiability problems without more assumptions being placed upon the analysis. Lyles et al. (2005) extend the idea of McNemar's test for matched pair 2×2 tables in a longitudinal study using both internal and external validation data for odds ratio estimates. Extending both of these contributions for longitudinal studies involving repeatedly measured error prone responses, Tang et al. (2015b) use internal validation subsamples at different study time points to find valid and computationally efficient maximum likelihood estimates of regression parameters.

In this chapter we extend the work by Tang et al. (2015b) to the Bayesian approach. We mirror their frequentist methods and develop corresponding Bayesian methods (as described in section 4.3.4) to motivate a comparison between the two approaches. We aim, as they did, to find methods that are computationally efficient and to show the importance of accounting for differential misclassification. This chapter is organized as follows: in Section 4.2 we introduce the frequentist methods proposed by Tang et al. (2015b) and the four models used to convey the validity of the assumption of differential misclassification. In Section 4.3 we describe the process by which we analyze our simulated data under the frequentist and Bayesian approaches. We lastly discuss the results and the comparisons of those results for both approaches in 4.4.

4.2 Frequentist Methods for Differential Misclassification

Suppose we wish to examine the results of a test given to patients over repeated visits to a clinic. Let Y_{ij} denote the gold-standard response measurement for subject i at time point j and denote Y_{ij}^* as the error prone measurement for subject i at time point j. Here the gold-standard response, Y_{ij}, is assumed to be prohibitively expensive in time or resources while the error prone response, Y_{ij}^*, is readily available. We will let our covariates be denoted as X_{ip}, where p is the total number of covariates relating to the response. The covariates are not dependent on the time of the visit but are dependent upon the subject. We denote the total number of subjects as n, however, the amount of validated data and main study data can vary depending on the validation size setting. We denote the validated data sample size as n_v and our main data sample size (data that has not been validated) as n_m.

To incorporate correlation between the visits, we include a correlated random error structure in the models, denoted by u_i. This is expanded upon in Section 4.2.1. Further, we assume that the responses within each subject, i, are conditionally independent, so that as Breslow and Clayton (1993) find, the likelihood can be fully specified from our coefficients and our subject specific random effect. For generalized linear mixed models, it is commonly assumed that $(u_i, u_i^*)^T \sim N(\mathbf{0}, \mathbf{\Sigma})$; we will assume this here without loss of generality.

We consider four models with a correlated structure: "General Misclassification" where we assume correlated binary outcomes have dependent differential misclassification (GEN), independent correlated differential (ICD) misclassification, independent uncorrelated differential (IUD) misclassification, and completely non-differential (ND) misclassification. For each model, we assume the generalized logistic models are appropriate, although, other link functions could be used instead of the "logit" function.

4.2.1 General Misclassification (GEN)

The assumption under "General Misclassification" is that the outcomes are correlated binary responses and that there exists dependent differential misclassification in the

responses. To account for the misclassification, we use two logistic regression models to represent our response model, the first of which is:

$$\text{logit}\{Pr(Y_{ij} = 1|X_{i1} \dots X_{iP}, u_i)\} = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ip} + u_i. \tag{4.1}$$

Equation 4.1 is referred to as the response or "main" model, since it is a model for the gold-standard measurement. The other logistic model incorporates the misclassification and its dependence on the individual covariate information. Here, we note that the covariates may not be the same as they were for the main model, thus we define C_{iq} as the covariates for the misclassified response. This gives:

$$\text{logit}\{Pr(Y^*_{ij} = 1|C_{i1} \dots C_{iQ}, Y_{ij}, u^*_i)\} = \gamma_0 + \sum_{q=1}^{Q} \gamma_q C_{iq} + \gamma_{Q+1} Y_{ij} + u^*_i. \tag{4.2}$$

We refer to equation 4.2 as the "joint" model since it incorporates knowledge of both the error prone and the gold-standard response variable. To ease our notation, we will denote $\mathbf{C}_{iq} = (C_{i1} \dots C_{iQ})$ and, likewise, $\mathbf{X}_{ip} = (X_{i1} \dots X_{iP})$.

Using the joint model, we define the sensitivity and specificity as:

$$SE(\mathbf{C}_{iq}) = Pr(Y^*_{ij} = 1|Y_{ij} = 1, \mathbf{C}_{iq}, u^*_i) \tag{4.3}$$

$$SP(\mathbf{C}_{iq}) = Pr(Y^*_{ij} = 0|Y_{ij} = 0, \mathbf{C}_{iq}, u^*_i). \tag{4.4}$$

Through these models we allow for the misclassification to be correlated within the same subject in addition to the response. Thus, we must also define the assumed structure for the subject-specific random effect vector that will allow for correlation among the models:

$$(u_i, u^*_i)^T \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma^2_{u_i} & \psi \\ \psi & \sigma^2_{u^*_i} \end{pmatrix}\right]. \tag{4.5}$$

Using the random error structure in 4.5, we have that the likelihood contribution for subject i at time j is:

$$Pr(Y^*_{ij}|\mathbf{X}_{ip}, \mathbf{C}_{iq}, u_i, u^*_i) = \sum_{Y_{ij}=0}^{1} Pr(Y^*_{ij}|Y_{ij}, \mathbf{C}_{iq}, u^*_i) Pr(Y_{ij}|\mathbf{X}_{ip}, u_i) \tag{4.6}$$

This can then be further broken down depending on whether validation data is available.

The first formula is associated with the main data, where the true value of y_{ij} is unknown; this gives:

$$\begin{aligned} Pr(Y^*_{ij}|\mathbf{X}_{ip}, \mathbf{C}_{iq}, u_i, u^*_i) =& Pr(Y^*_{ij}|Y_{ij}=1, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=1|\mathbf{X}_{ip}, u_i) \\ &+ Pr(Y^*_{ij}|Y_{ij}=0, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=0|\mathbf{X}_{ip}, u_i) \\ =& [Pr(Y^*_{ij}=1|Y_{ij}=1, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=1|\mathbf{X}_{ip}, u_i) \\ &+ Pr(Y^*_{ij}=1|Y_{ij}=0, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=0|\mathbf{X}_{ip}, u_i)]^{Y^*_{ij}} \\ &[Pr(Y^*_{ij}=0|Y_{ij}=1, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=1|\mathbf{X}_{ip}, u_i) \\ &+ Pr(Y^*_{ij}=0|Y_{ij}=0, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=0|\mathbf{X}_{ip}, u_i)]^{(1-Y^*_{ij})}. \end{aligned} \tag{4.7}$$

Equation 4.7, is entirely dependent on the value of Y^*_{ij}; the unobserved true values are summed out.

The second formula is associated with the validated data, where the true value of y_{ij} is known, which gives:

$$\begin{aligned} Pr(Y^*_{ij}|\mathbf{X}_{ip}, \mathbf{C}_{iq}, u_i, u^*_i) =& [Pr(Y^*_{ij}=0|Y_{ij}=0, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=0|\mathbf{X}_{ip}, u_i)]^{(1-Y^*_{ij})(1-Y_{ij})} \\ &[Pr(Y^*_{ij}=1|Y_{ij}=0, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=0|\mathbf{X}_{ip}, u_i)]^{Y^*_{ij}(1-Y_{ij})} \\ &[Pr(Y^*_{ij}=0|Y_{ij}=1, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=1|\mathbf{X}_{ip}, u_i)]^{(1-Y^*_{ij})Y_{ij}} \\ &[Pr(Y^*_{ij}=1|Y_{ij}=1, \mathbf{C}_{iq}, u^*_i)Pr(Y_{ij}=1|\mathbf{X}_{ip}, u_i)]^{Y^*_{ij}Y_{ij}}. \end{aligned} \tag{4.8}$$

You can see that equation 4.8 depends on both the values of Y_{ij} and Y^*_{ij} in that for each validated subject for each time, only one of the four components of this model equation will contribute to the likelihood at a time.

From 4.7 and 4.8 we can develop the likelihoods in terms of the sensitivity and specificity (4.3 and 4.4); we have two components that make up the likelihood. We denote the validation portion of the likelihood as L_v where we have both the error prone and gold-standard responses available. L_m denotes the likelihood for the main data, in which we

only have the error prone response measurements available. These equations can be broken down into the subject and time specific contributions and are defined as follows:

$$\begin{aligned} L_m =&[SE(\mathbf{C}_{iq})Pr(Y_{ij} = y_{ij}|\mathbf{X}_{ip}, u_i)+ \\ &(1 - SP(\mathbf{C}_{iq}))(1 - Pr(Y_{ij} = y_{ij}|\mathbf{X}_{ip}, u_i))]^{y^*_{ij}} \\ &[(1 - SE(\mathbf{C}_{iq}))Pr(Y_{ij} = y_{ij}|\mathbf{X}_{ip}, u_i)+ \\ &SP(\mathbf{C}_{iq})(1 - Pr(Y_{ij} = y_{ij}|\mathbf{X}_{ip}, u_i))]^{(1-y^*_{ij})} f(u_i, u^*_i) \end{aligned} \tag{4.9}$$

$$\begin{aligned} L_v =&[SE(\mathbf{C}_{iq})Pr(Y_{ij} = y_{ij}|\mathbf{X}_{ip}, u_i)]^{y^*_{ij}y_{ij}} \\ &[(1 - SE(\mathbf{C}_{iq}))Pr(Y_{ij} = y_{ij}|\mathbf{X}_{ip}, u_i)]^{(1-y^*_{ij})y_{ij}} \\ &[(1 - SP(\mathbf{C}_{iq}))(1 - Pr(Y_{ij} = y_{ij}|\mathbf{X}_{ip}, u_i))]^{y^*_{ij}(1-y_{ij})} \\ &[SP(\mathbf{C}_{iq}))(1 - Pr(Y_{ij} = y_{ij}|\mathbf{X}_{ip}, u_i))]^{(1-y^*_{ij})(1-y_{ij})} f(u_i, u^*_i) \end{aligned} \tag{4.10}$$

The resulting likelihood is thus:

$$L \propto L_m^{(1-val)} L_v^{val} \tag{4.11}$$

where val is an indicator for whether or not the subject/time observation has been validated or not. We next examine some of the changes to this model when the assumptions are modified.

4.2.2 Independent Correlated Differential Misclassification (ICD)

Independent correlated differential misclassification (ICD) relaxes the assumptions of our general misclassification model (models 4.1 and 4.2). For this model, the misclassification process is assumed to be correlated within the same subject, however the subject specific random effects, u_i and u^*_i, are independent. For this, we have that the distribution of our random error terms is as follows:

$$(u_i, u^*_i)^T \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma^2_{u_i} & 0 \\ 0 & \sigma^2_{u^*_i} \end{pmatrix}\right]$$

4.2.3 *Independent Uncorrelated Differential Misclassification (IUD)*

Independent uncorrelated differential misclassification (IUD) relaxes the assumptions of the GEN and ICD models (models 4.1 and 4.2 with the error structure of the ICD model). For this model, the random errors are assumed to be independent and uncorrelated. We now omit u_i^* from the joint model (4.2), yielding the following model definition:

$$\text{logit}\{Pr(Y_{ij}^* = 1|C_{i1} \dots C_{iQ}, Y_{ij}, u_i^*)\} = \gamma_0 + \sum_{q=1}^{Q} \gamma_q C_{iq} + \gamma_{Q+1} Y_{ij}. \tag{4.12}$$

Our main model is still as described for the GEN and IUD models (model 4.1), although the distribution of our random error is now $u_i \sim N(0, \sigma^2)$.

4.2.4 *Non-Differential Misclassification (ND)*

Non-differential misclassification relaxes the assumptions under our model of general misclassification (models 4.1 and 4.2) considerably. We now assume that there is no correlation between the two models, and that the covariates do not have an impact on the error prone response. This model reduces the original model equations to:

$$\text{logit}\{Pr(Y_{ij} = 1|\mathbf{X}_{ip}, u_i)\} = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ip} + u_i \tag{4.13}$$

$$\text{logit}\{Pr(Y_{ij}^* = 1|Y_{ij})\} = \gamma_0 + \gamma_1 Y_{ij} \tag{4.14}$$

Notice that the joint model no longer includes $\mathbf{C_{iq}}$ or u_i^*.

4.3 *Simulation Study based on Frequentist Methods*

A simulation was performed to analyze the differences in the two methods using all four models described in Section 4.2. This simulation study was based upon the potential real-world scenario of a study in which patients are tested for a certain disease over a long period of time, in which several visits to a clinic are performed. There are two tests for this disease, a very expensive gold-standard test, which is only given to a small subsample of the study participants, and a cheap, but error-prone test which is given to all study participants. Thus, some study participants will be "validated" in that they will be tested using both tests

at each visit. Each visit involves the patient being re-tested for the disease in question and certain covariates being noted (e.g. gender, age at first visit, race, etc.).

We set $J = 4$, thus each subject visited the clinic and was tested at four distinct times. Suppose we have two covariates: a binary covariate generated from a Bernoulli distribution with probability of success as 0.5, and a continuous covariate generated from a Normal distribution with a mean of 0 and a standard deviation of 4 (variance of 16). The covariates are not dependent on the time of the visit but are, clearly, dependent upon the subject. As for the number of subjects, we will use an overall sample size of $n = 1,000$; the amount of validated data and main data can vary depending on the validation size setting. For this example, we will use a validation size of 20%. We will denote the validated data sample size as $n_v = 200$ and our main data sample size (data that has not been validated) as $n_m = 800$.

4.3.1 Simulation Specific Frequentist Models

Based upon this study's specifics, our models from Section 4.2 can be adapted to fit the data at hand. The likelihood structure does not change, only the joint model, main model, and random error structure change depending on the model.

"General Misclassification" (GEN) assumes the outcomes are correlated binary responses subject to dependent differential misclassification. For this, we assume the response is only related to our two covariates, X_1 and X_2, so our main model is:

$$\text{logit}\{Pr(Y_{ij} = 1|\mathbf{X}_{ip}, u_i)\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i. \tag{4.15}$$

We also assume that our error prone response model is dependent upon the time of the visit, and only the first covariate, X_1. We will let our reference time point be the first visit and we incorporate the visit information using indicator functions. Using this, we can define our joint model as:

$$\begin{aligned}\text{logit}\{Pr(Y^*_{ij} = 1|\mathbf{C}_{iq}, Y_{ij}, u^*_i)\} = \gamma_0 &+ \gamma_1 I(T_{ij} = 2) + \gamma_2 I(T_{ij} = 3) \\ &+ \gamma_3 I(T_{ij} = 4) + \gamma_4 X_{i1} + \gamma_5 Y_{ij} + u^*_i. \end{aligned} \tag{4.16}$$

We assume that the errors are dependent, thus the distribution of our random error terms is $(u_i, u_i^*)^T \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u_i} & \psi \\ \psi & \sigma^2_{u_i^*} \end{pmatrix}\right)$.

Independent correlated differential misclassification (ICD) assumes the misclassification process is correlated within the same subject, however the subject specific random effects, u_i and u_i^*, are independent. For this, we have that the distribution of our random error terms is $(u_i, u_i^*)^T \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u_i} & 0 \\ 0 & \sigma^2_{u_i^*} \end{pmatrix}\right)$, and we have the same joint and main models as in the GEN model (models 4.15 and 4.2).

Independent uncorrelated differential misclassification (IUD) relaxes the assumptions of the GEN and ICD models giving us the following model definition:

$$\begin{aligned} \text{logit}\{Pr(Y_{ij}^* = 1|\mathbf{T}, \mathbf{X}_{ip}, Y_{ij})\} =& \gamma_0 + \gamma_1 I(T_{ij} = 2) + \gamma_2 I(T_{ij} = 3) \\ &+ \gamma_3 I(T_{ij} = 4) + \gamma_4 X_{i1} + \gamma_5 Y_{ij}. \end{aligned} \tag{4.17}$$

Notice that this joint model no longer incorporates the random error related to the error-prone response measurement (u_i^*). The main model for IUD misclassification is still as described for the GEN and ICD models (model 4.1), although the distribution of our random error is now $u_i \sim N(0, \sigma^2)$.

For the non-differential misclassification models, our joint and main models become:

$$\text{logit}\{Pr(Y_{ij} = 1|\mathbf{X}_{ip}, u_i)\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i \tag{4.18}$$

$$\text{logit}\{Pr(Y_{ij}^* = 1|Y_{ij})\} = \gamma_0 + \gamma_5 Y_{ij}. \tag{4.19}$$

The joint model, like the IUD joint model, does not include the correlated random error term, u_i^*, nor do we include the time point indicator functions. Again, the distribution of our random error is $u_i \sim N(0, \sigma^2)$.

4.3.2 Data Generation and Frequentist Analysis Process

To begin our simulations, we first generate datasets to use; the generation process is as follows:

(1) Set the true estimates of the parameters based on general examples:

(a) $\beta_0 = 0$, $\beta_0 = 1$, $\beta_0 = 0.5$,

(b) $\gamma_0 = -3$, $\gamma_1 = 0.05$, $\gamma_2 = 0.2$, $\gamma_3 = 0.4$, $\gamma_4 = 1.5$, $\gamma_5 = 3$

(c) $\sigma^2_{u_i} = 1$, $\sigma^2_{u_i^*} = 1$, and $\psi = 0.5$.

(2) Generate $n = 1,000$ values of $X_{i1} \sim Bern(0.5)$ and $X_{i2} \sim N(0, \sigma = 4)$.

(3) Generate $(u_i, u_i^*)^T \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u_i}=1 & 0.5 \\ 0.5 & \sigma^2_{u_i}=1 \end{pmatrix}\right)$.

(4) Generate Y_{ij} and Y^*_{ij} following the models in Section 4.3.1 outlined under General Misclassification.

(5) Create an indicator vector for which subjects have been validated.

This dataset will be used in both the Frequentist and Bayesian analysis processes outlined next.

4.3.3 Frequentist Estimation

To estimate our parameters under the frequentist paradigm we use numerical integration to integrate out the random effects in our models. Utilizing SAS/STAT® software and the `NLMIXED` (1999) procedure we can then maximize the likelihood with quasi-Newton optimization. To use these methods, we can supply initial values for our parameters. Initial values are set assuming we would have some intuition about how these parameters would behave; this intuition could come from a variety of sources like previous studies or experts within the field we are working in.

The overall process of simulation for the frequentist method is:

(1) Generate a full dataset as described in Section 4.3.

(2) We use initial values for the parameters that are on the appropriate side of zero and within a similar magnitude to the truth:

(a) $\beta_0 = 0$, $\beta_1 = 2$, $\beta_2 = 1$,

(b) $\gamma_0 = -2$, $\gamma_1 = 0$, $\gamma_2 = 0.5$, $\gamma_3 = 0.5$, $\gamma_4 = 2$, $\gamma_5 = 1$

(c) $\sigma^2_{u_i} = 0.5$, $\sigma^2_{u_i^*} = 0.5$, and $\psi = 0.1$.

(3) Analyze the data under each of the models described in section 4.2.

(4) Compare the estimated parameters to the "true estimates" used to generate the data in specified in Section 4.3 under step 1.

(5) Repeat these steps 500 times and report the mean, standard error, coverage, and convergence rate for each parameter.

4.3.4 Bayesian Priors and Estimation

The Bayesian approach requires prior distributions for all parameters. We focus on mildly informative priors that do not allow for highly unlikely, extreme values to be probable.

Similar to the frequentist methods described in Section 4.2, we have chosen to use a logistic link function for our regression components (eqs. 4.15 and 4.16). This type of function will predict response estimates that range between zero and one, when transformed appropriately, to represent a probability of occurrence. Objectively (before data has been collected/analyzed), a one-unit change in a predictor would rarely result in a response level change from 0.01 (1% chance of a positive response) to 0.99 (99% chance of a positive response). This change in the response is equivalent to the appropriate back-transformed potential coefficient estimate of a magnitude of 10 that Gelman et al. discuss (2008). An effect size this large is highly unlikely, thus our prior suitably keeps low probability on this possibility.

For each of the models described in the frequentist method (Section 4.2), we will add in prior information and update our prior beliefs via Bayes theorem. For this method, we will use the same general prior for all coefficient parameters involved:

$$\beta_0, \ldots \beta_2, \gamma_0, \ldots \gamma_5 \sim N(0, \sigma^2 = 10). \tag{4.20}$$

In addition to the priors needed for the coefficients on the predictors, we also require prior distributions for the random error components of the models. Recall that in section 4.2 we

described four different misclassification models, each of which had a different random error structure assumption.

The simplest model (the model with the fewest predictors and the simplest random error structure) is the non-differential misclassification (ND) model. Here, we are assuming that there is only random error in the model to predict our gold-standard response, model 4.18. Since we would not expect that by random error alone our response, when appropriately transformed, would increase from 1% to 50% (or result in an odds ratio increase of 5), we limit our prior distribution to:

$$\begin{aligned} u_i &\sim N(0, \sigma^2), \\ \frac{1}{\sigma^2} &\sim Gamma(2, 2). \end{aligned} \tag{4.21}$$

These distributions allow for considerable variability, but still within a realm of practical possibility. We use the Gamma distribution because it limits our responses to only positive values (precision cannot be negative) and limits the probability of getting extremely small or large values for the variance of u_i. See figure 4.3.1 for the density of this prior distribution for the precision. We could have chosen to use a Uniform(0, Max) prior on the standard deviations, however this is an improper prior and relies heavily on the choice of the maximum value except for arbitrarily large values (see Gelman 2004). Likewise, we could have used a half-Cauchy, although the limiting distribution becomes the Uniform distribution proving to, again, depend on the value of the parameter, in this case relating to scale. Therefore, we will use this same prior distribution (distribution 4.21) for the independent uncorrelated differential (IUD) case, which has the assumption that our main model is now reliant upon covariate information.

For the correlated models, our priors need to be altered slightly due to the fact that each of the models introduces a new correlated random error structure, thus the need for a multivariate prior distribution. Recall, we use two regression models here utilizing $J = 4$

time-points with the first time-point as our reference:

$$\text{logit}\{Pr(Y_{ij}=1|\mathbf{X}_{ip},u_i)\} = \beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+u_i \tag{4.22}$$

$$\begin{aligned}\text{logit}\{Pr(Y^*_{ij}=1|\mathbf{T},\mathbf{X}_{ip},Y_{ij},u^*_i)\} =& \gamma_0+\gamma_1 I(T_{ij}=2)+\gamma_2 I(T_{ij}=3)\\ &+\gamma_3 I(T_{ij}=4)+\gamma_4 X_{i1}+\gamma_5 Y_{ij}+u^*_i.\end{aligned} \tag{4.23}$$

For the independent correlated differential misclassification (ICD) models, we need a multivariate prior on the error terms for both the joint and main models shown above that incorporates the assumption of independence between the error terms but still allows for some correlation. Thus, we will again use a gamma distribution on the precisions, but will combine this information as hyper-priors to a multivariate normal distribution, as follows:

$$\frac{1}{\sigma^2_{u_i}} \sim Gamma(2,2) \tag{4.24}$$

$$\frac{1}{\sigma^2_{u^*_i}} \sim Gamma(2,2) \tag{4.25}$$

$$(u_i,u^*_i)^T \sim N\left[\begin{pmatrix}0\\0\end{pmatrix},\ \begin{pmatrix}\sigma^2_{u_i} & 0\\ 0 & \sigma^2_{u^*_i}\end{pmatrix}\right]. \tag{4.26}$$

The last model, dependent correlated differential misclassification (GEN), specifies one more assumption from the ICD model in that the errors are now assumed to be dependent upon one another. This will require that we have an added hyper-prior for the

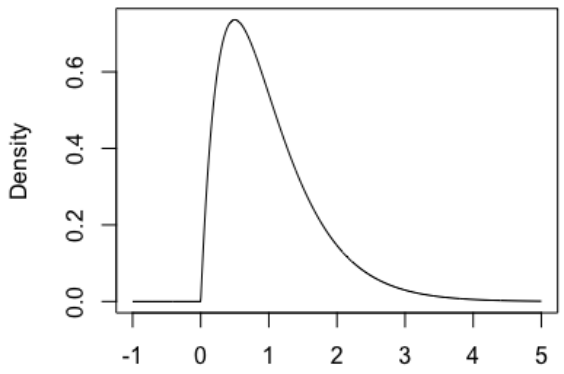

Figure 4.3.1. Bayesian Prior Distribution for $\frac{1}{\sigma^2} \sim Gamma(2,2)$

dependence between the two error terms. For this, we assume our priors to be:

$$\frac{1}{\sigma^2_{u_i}} \sim Gamma(2,2) \tag{4.27}$$

$$\frac{1}{\sigma^2_{u_i^*}} \sim Gamma(2,2) \tag{4.28}$$

$$\psi \sim Unif(-2,2), \tag{4.29}$$

$$(u_i, u_i^*)^T \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma^2_{u_i} & \psi\\ \psi & \sigma^2_{u_i^*}\end{pmatrix}\right], \tag{4.30}$$

where the added parameter, ψ, is used to represent the amount of covariance between the two random error components of the joint and main models.

Analysis for the Bayesian approach was completed using `NIMBLE` (2017), a package which uses Markov Chain Monte Carlo (MCMC) simulation approximate the posterior distribution. MCMC is, simply put, the idea that you can define a sequence of random vectors in which the vectors generated towards the end of this sequence tend towards a particular "posterior" distribution (Christensen et al. 2011). Understanding that the sequence may need a large number of vectors, a suitable set of vectors can be removed from the beginning of the sequence to provide more accurate estimations of the posterior distribution (burning in); since the sequence of vectors can be shown to be correlated, a pattern of vectors can be set to be kept versus discarded so as to make consecutive random vectors, nearly independent (thinning). More than one sequence can be run to ensure that each chain, while random in nature, behaves similarly in its posterior quantities; all of these principles can be checked and fine-tuned through convergence diagnostics.

4.3.5 Convergence Criteria and Bayesian Simulation Settings

Dataset convergence for the frequentist results varied between models: The ideal model had 479/500 converged datasets; the naïve model had 71/500 datasets that converged; the general model had 473/500 converged datasets; the ICD model had 485/500 converged datasets; the IUD model had 458/500 converged datasets; and the ND model

had 417/500 converged datasets. Datasets that did not converge were removed from the tabulation of our results. The naïve model had the worst rate of convergence, while the remainder of the models all had relatively large rates of convergence.

Using the Bayesian approach, convergence can be assessed using diagnostic plots. To do this, we examine three main plots: a marginal density plot, a trace plot, and an auto-correlation plot. Each of these needs to show certain qualities to demonstrate convergence. The marginal density plot should show a bell-shaped curve over the mode of the distribution, although it need not be symmetric. The trace plot should show "good mixing" over the mode of the distribution; this means we want to see no clear pattern to the values as they explore the parameter space. Lastly, the auto-correlation plot should show a decrease in correlation towards 0 as the lag increases.

For the sake of brevity, we will discuss the convergence plots for the GEN model only as it relates to our parameter of interest, β_1, and the variance of the random error term, $\sigma^2_{u_i}$. This model has the highest number of parameters to be estimated and should thus be the most difficult to meet convergence; if we can show convergence for this model, it should follow that we can establish convergence for the remaining, more simplified, models. Refer to Appendix A.8 and A.9 for a complete display of the convergence plots for all models discussed here.

Figure 4.3.2 shows the marginal density plots for β_1, $\tau_i = 1/\sigma^2_{u_i}$, and $\tau_i^* = 1/\sigma^2_{u_i^*}$ for the GEN model. Here we see that the densities do appear to be moderately bell-shaped, although in the cases of τ_i and τ_i^* we see that the densities are asymmetric (not surprising for a distribution of precision). The asymmetric distributions should not be a problem, as long as we see that the mode of the distribution is not an extreme value. In this case, we see that the mode for the distribution of τ_i and τ_i^* are about 1.10 and 0.90 respectively.

Figure 4.3.3 shows the trace plots for β_1, $\tau_i = 1/\sigma^2_{u_i}$, and $\tau_i^* = 1/\sigma^2_{u_i^*}$ for the GEN model. Here we see that the trace plot does not appear to have a discernible pattern for β_1. For the plots related to τ_i and τ_i^* we see that we have a skewed amount of data falling above

the modes of about 1.10 and 0.90, respectively. This skewness can be explained by the fact that we have constrained these values to be positive values only.

Figure 4.3.4 shows the auto-correlation plots for β_1, $\tau_i = 1/\sigma^2_{u_i}$, and $\tau^*_i = 1/\sigma^2_{u^*_i}$ for the GEN model. Here we see that the autocorrelations do tend to zero. As with the previous diagnostics, we have questionable convergence for both τ_i and τ^*_i since plot shows moderate autocorrelation. The autocorrelation plot for β_1 does decrease to zero for both chains, however there is still some variation above and below zero. Given the complexity of the models, these plots do not indicate a lack of convergence.

For these simulations we used two chains instead of one because we wanted to ensure convergence could be justified from our results. In the previous graphics you can see the benefit in the chains because each chain behaves similarly. For example, comparing the density plots in Figure 4.3.2 shows that both chains have a mode at about the same values. If

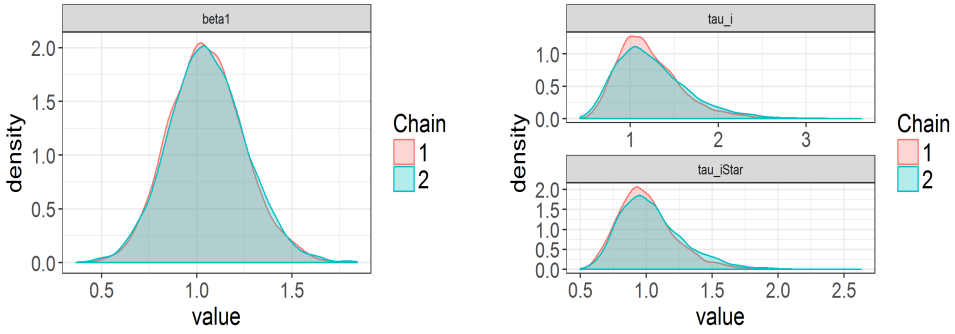

Figure 4.3.2. Density Plots for β_1, τ_i, and τ^*_i under the GEN Model

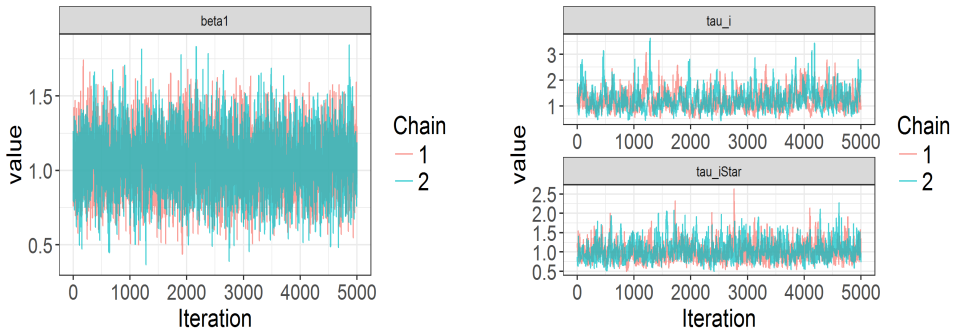

Figure 4.3.3. Trace Plots for β_1, τ_i, and τ^*_i under the GEN Model

Table 4.1. Settings for Bayesian Results for Correlated Binary Response Models

Model	Iterations	Burn-in	Thin	Chains
Ideal	200,000	185,000	3	2
Naïve	200,000	185,000	3	2
GEN	900,000	750,000	30	2
ICD	400,000	350,000	10	2
IUD	375,000	350,000	5	2
ND	350,000	330,000	4	2

we saw two different modes in our chains this would be a good indication that convergence had not been met. The same idea holds for the trace plots shown in Figure 4.3.3; here we see that each chain centers (without pattern) over similar values. If we had seen that the chains centered over different values, then this would have been a clear violation of convergence that would require us to adjust our settings or adjust our methods. Lastly, the auto-correlation plots in Figure 4.3.4 would show a violation if the two correlations did not fall to zero or if they behaved differently from one another. We see that the chains behave similarly and thus, again, the addition of a second chain is helpful to assess convergence.

Based upon these plots (and the plots for the remainder of models shown in Appendix A.8 to A.9) we have not seen a violation in convergence. For the Bayesian approach, unlike the frequentist method, all 500 iterations were used for each model under the settings described in Table 4.1.

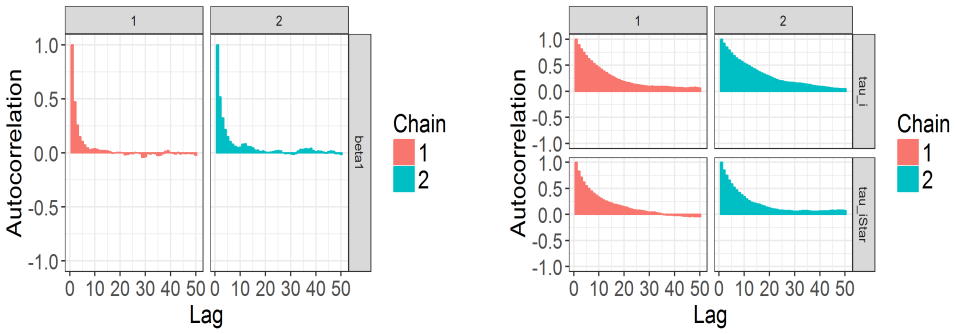

Figure 4.3.4. Auto-Correlation Plots for β_1, τ_i, and τ_i^* under the GEN Model

Table 4.2. Results Comparison for Bayesian and Frequentist Techniques Assuming Dependent Correlated Differential Misclassification with Validation Size of 20%

Model	Frequentist			Bayesian		
True Value	Estimate	(SE)	Coverage	Estimate	(SD)	Coverage
Ideal						
$\beta_0 = 0.000$	0.001	(0.076)	95.0	0.005	(0.078)	94.0
$\beta_1 = 1.000$	0.992	(0.112)	94.6	1.000	(0.114)	95.6
$\beta_2 = 0.500$	0.498	(0.021)	95.2	0.504	(0.021)	94.6
$\sigma^2_{u_i} = 1.000$	0.941	(0.152)	92.3	1.031	(0.172)	95.6
Naïve						
$\beta_0 = 0.000$	-1.007	(0.075)	0.0	-1.026	(0.077)	0.0
$\beta_1 = 1.000$	1.540	(0.105)	0.0	1.560	(0.108)	0.0
$\beta_2 = 0.500$	0.198	(0.014)	0.0	0.200	(0.014)	0.0
$\sigma^2_{u_i} = 1.000$	1.014	(0.138)	91.5	0.911	(0.125)	84.2
GEN						
$\beta_0 = 0.000$	0.001	(0.143)	94.7	0.003	(0.147)	94.4
$\beta_1 = 1.000$	1.007	(0.206)	95.2	1.032	(0.213)	93.6
$\beta_2 = 0.500$	0.501	(0.041)	96.2	0.510	(0.041)	96.8
$\sigma^2_{u_i} = 1.000$	0.964	(0.320)	94.1	1.117	(0.332)	97.0
$\sigma^2_{u_i^*} = 1.000$	0.996	(0.229)	94.1	1.035	(0.222)	97.8
$\psi = 0.500$	0.517	(0.168)	96.6	0.517	(0.244)	96.2
ICD						
$\beta_0 = 0.000$	0.037	(0.154)	94.7	0.026	(0.159)	94.2
$\beta_1 = 1.000$	0.984	(0.222)	96.3	1.026	(0.230)	94.6
$\beta_2 = 0.500$	0.502	(0.041)	96.1	0.512	(0.043)	96.6
$\sigma^2_{u_i} = 1.000$	1.119	(0.337)	96.1	1.300	(0.382)	91.6
$\sigma^2_{u_i^*} = 1.000$	1.292	(0.242)	83.5	1.287	(0.242)	79.2
IUD						
$\beta_0 = 0.000$	-0.005	(0.148)	94.8	-0.108	(0.179)	93.2
$\beta_1 = 1.000$	1.048	(0.210)	94.8	1.238	(0.256)	84.6
$\beta_2 = 0.500$	0.483	(0.039)	87.6	0.520	(0.046)	96.2
$\sigma^2_{u_i} = 1.000$	1.309	(0.272)	93.2	3.003	(0.639)	0.6
ND						
$\beta_0 = 0.000$	-0.751	(0.129)	0.0	-0.898	(0.161)	0.0
$\beta_1 = 1.000$	2.308	(0.197)	0.0	2.611	(0.253)	0.0
$\beta_2 = 0.500$	0.442	(0.034)	50.6	0.485	(0.042)	91.8
$\sigma^2_{u_i} = 1.000$	1.575	(0.280)	67.9	3.243	(0.628)	0.0

4.4 Comparing Frequentist and Bayesian Methods

Table 4.2 shows the results from each method for all β parameters and the parameters denoting the variability in our random error terms ($\sigma^2_{u_i}$, $\sigma^2_{u_i^*}$, or ψ) for both methods.

For the Ideal model we see that the estimates for both methods fall relatively close to the true values (small bias). The frequentist standard error is slightly smaller than the Bayesian standard deviations for all parameters. The coverages for the Bayesian method all hover around the nominal level of 95% whereas the frequentist methods do have one coverage probability slightly below nominal at 92.3%.

For the Naïve model we see that the estimates for both methods are very biased. We see slightly smaller frequentist standard errors and Bayesian standard deviations than we saw in the Ideal model. The coverages for β_0, β_1, and β_2 are zero, regardless of the method. The coverage for $\sigma^2_{u_i}$ is still below nominal, but is around 92% for the frequentist approach and 84% for the Bayesian. Thus, the naïve model is poor for both the Bayesian and frequentist approach.

For the GEN model we see that both methods have estimates close to the true values (small bias); we see a bit more bias in the Bayesian estimates however, the magnitude of this offset is not so much that the results should be regarded as inferior. Again, the frequentist standard error is slightly smaller than the Bayesian standard deviations for all but one of the parameters, $\sigma^2_{u_i^*}$. The coverages for both methods hover near the nominal level of 95% with the frequentist coverages a bit closer than the Bayesian coverages overall.

For the ICD model we see that the estimates for both methods have slightly more bias than in the GEN model. The Bayesian standard errors are slightly larger than the frequentist standard deviations. The most dramatic change is in the coverage probabilities. We now see that the coverage for $\sigma^2_{u_i^*}$ is much less than the nominal coverage of 95% for both methods. We see a discrepancy between the frequentist and Bayesian coverages for $\sigma^2_{u_i}$. The frequentist methods have slightly better coverage than the Bayesian approach; however, recall this model is incorrect for the data.

For the IUD model we see slightly more bias in the parameter estimates than in previous models. The standard errors and deviations are similar in magnitude to previous models. Interestingly, both of these statements are much more dramatic for the Bayesian es-

timate of $\sigma^2_{u_i}$ under this model; our estimate is now much larger than the truth ($\hat{\sigma}^2_{u_i} = 3.003$ vs. $\sigma^2_{u_i} = 1.000$). The standard deviation (0.639) is now much larger than the previous model (0.382) as well as the frequentist standard error for the same model (0.272). Interestingly, each method has a coverage probability for a regression parameter that is markedly lower than the others; the frequentist method has a smaller than nominal coverage for β_2, while the Bayesian method has a smaller than nominal coverage for β_1. The Bayesian method also, unsurprisingly because of the biased estimates, has a coverage probability of 0.6% for the estimate of $\sigma^2_{u_i}$. This means only 3 out of the 500 iterations actually covered the true value of $\sigma^2_{u_i}$.

For the ND model we see very obvious problems with both methods. Both methods provide extremely biased estimates for all four parameters shown. The magnitudes of the standard errors and deviations are similar to the respective spreads in the IUD models. The coverages are now zero in the frequentist method for both β_0 and β_1; the coverages for β_2 and $\sigma^2_{u_i}$ are not much better at 50.6% and 67.9%, respectively. For the Bayesian coverage probabilities, the only parameter with a non-zero coverage is β_2, which has a surprisingly high coverage of 91.8%.

We can also utilize model selection criteria to assess whether the correct model (the GEN model) is proving to be the "best" model beyond simply the comparison of parameter estimates. The frequentist methods use Akaike Information Criteria (AIC) to assess models; this statistic allows models to be compared to another and the "best" model (or the model to be selected) is the model with the lowest AIC value. Likewise, for the Bayesian methods you can use the Widely Applicable Information Criteria (WAIC; refer to Watanabe (2010) for a thorough discussion of this statistic) to compare and select models. Again, the model with the lowest WAIC is selected as the "best" model.

Interestingly, in a study of 20 datasets, the frequentist approach selected the ICD model 14 times and in each case the GEN model had a slightly higher AIC value. The GEN model was only selected 2 times; the remaining 4 datasets had convergence problems

under the frequentist approach and resulted in the IUD and ND models being selected twice, each. The Bayesian approach, on the other hand, used a model diagnostic that hinges upon the data correlated or not (refer to Ando and Tsay, 2010); for this, our models have varying correlation assumptions and thus the four models must be grouped based off of the assumed correlation structure. Under this approach we compare the GEN and ICD models to one another and separately compare the IUD and ND models to one another. Here we see that in all 20 of the models the Bayesian approach selected the GEN model over the ICD model, and the IUD model over the ND model.

The entirety of these results show that the Bayesian methods do just as well as the frequentist methods in model estimation, but completely out-perform the frequentist methods in terms of model selection ability. We also see that a slight adjustment in the assumptions governing our analysis models results in biased or imprecise estimates for both methods. As Tang et al. (2015b) showed for frequentist methods only, we see that the assumption of dependence and correlation between the two error terms is a required assumption to produced accurate and precise estimates based on data of this kind for both analysis approaches. While the effect is not necessarily strong for our logistic regression parameters of interest, we should still be concerned that the estimate of the error is affected by this dropped assumption (as seen in the changes from the GEN model to the ICD model results on $\sigma^2_{u_i^*}$).

CHAPTER FIVE

Conclusions

In this dissertation we examined the effects of analysis approach on misclassification with partially validated data. We studied two main scenarios, one with both the response and exposure being misclassified and one with correlated responses that are subject to misclassification. We introduced both frequentist and Bayesian approaches for inference on logistic regression parameters that accounted for misclassification. We discussed the Bayesian priors needed for all models considered and thoroughly discussed convergence diagnostics relating to the models. Finally, we varied the original settings for each misclassification scenario to substantiate the need for our assumptions.

In Chapter Two we expanded upon work by Tang et al. (2015a) to account for misclassification that can affect binary logistic regression. We used the HIV Epidemiology Research Study (HERS) data as a template for a simulation study based on misclassification in both the response and the exposure variables (Smith et al. 1997). We examined the differences between the frequentist and Bayesian approaches assuming dependent differential misclassification under a variety of validation sizes. We found that as the validation size decreased, the diffuse priors used in the Bayesian approach greatly benefited the results compared to the frequentist results which were inadequate to handle the reduced sizes.

In Chapter Three we ran the simulation study from Chapter Two under a selection of different settings. We changed the underlying assumption of dependent differential misclassification to independent differential misclassification. The premise was to understand the effect of assuming there exists a relationship between the error prone exposure measurement and the gold-standard response measurement. We saw that this relationship, while counter-intuitive, does provide adequate information in the estimation of the parameters for the response model, and thus this assumption is necessary. We then increased the over-

all sample size ten-fold; our aim was to understand whether the results from Chapter Two were dependent solely on the size of the validation data, or if they also depended on the overall sample size. We learned that increasing the overall sample size diminished the effect of decreasing the validation size, however the Bayesian approach still provided less bias and more precise estimates than the frequentist approach, especially as the validation size decreased. Lastly, this chapter explored the effect of prior information on the Bayesian approach. We found that if we could incorporate expert opinion or prior study information to not only center our priors but reduce the variation in our prior distributions, we could greatly improve the bias and precision in our estimates. We also found that if we greatly reduced the variation in our prior distributions we provided priors that were "too informative" in that the posteriors began to reflect the prior distribution rather than the data.

In Chapter Four we studied the setting in which the response variable is correlated over time; i.e. we had multiple visits from the same patient and measured the response and covariates each time. We explored the effect of correlated responses with partially validated data and dependent errors using both frequentist and Bayesian methods. The frequentist methods in this chapter were originally developed by Tang et al. (2015b). The results of our simulation study show that the Bayesian approach does just as well as the frequentist methods in terms of parameter estimation. We also found that a slight adjustment in the assumptions governing our analysis models results in biased or imprecise estimates for both methods. We learned that the assumption of dependence and correlation between the error terms is a required assumption to produce accurate and precise estimates for both analysis methods. In terms of appropriate model selection, we found that the Bayesian approach out-performed the frequentist approach by effectively selecting the correct model more often than the frequentist approach.

In the future, we will extend the work done in Chapter Four; we will test the models under a multitude of scenarios for which we will let the validation size vary, the overall sample size vary, and explore more informative priors. We will also examine the intertwin-

ing of the two scenarios described in Chapter Two and Chapter Four to understand how to adequately account for misclassification when the response is correlated, misclassified, and also dependent upon a misclassified exposure. This scenario has not been studied under either the frequentist or Bayesian approach.

APPENDICES

APPENDIX A

Convergence Diagnostics

A.1 Bayesian Analysis of Simulation Study based on work from Tang, et al[38] Discussed in Chapter Two

(A) 25% Validation Size; Iterations: 14,000 Burn-in: 4,000 Thin: 1

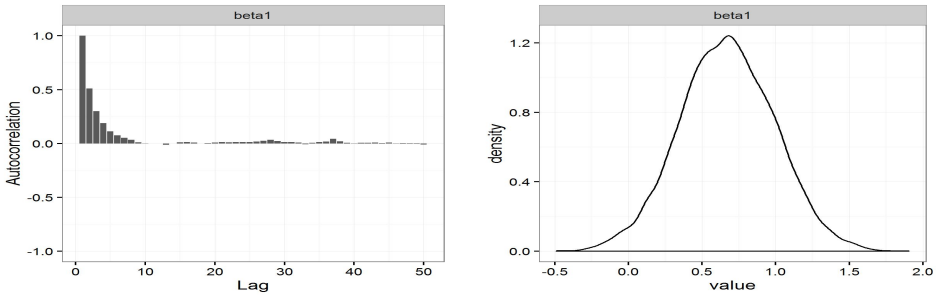

Figure A.1.1. Auto-correlation and density plot for 25% validation sizes

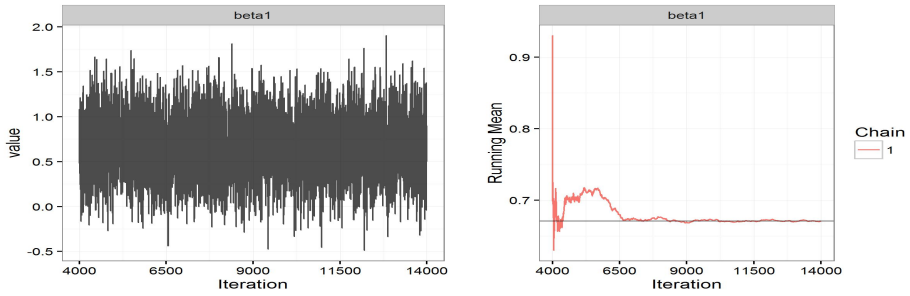

Figure A.1.2. Trace and running means plot for 25% validation sizes

(B) 15% Validation Size; Iterations: 27,000 Burn-in: 7,000 Thin: 2

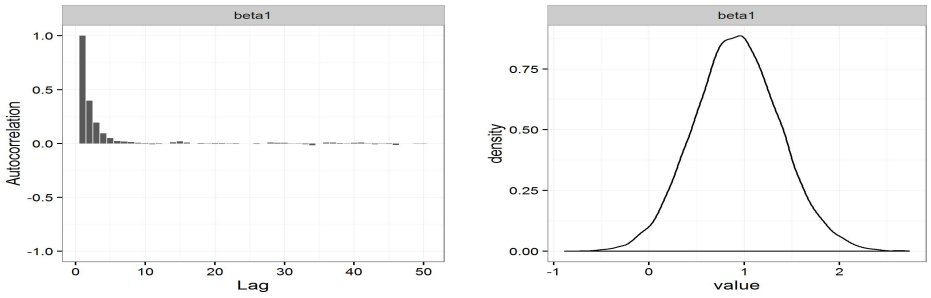

Figure A.1.3. Auto-correlation and density plot for 15% validation sizes

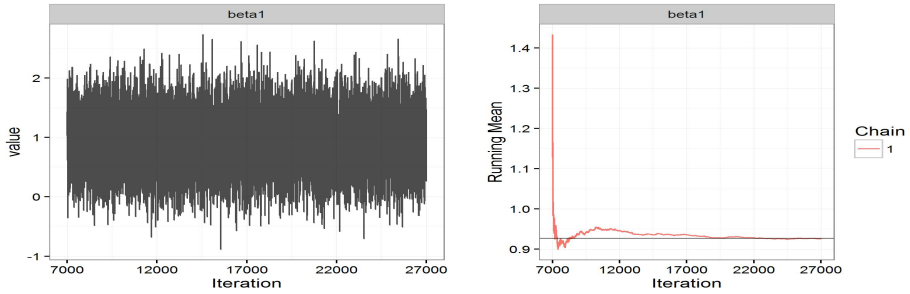

Figure A.1.4. Trace and running means plot for 15% validation sizes

(C) 10% Validation Size; Iterations: 39,000 Burn-in: 9,000 Thin: 3

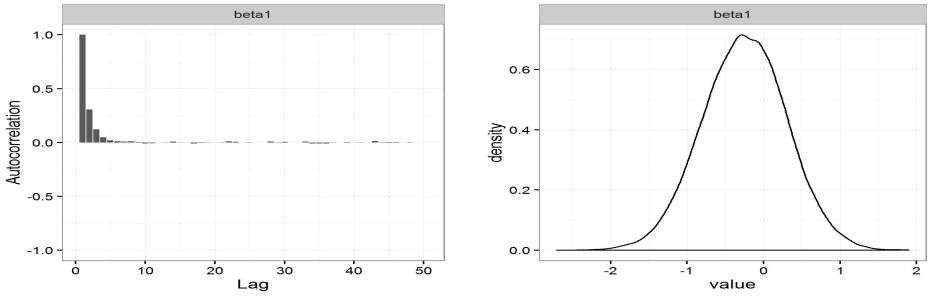

Figure A.1.5. Auto-correlation and density plot for 10% validation sizes

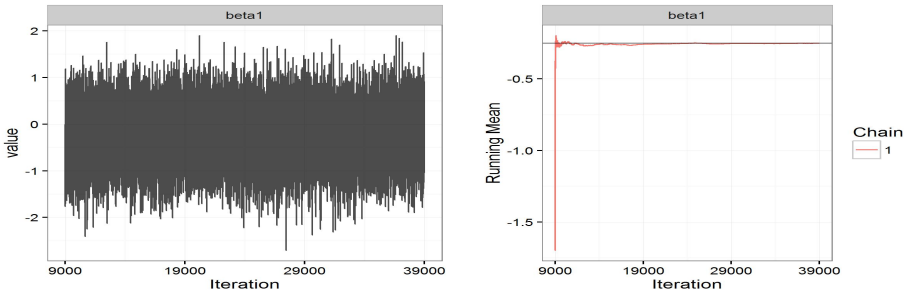

Figure A.1.6. Trace and running means plot for 10% validation sizes

(D) 5% Validation Size; Iterations: 60,000 Burn-in: 10,000 Thin: 5

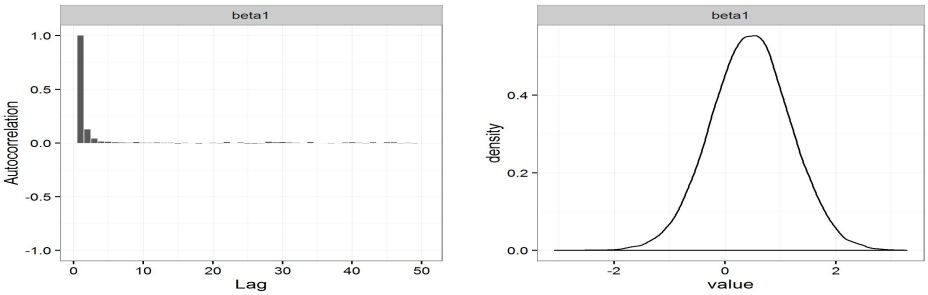

Figure A.1.7. Auto-correlation and density plot for 5% validation sizes

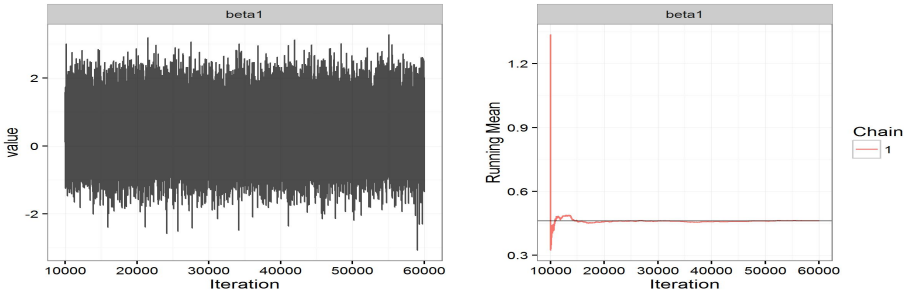

Figure A.1.8. Trace and running means plot for 5% validation sizes

(E) 2.5% Validation Size; Iterations: 225,000 Burn-in: 105,000 Thin: 12

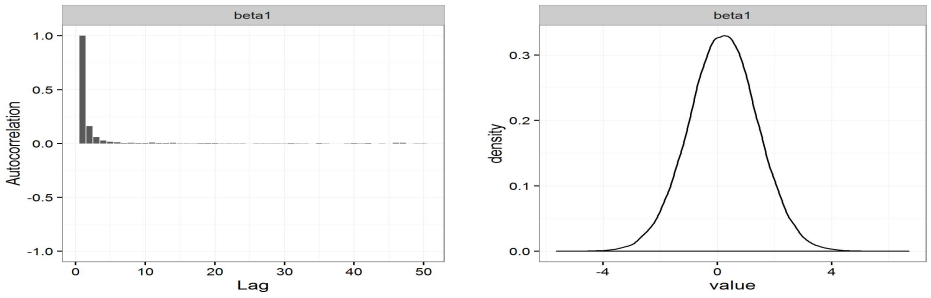

Figure A.1.9. Auto-correlation and density plot for 2.5% validation sizes

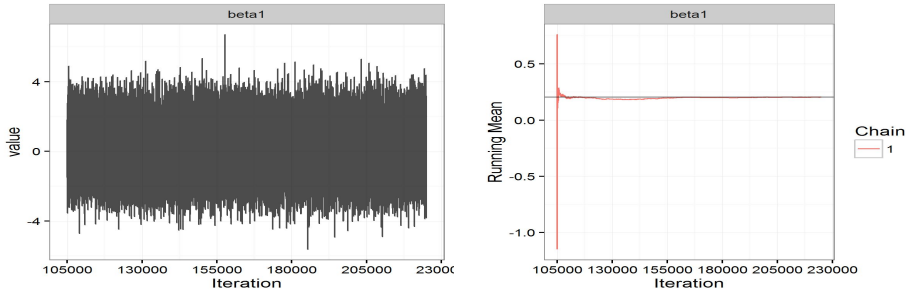

Figure A.1.10. Trace and running means plot for 2.5% validation sizes

A.2 Baseline Bayesian Analysis of Simulation Study Discussed in Section 3.1

(A) 25% Validation Size; Iterations: 25,000 Burn-in: 15,000 Thin: 1

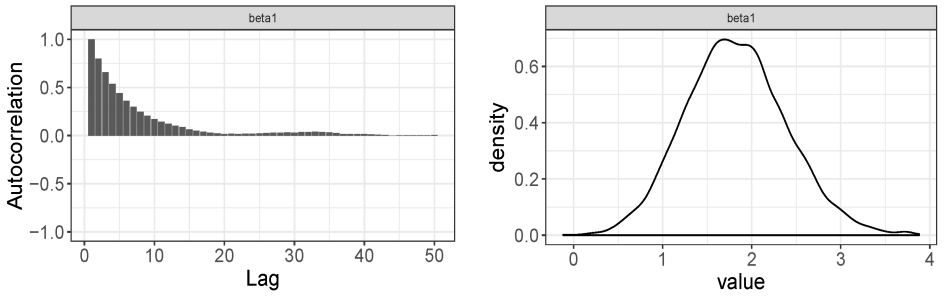

Figure A.2.1. Auto-correlation and density plot for 25% validation sizes

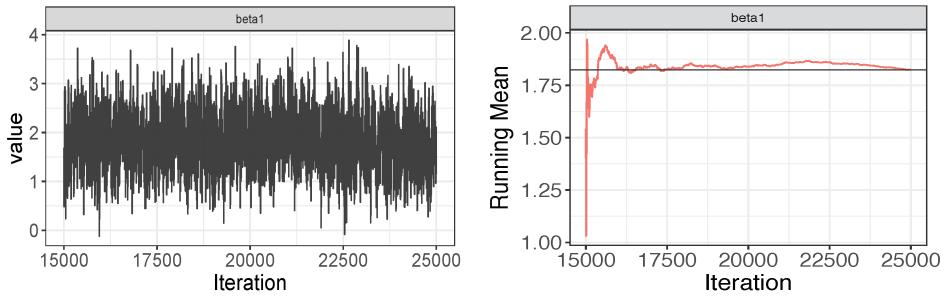

Figure A.2.2. Trace and running means plot for 25% validation sizes

(B) 15% Validation Size; Iterations: 40,000 Burn-in: 20,000 Thin: 2

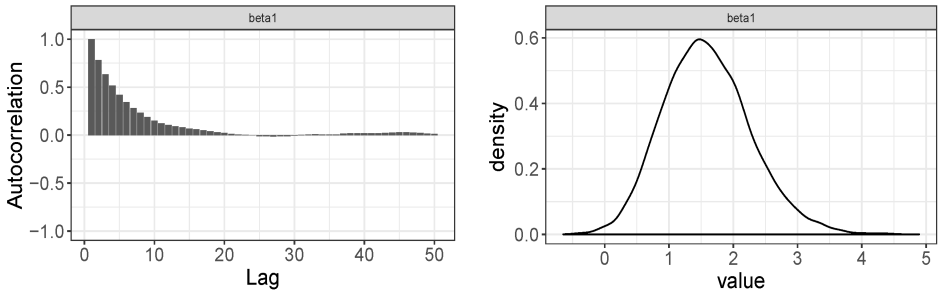

Figure A.2.3. Auto-correlation and density plot for 15% validation sizes

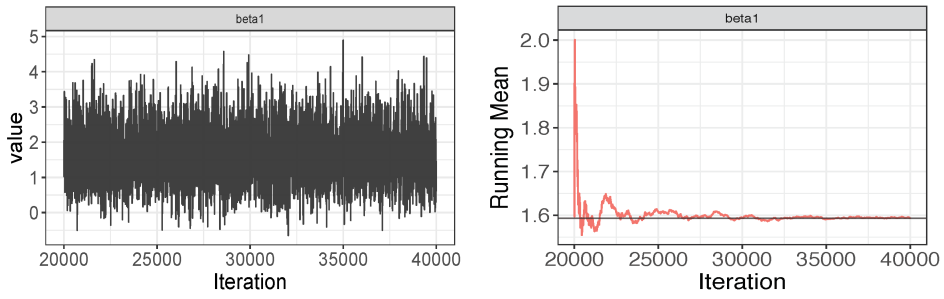

Figure A.2.4. Trace and running means plot for 15% validation sizes

(C) 10% Validation Size; Iterations: 45,000 Burn-in: 25,000 Thin: 2

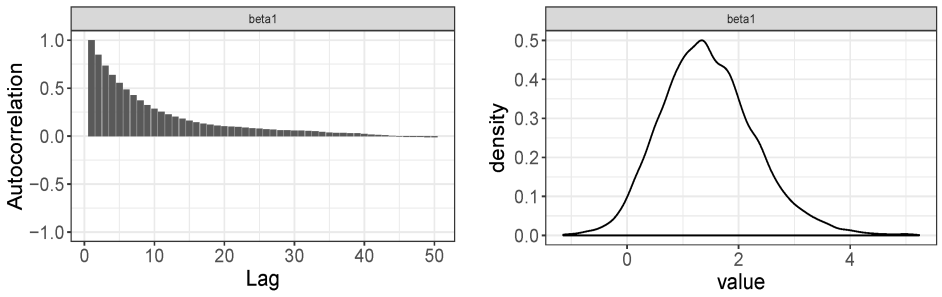

Figure A.2.5. Auto-correlation and density plot for 10% validation sizes

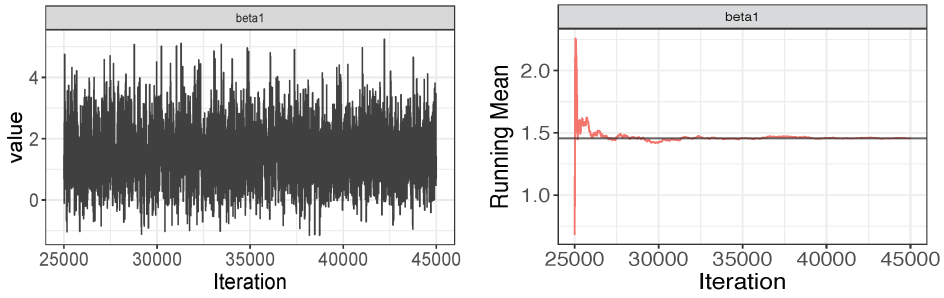

Figure A.2.6. Trace and running means plot for 10% validation sizes

(D) 5% Validation Size; Iterations: 70,000 Burn-in: 40,000 Thin: 3

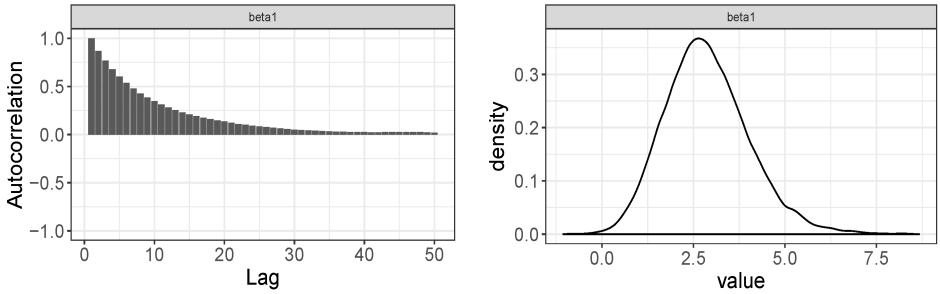

Figure A.2.7. Auto-correlation and density plot for 5% validation sizes

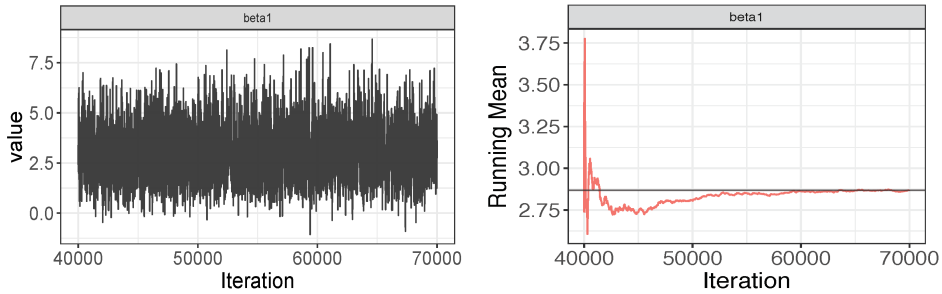

Figure A.2.8. Trace and running means plot for 5% validation sizes

(E) 2.5% Validation Size; Iterations: 110,000 Burn-in: 60,000 Thin: 5

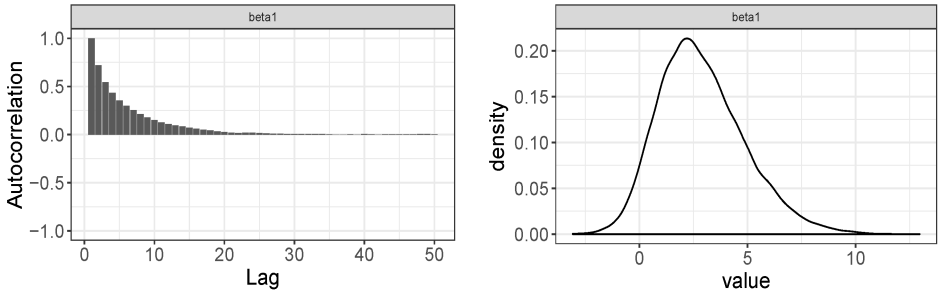

Figure A.2.9. Auto-correlation and density plot for 2.5% validation sizes

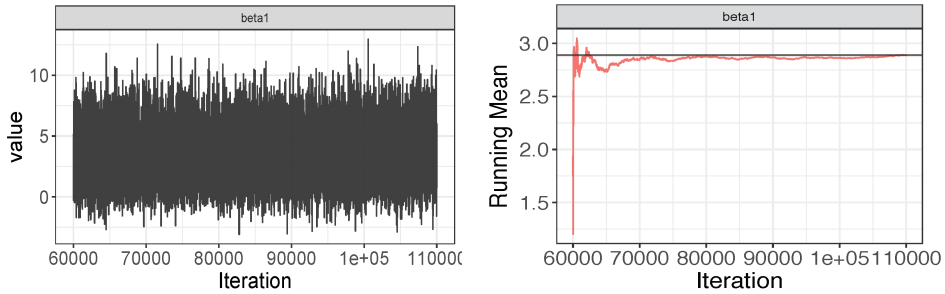

Figure A.2.10. Trace and running means plot for 2.5% validation sizes

A.3 Bayesian Analysis of Simulation Study Discussed in Section 3.2 Assuming Independent Differential Misclassification

(A) 25% Validation Size; Iterations: 25,000 Burn-in: 15,000 Thin: 1

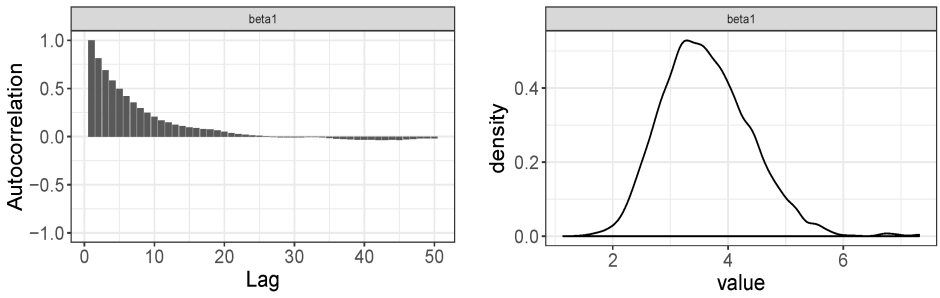

Figure A.3.1. Auto-correlation and density plot for 25% validation sizes

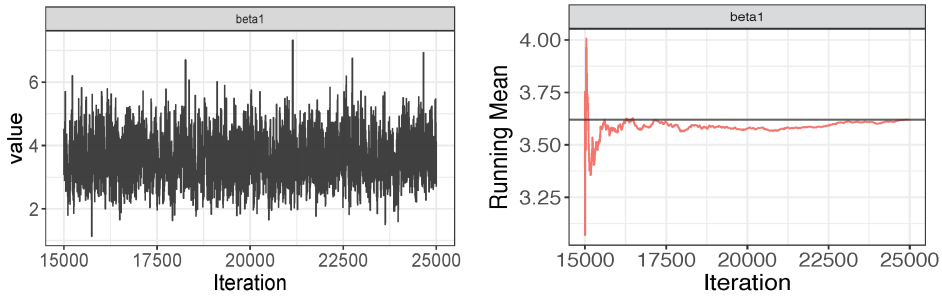

Figure A.3.2. Trace and running means plot for 25% validation sizes

(B) 15% Validation Size; Iterations: 40,000 Burn-in: 20,000 Thin: 2

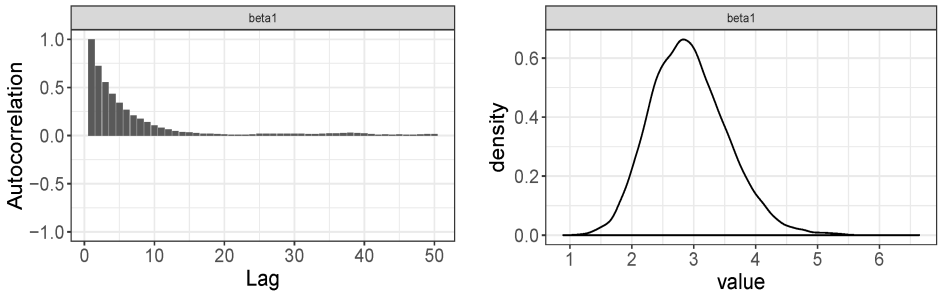

Figure A.3.3. Auto-correlation and density plot for 15% validation sizes

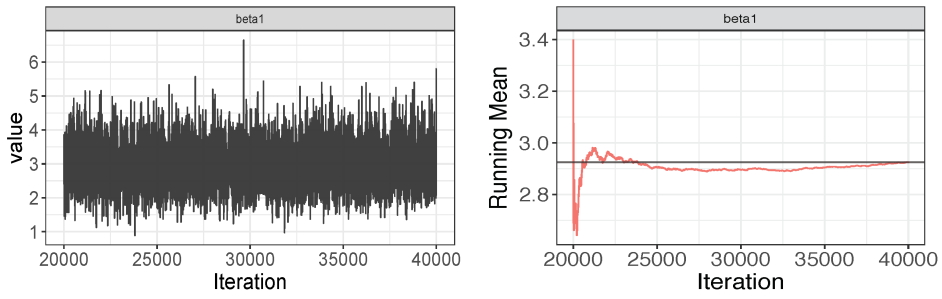

Figure A.3.4. Trace and running means plot for 15% validation sizes

(C) 10% Validation Size; Iterations: 45,000 Burn-in: 25,000 Thin: 2

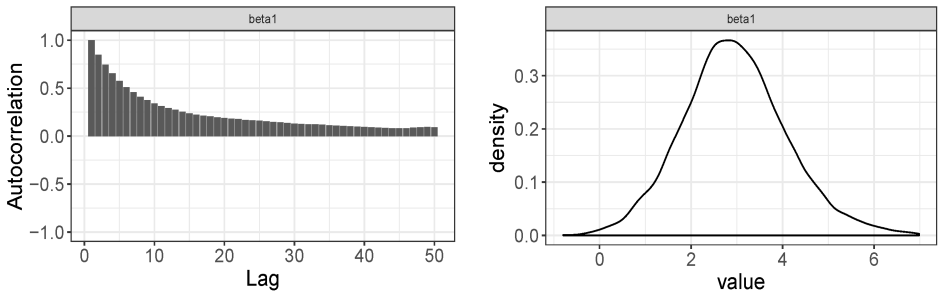

Figure A.3.5. Auto-correlation and density plot for 10% validation sizes

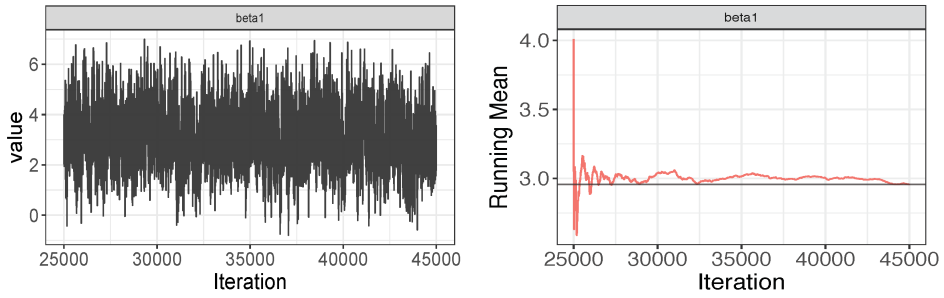

Figure A.3.6. Trace and running means plot for 10% validation sizes

(D) 5% Validation Size; Iterations: 70,000 Burn-in: 40,000 Thin: 3

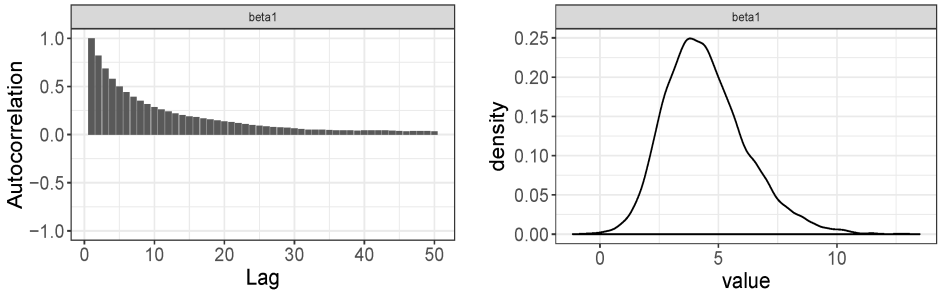

Figure A.3.7. Auto-correlation and density plot for 5% validation sizes

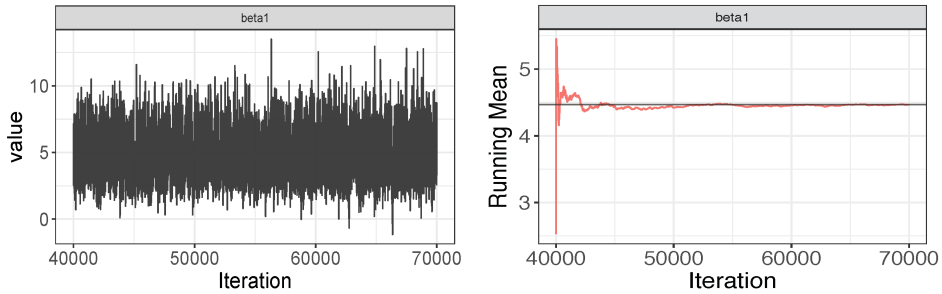

Figure A.3.8. Trace and running means plot for 5% validation sizes

(E) 2.5% Validation Size; Iterations: 110,000 Burn-in: 60,000 Thin: 5

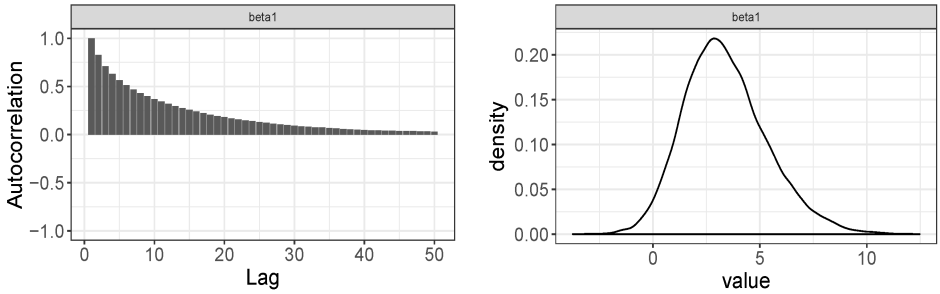

Figure A.3.9. Auto-correlation and density plot for 2.5% validation sizes

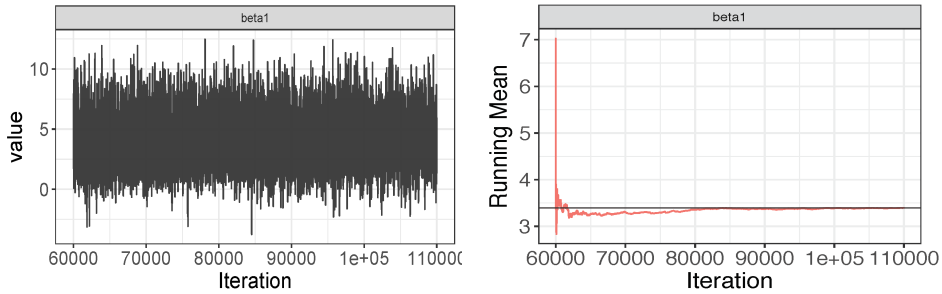

Figure A.3.10. Trace and running means plot for 2.5% validation sizes

A.4 Bayesian Analysis of Simulation Study Discussed in Section 3.3 with Overall Sample Sizes of $n = 10,000$

(A) 25% Validation Size; Iterations: 25,000 Burn-in: 15,000 Thin: 1

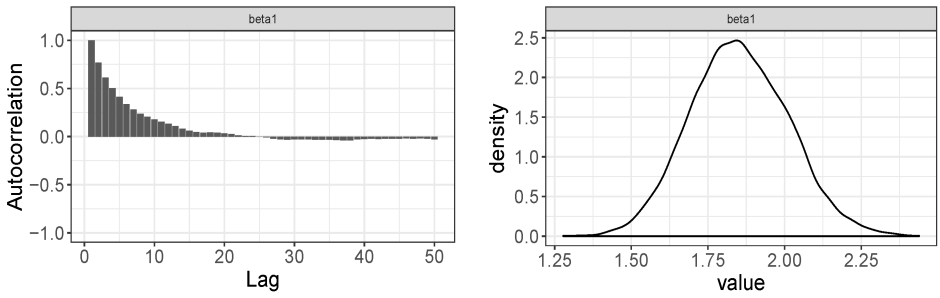

Figure A.4.1. Auto-correlation and density plot for 25% validation sizes

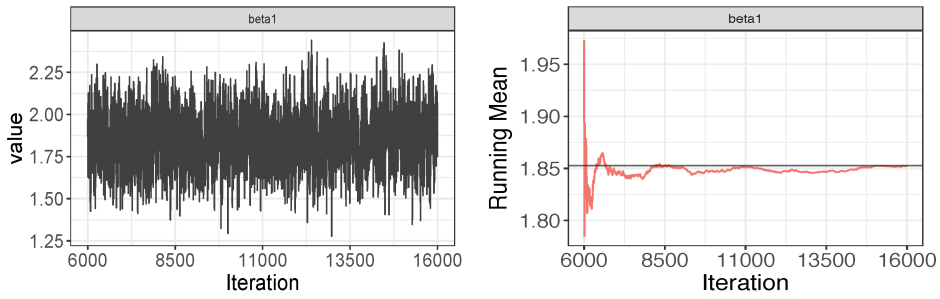

Figure A.4.2. Trace and running means plot for 25% validation sizes

(B) 15% Validation Size; Iterations: 40,000 Burn-in: 20,000 Thin: 2

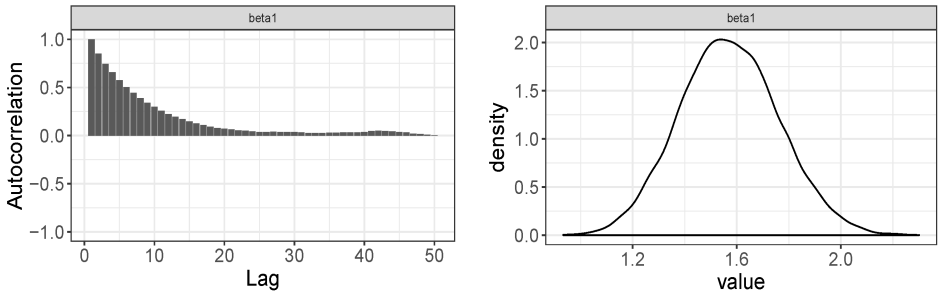

Figure A.4.3. Auto-correlation and density plot for 15% validation sizes

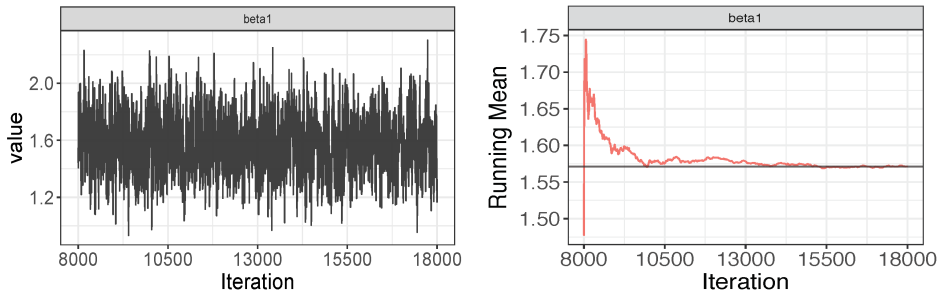

Figure A.4.4. Trace and running means plot for 15% validation sizes

(C) 10% Validation Size; Iterations: 45,000 Burn-in: 25,000 Thin: 2

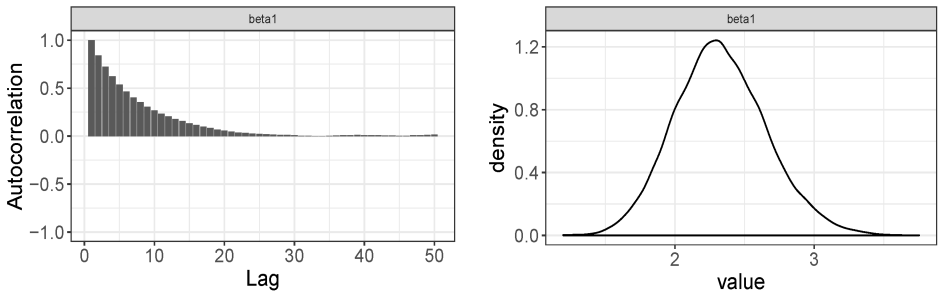

Figure A.4.5. Auto-correlation and density plot for 10% validation sizes

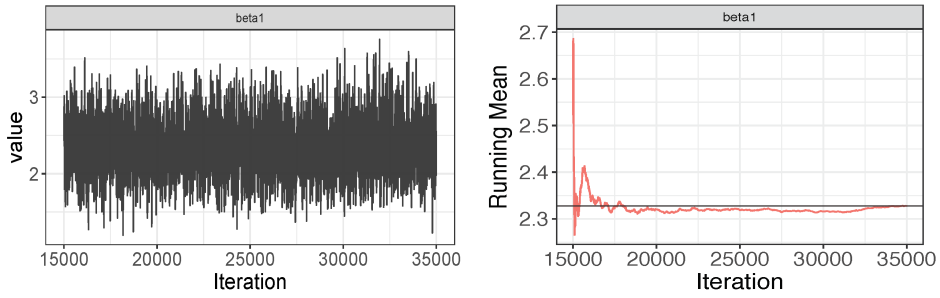

Figure A.4.6. Trace and running means plot for 10% validation sizes

(D) 5% Validation Size; Iterations: 70,000 Burn-in: 40,000 Thin: 3

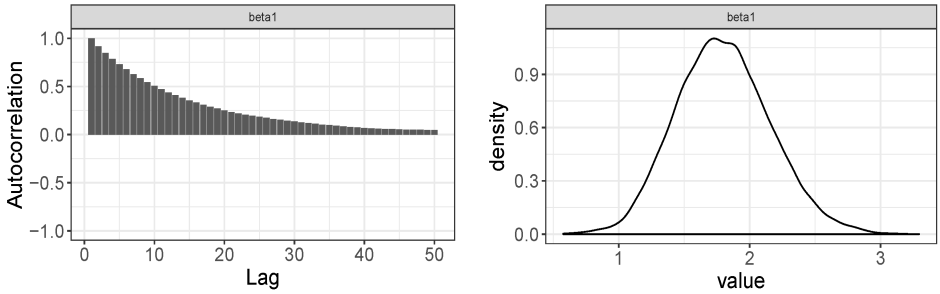

Figure A.4.7. Auto-correlation and density plot for 5% validation sizes

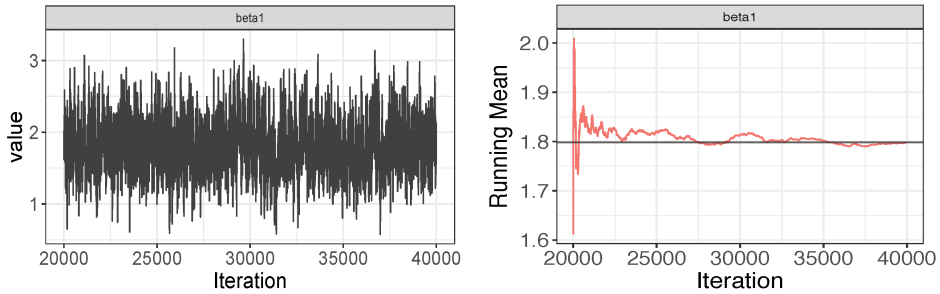

Figure A.4.8. Trace and running means plot for 5% validation sizes

(E) 2.5% Validation Size; Iterations: 110,000 Burn-in: 60,000 Thin: 5

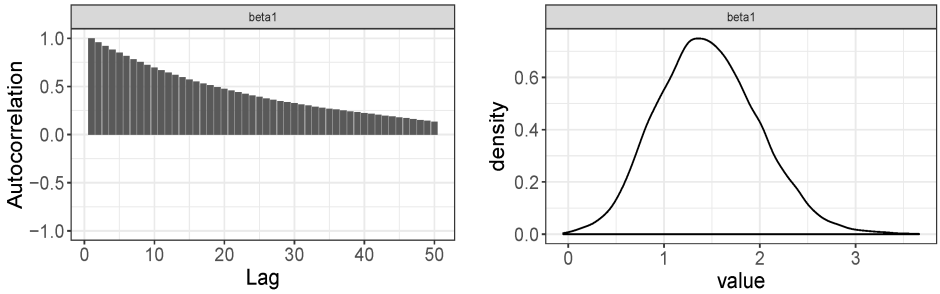

Figure A.4.9. Auto-correlation and density plot for 2.5% validation sizes

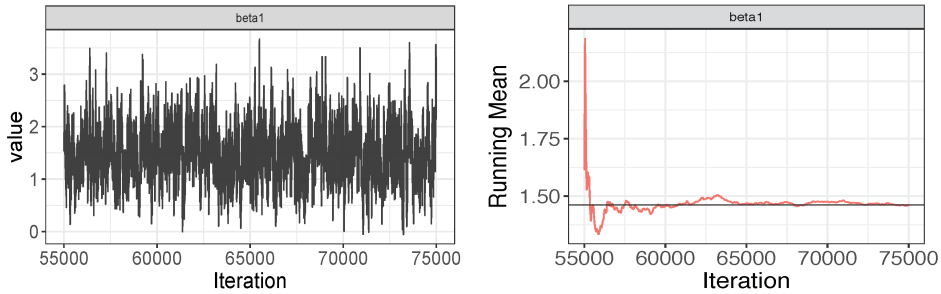

Figure A.4.10. Trace and running means plot for 2.5% validation sizes

A.5 Bayesian Analysis of Simulation Study Discussed in Section 3.4.3 Utilizing Informative Priors-Centered

(A) 25% Validation Size; Iterations: 15,000 Burn-in: 5,000 Thin: 1

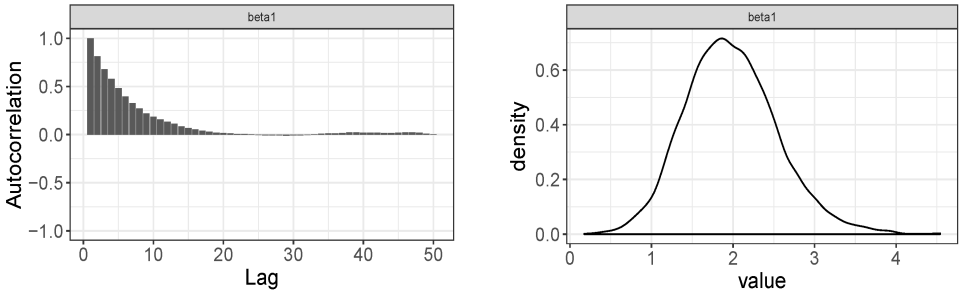

Figure A.5.1. Auto-correlation and density plot for 25% validation sizes

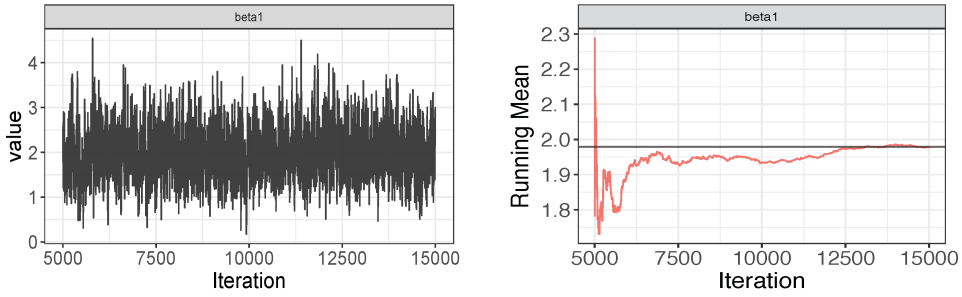

Figure A.5.2. Trace and running means plot for 25% validation sizes

(B) 15% Validation Size; Iterations: 35,000 Burn-in: 25,000 Thin: 1

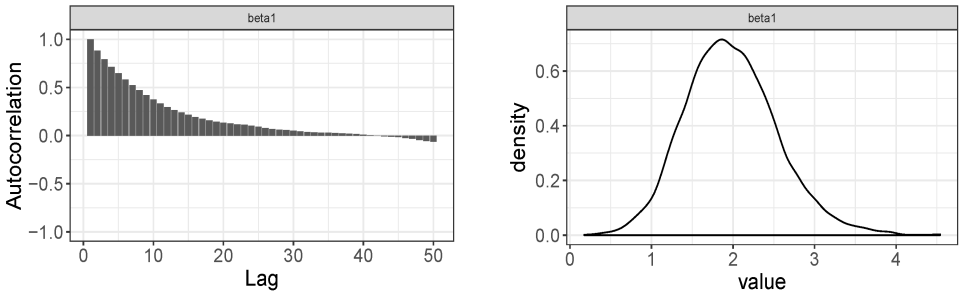

Figure A.5.3. Auto-correlation and density plot for 15% validation sizes

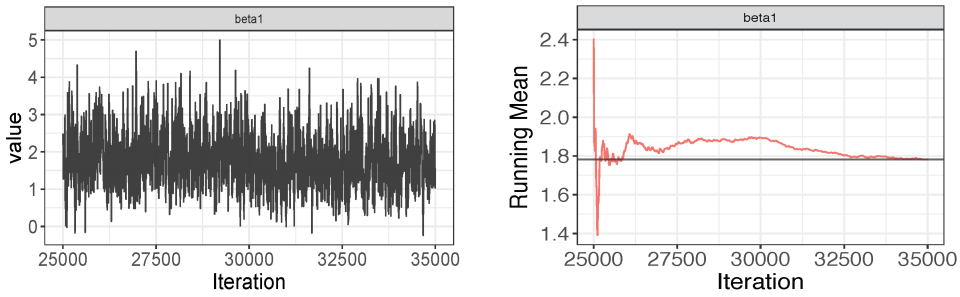

Figure A.5.4. Trace and running means plot for 15% validation sizes

(C) 10% Validation Size; Iterations: 50,000 Burn-in: 40,000 Thin: 1

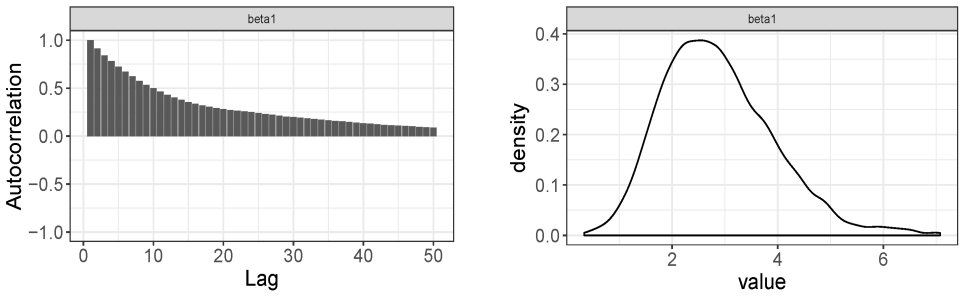

Figure A.5.5. Auto-correlation and density plot for 10% validation sizes

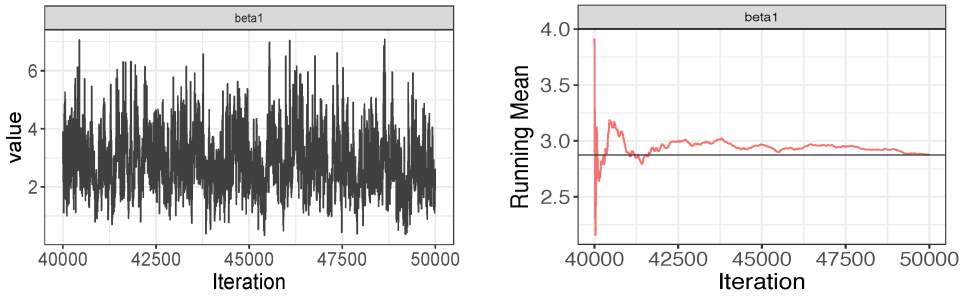

Figure A.5.6. Trace and running means plot for 10% validation sizes

(D) 5% Validation Size; Iterations: 80,000 Burn-in: 60,000 Thin: 2

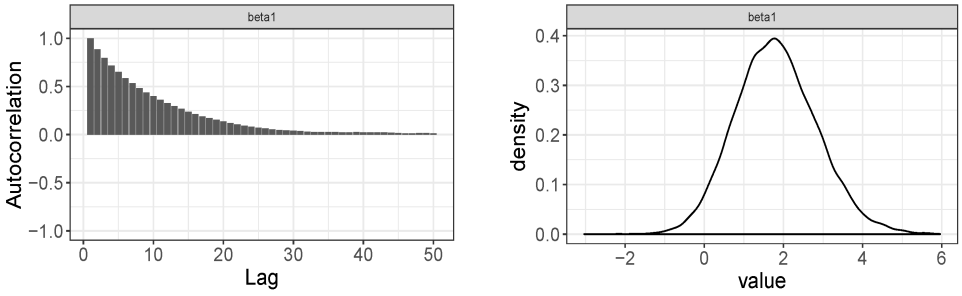

Figure A.5.7. Auto-correlation and density plot for 5% validation sizes

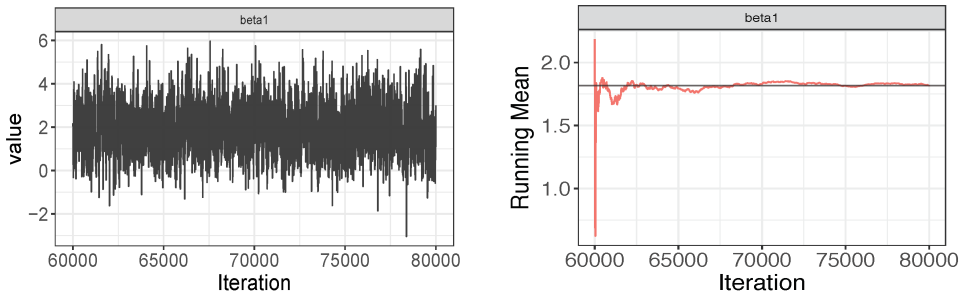

Figure A.5.8. Trace and running means plot for 5% validation sizes

(E) 2.5% Validation Size; Iterations: 105,000 Burn-in: 85,000 Thin: 2

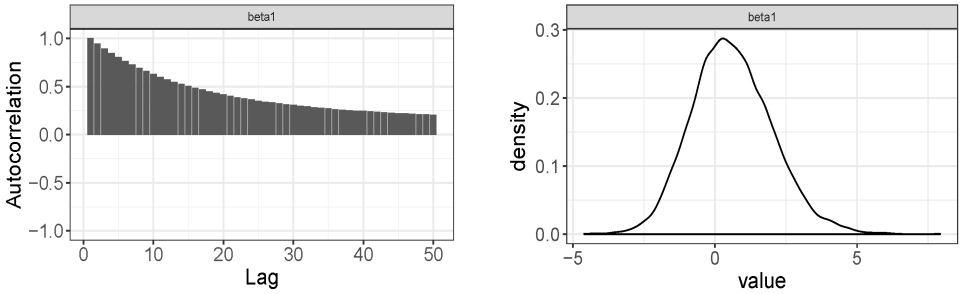

Figure A.5.9. Auto-correlation and density plot for 2.5% validation sizes

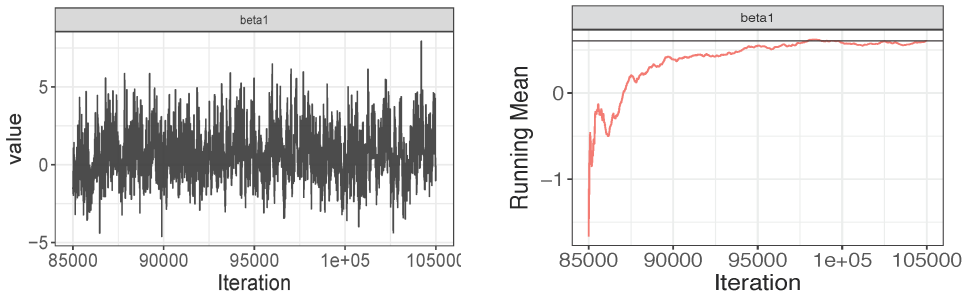

Figure A.5.10. Trace and running means plot for 2.5% validation sizes

A.6 Bayesian Analysis of Simulation Study Discussed in Section 3.4.3 Utilizing Informative Priors-Centered and Slightly Narrowed

(A) 25% Validation Size; Iterations: 15,000 Burn-in: 5,000 Thin: 1

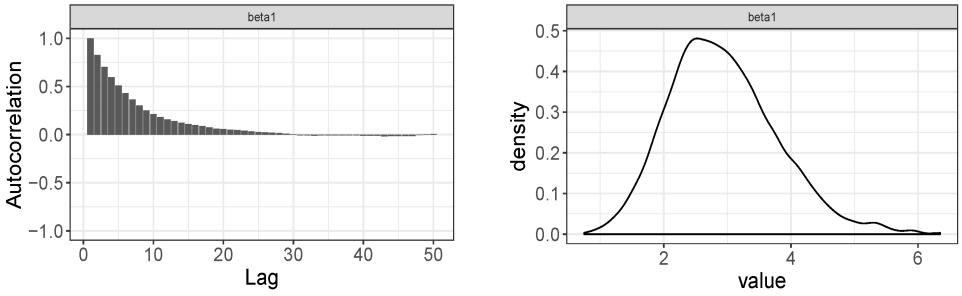

Figure A.6.1. Auto-correlation and density plot for 25% validation sizes

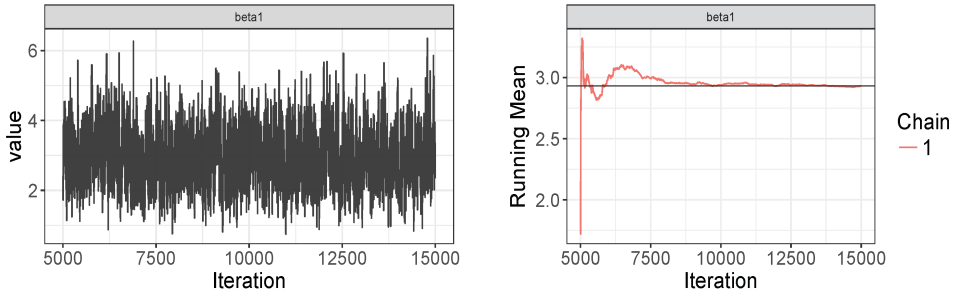

Figure A.6.2. Trace and running means plot for 25% validation sizes

(B) 15% Validation Size; Iterations: 35,000 Burn-in: 25,000 Thin: 1

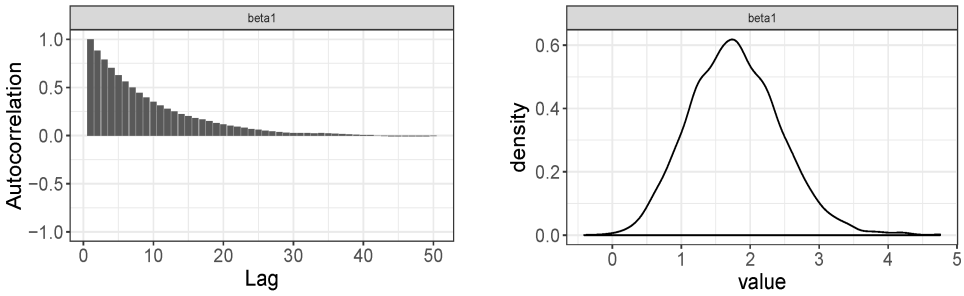

Figure A.6.3. Auto-correlation and density plot for 15% validation sizes

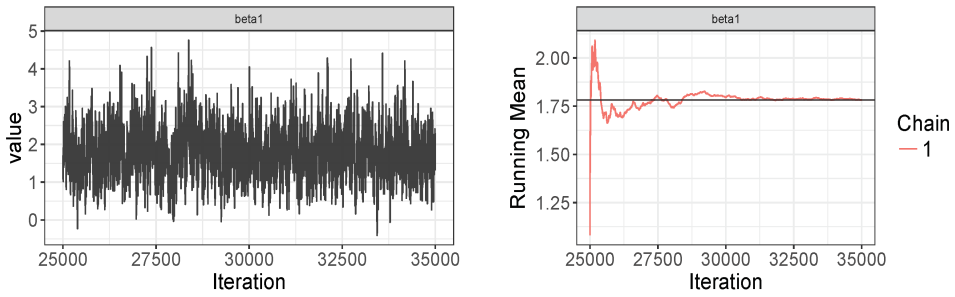

Figure A.6.4. Trace and running means plot for 15% validation sizes

(C) 10% Validation Size; Iterations: 50,000 Burn-in: 40,000 Thin: 1

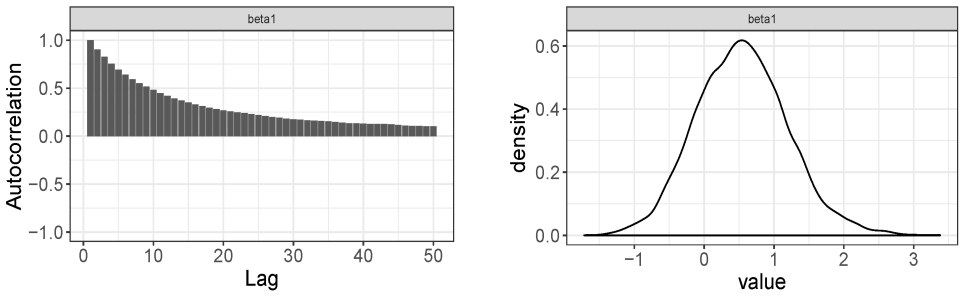

Figure A.6.5. Auto-correlation and density plot for 10% validation sizes

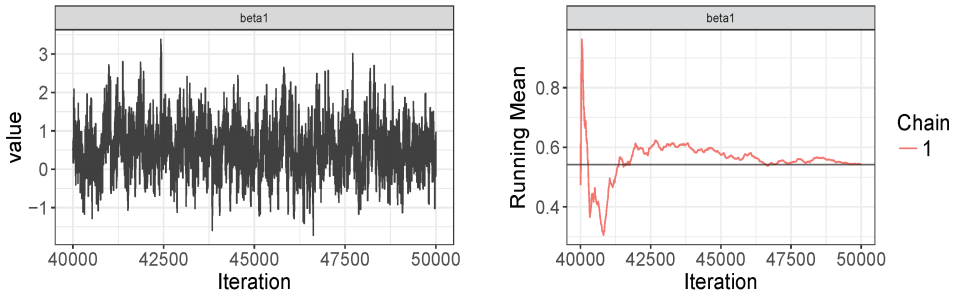

Figure A.6.6. Trace and running means plot for 10% validation sizes

(D) 5% Validation Size; Iterations: 80,000 Burn-in: 60,000 Thin: 2

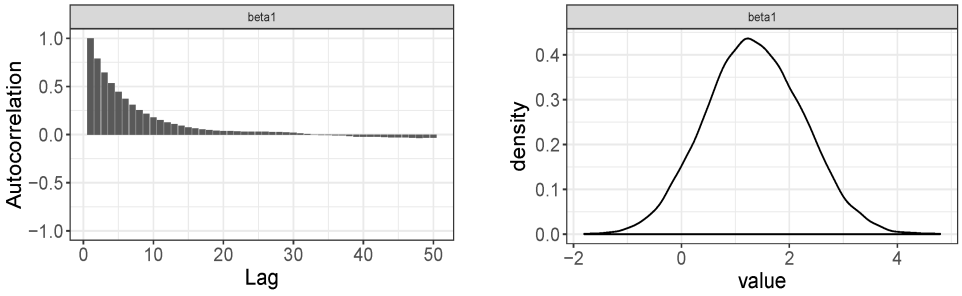

Figure A.6.7. Auto-correlation and density plot for 5% validation sizes

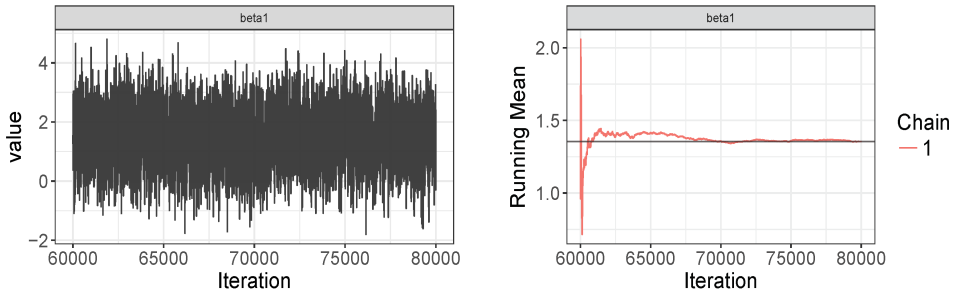

Figure A.6.8. Trace and running means plot for 5% validation sizes

(E) 2.5% Validation Size; Iterations: 105,000 Burn-in: 85,000 Thin: 2

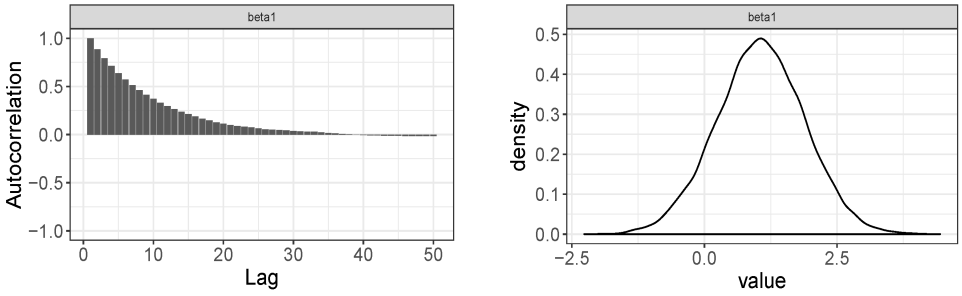

Figure A.6.9. Auto-correlation and density plot for 2.5% validation sizes

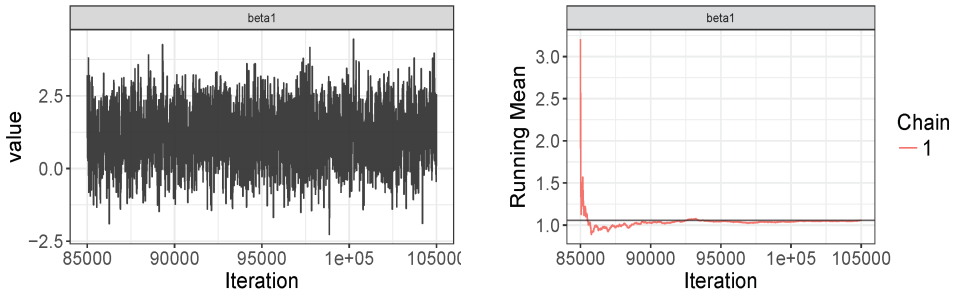

Figure A.6.10. Trace and running means plot for 2.5% validation sizes

A.7 Bayesian Analysis of Simulation Study Discussed in Section 3.4.3 Utilizing Informative Priors-Centered and Narrowed

(A) 25% Validation Size; Iterations: 15,000 Burn-in: 5,000 Thin: 1

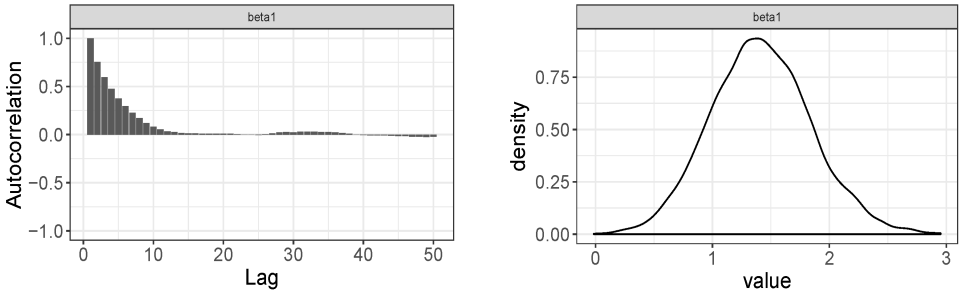

Figure A.7.1. Auto-correlation and density plot for 25% validation sizes

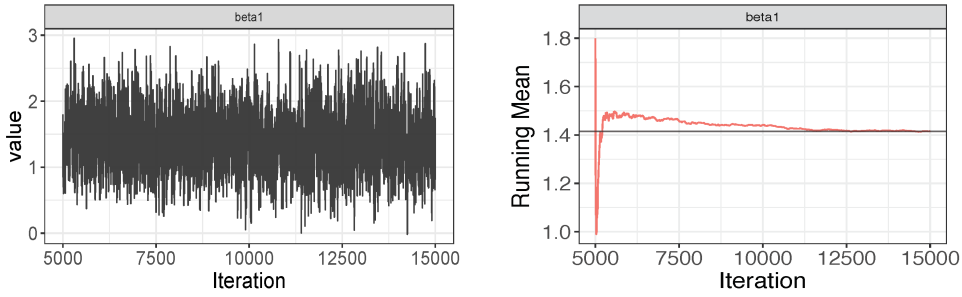

Figure A.7.2. Trace and running means plot for 25% validation sizes

(B) 15% Validation Size; Iterations: 35,000 Burn-in: 25,000 Thin: 1

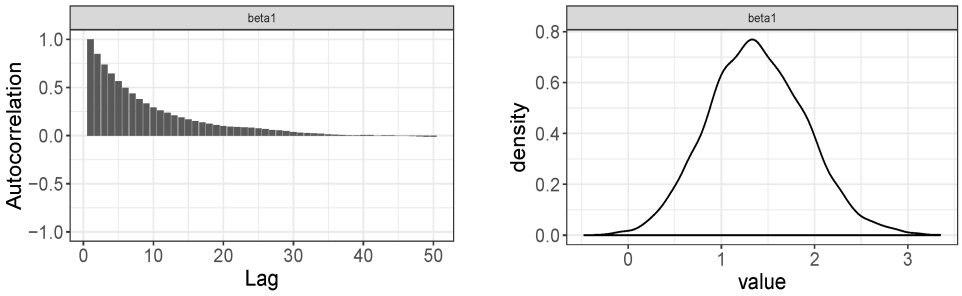

Figure A.7.3. Auto-correlation and density plot for 15% validation sizes

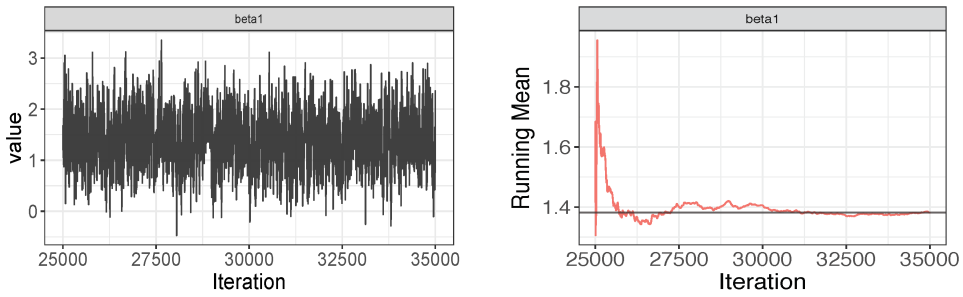

Figure A.7.4. Trace and running means plot for 15% validation sizes

(C) 10% Validation Size; Iterations: 50,000 Burn-in: 40,000 Thin: 1

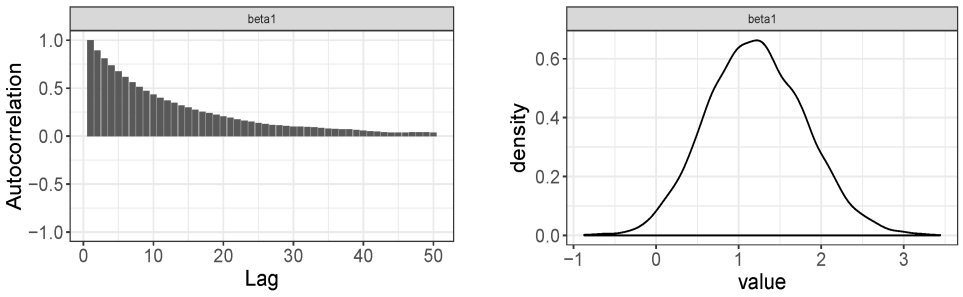

Figure A.7.5. Auto-correlation and density plot for 10% validation sizes

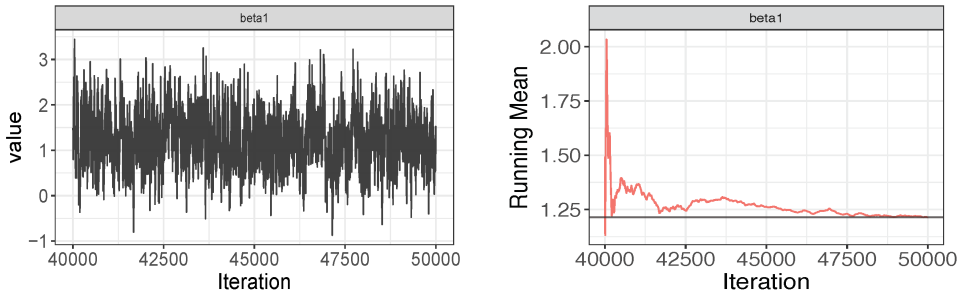

Figure A.7.6. Trace and running means plot for 10% validation sizes

(D) 5% Validation Size; Iterations: 80,000 Burn-in: 60,000 Thin: 2

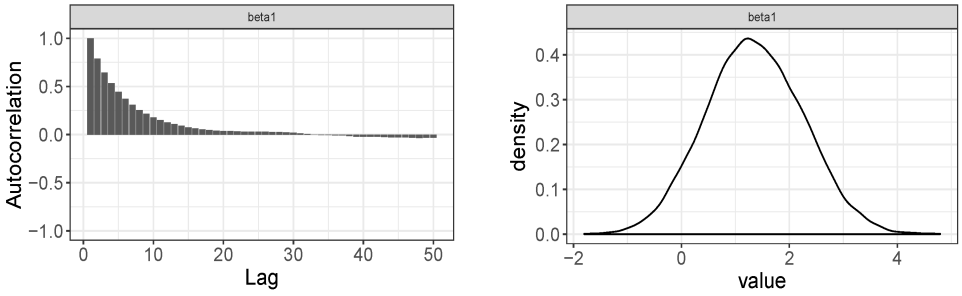

Figure A.7.7. Auto-correlation and density plot for 5% validation sizes

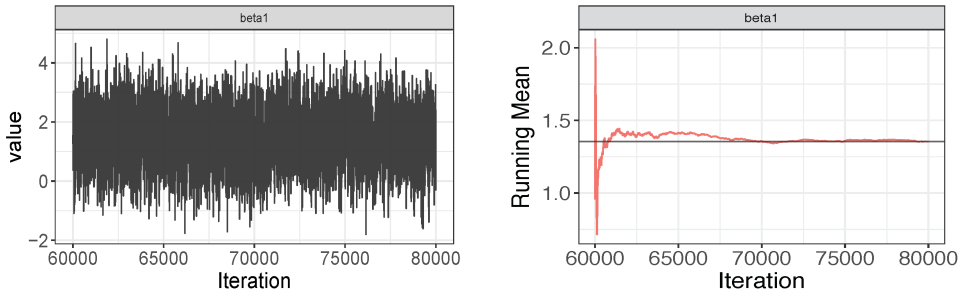

Figure A.7.8. Trace and running means plot for 5% validation sizes

(E) 2.5% Validation Size; Iterations: 105,000 Burn-in: 85,000 Thin: 2

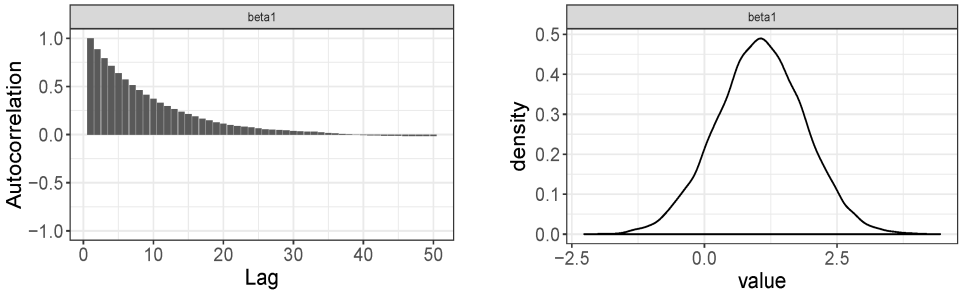

Figure A.7.9. Auto-correlation and density plot for 2.5% validation sizes

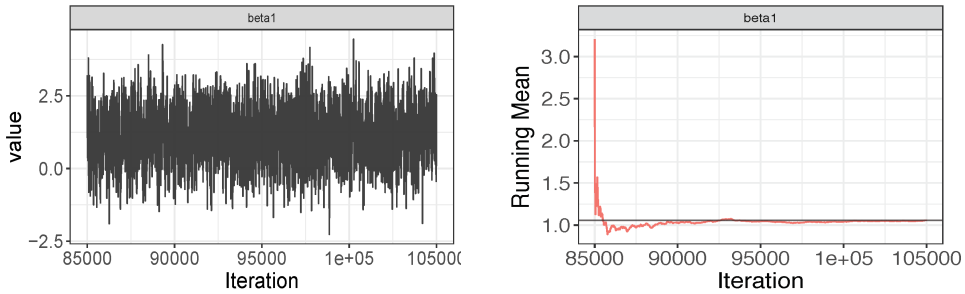

Figure A.7.10. Trace and running means plot for 2.5% validation sizes

A.8 Bayesian Analysis of Simulation Study Discussed in Section 4.3.5 for Correlated Binary Responses for β_1

(A) Naïve Model: Iterations: 200,000; Burn-in: 185,000; Thin: 3; Chains: 2

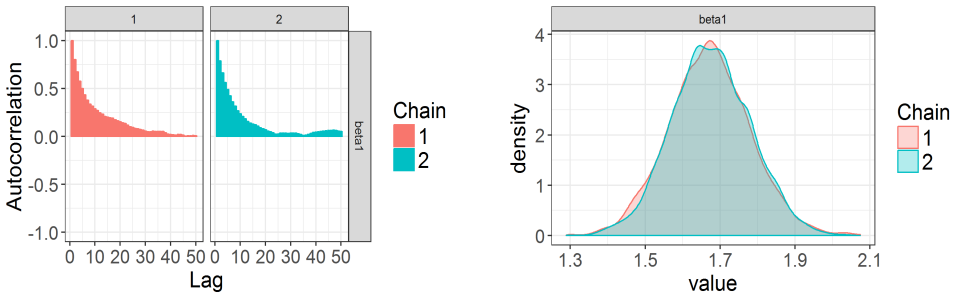

Figure A.8.1. Auto-correlation and density plot for β_1 under the Naïve Model

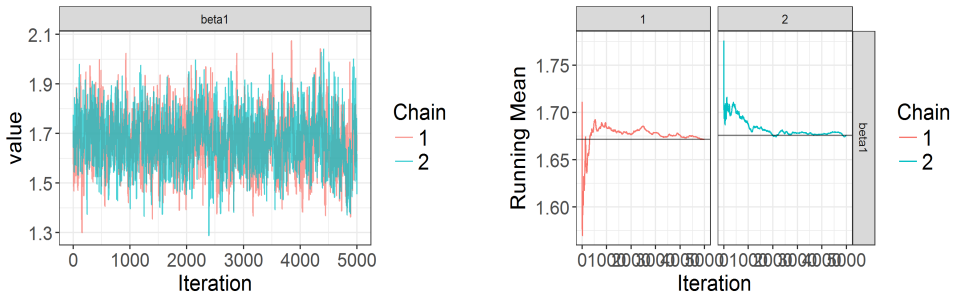

Figure A.8.2. Trace and running means plot for β_1 under the Naïve Model

(B) Ideal Model: Iterations: 200,000; Burn-in: 185,000; Thin: 3; Chains: 2

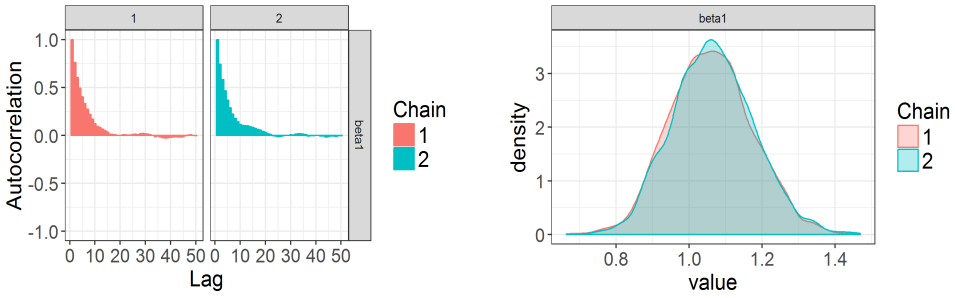

Figure A.8.3. Auto-correlation and density plot for β_1 under the Ideal Model

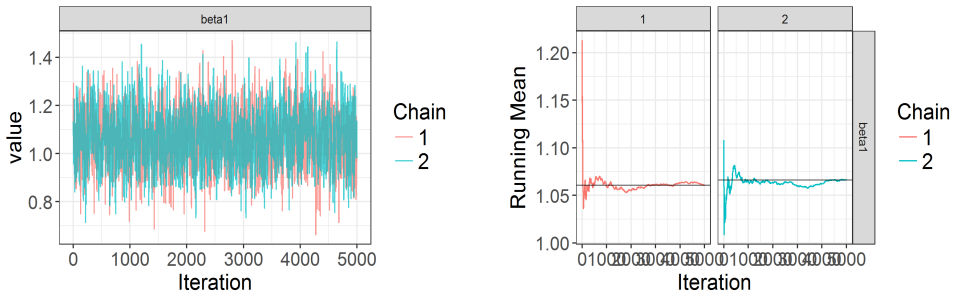

Figure A.8.4. Trace and running means plot for β_1 under the Ideal Model

(C) General Model: Iterations: 900,000; Burn-in: 750,000; Thin: 30; Chains: 2

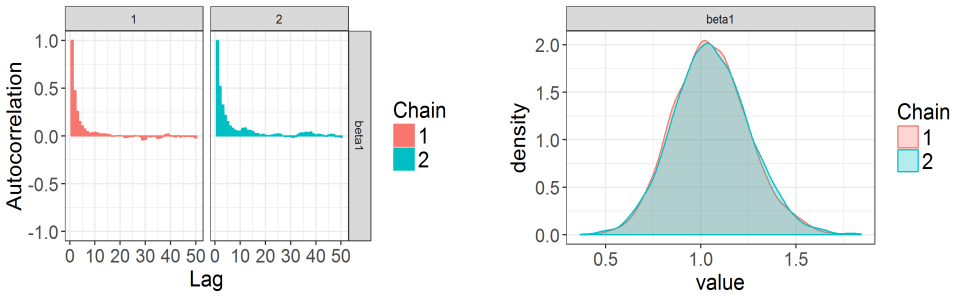

Figure A.8.5. Auto-correlation and density plot for β_1 under the General Model

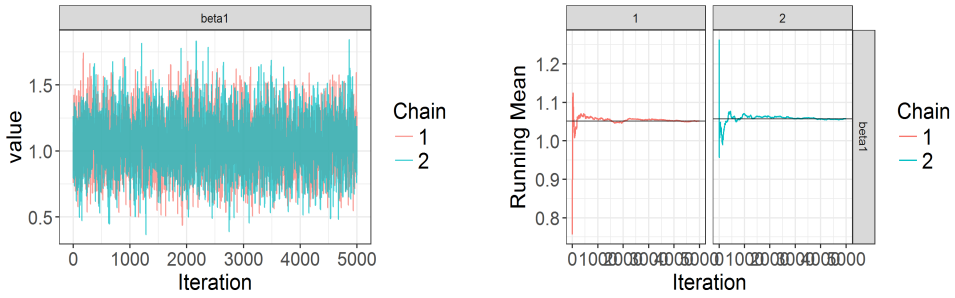

Figure A.8.6. Trace and running means plot for β_1 under the General Model

(D) Independent Correlated Model: Iterations: 400,000; Burn-in: 350,000; Thin: 10; Chains: 2

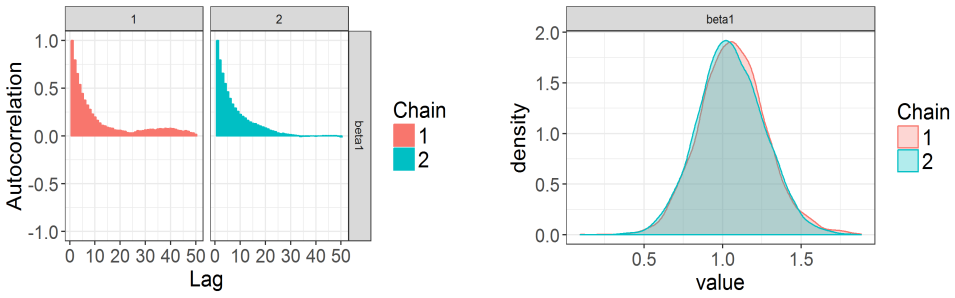

Figure A.8.7. Auto-correlation and density plot for β_1 under the Independent Correlated Model

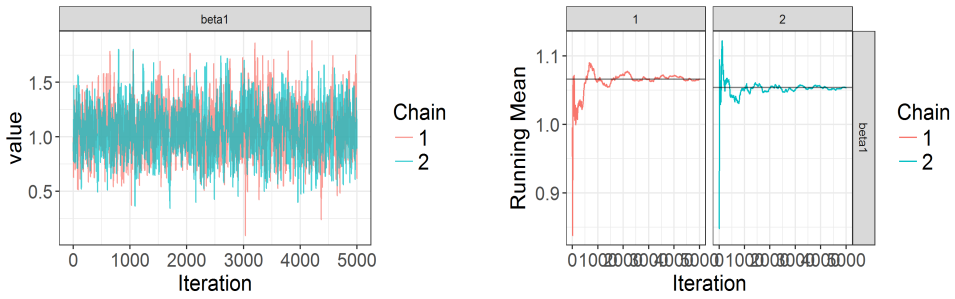

Figure A.8.8. Trace and running means plot for β_1 under the Independent Correlated Model

(E) Un-Correlated Differential Model: Iterations: 375,000; Burn-in: 350,000; Thin: 5; Chains: 2

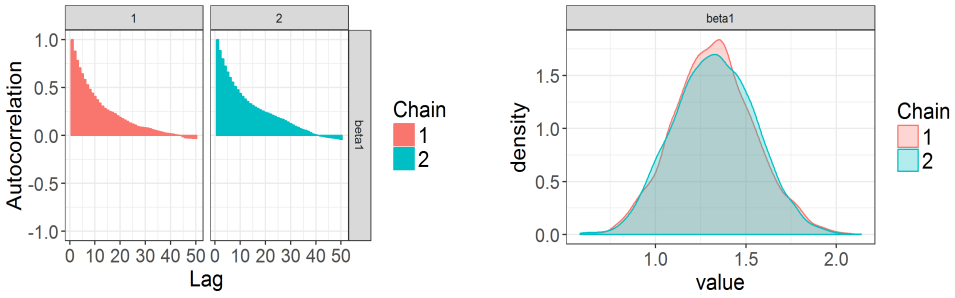

Figure A.8.9. Auto-correlation and density plot for β_1 under the Un-Correlated Differential Model

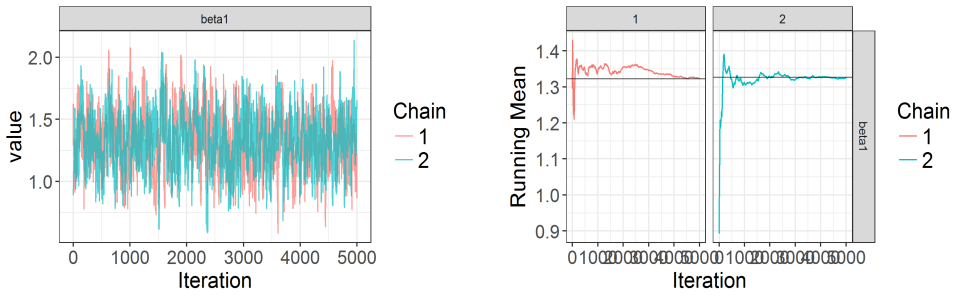

Figure A.8.10. Trace and running means plot for β_1 under the Un-Correlated Differential Model

(F) Un-Correlated Non-Differential Model: Iterations: 350,000; Burn-in: 330,000; Thin: 4; Chains: 2

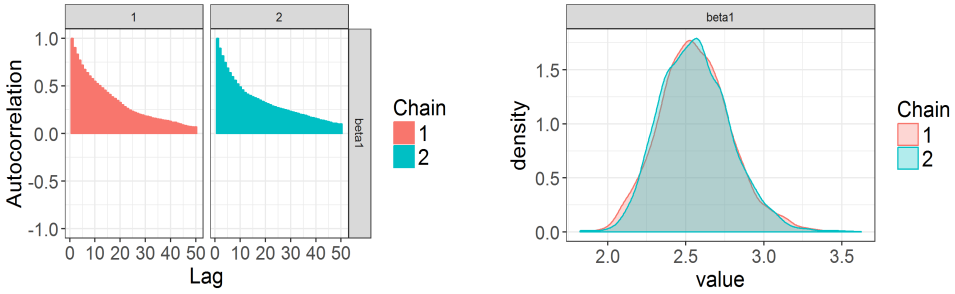

Figure A.8.11. Auto-correlation and density plot for β_1 under the Un-Correlated Non-Differential Model

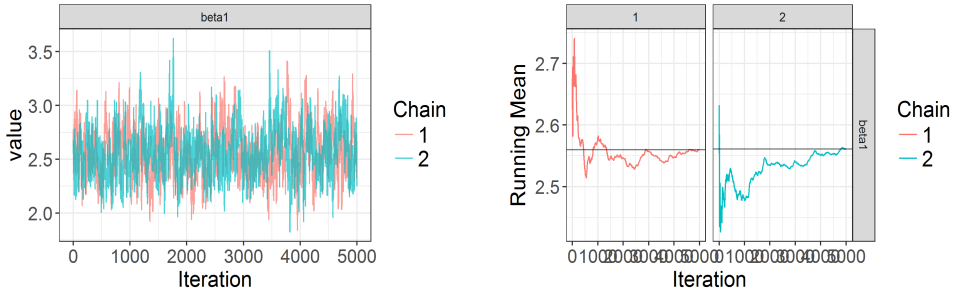

Figure A.8.12. Trace and running means plot for β_1 under the Un-Correlated Non-Differential Model

A.9 Bayesian Analysis of Simulation Study Discussed in Section 4.3.5 for Correlated Binary Responses for $\sigma^2_{u_i} = 1/\tau_i$

(A) Naïve Model: Iterations: 200,000; Burn-in: 185,000; Thin: 3; Chains: 2

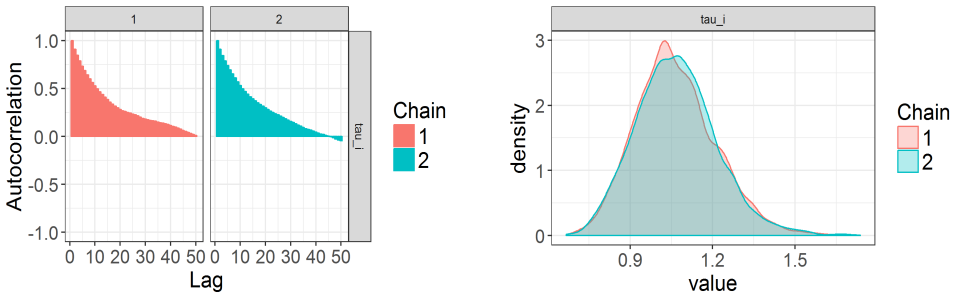

Figure A.9.1. Auto-correlation and density plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Naïve Model

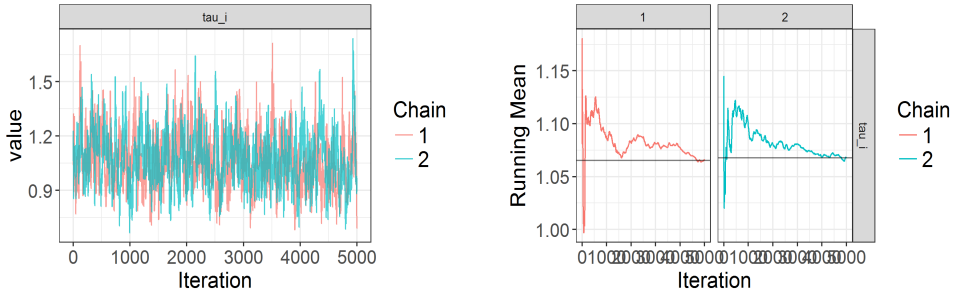

Figure A.9.2. Trace and running means plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Naïve Model

(B) Ideal Model: Iterations: 200,000; Burn-in: 185,000; Thin: 3; Chains: 2

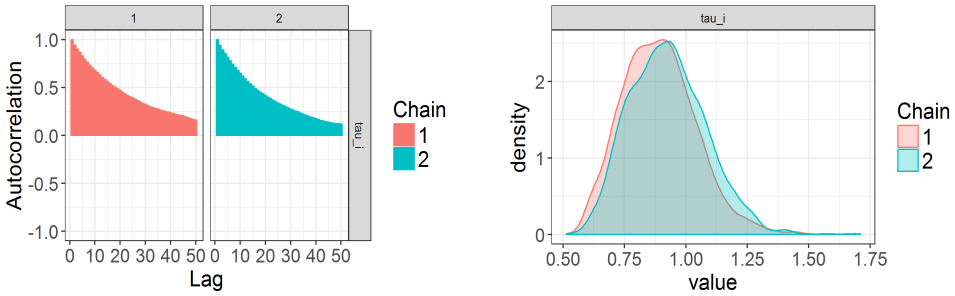

Figure A.9.3. Auto-correlation and density plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Ideal Model

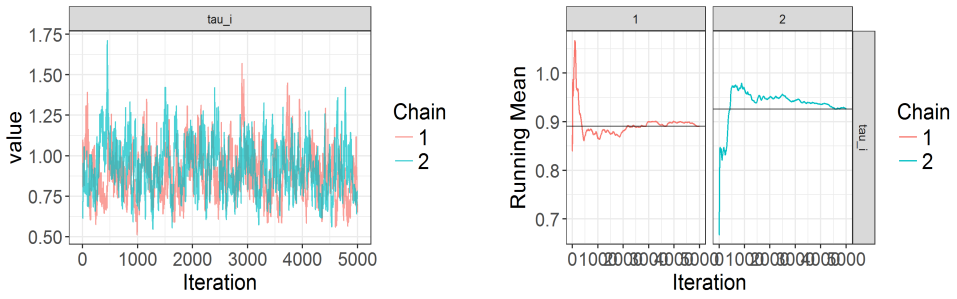

Figure A.9.4. Trace and running means plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Ideal Model

(C) General Model: Iterations: 900,000; Burn-in: 750,000; Thin: 30; Chains: 2

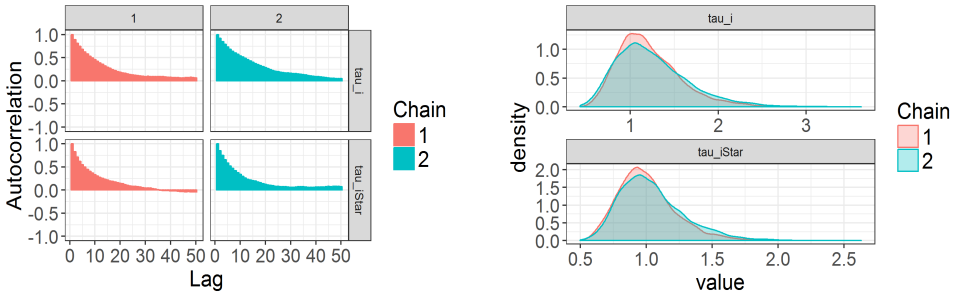

Figure A.9.5. Auto-correlation and density plot for $\sigma^2_{u_i} = 1/\tau_i$ under the General Model

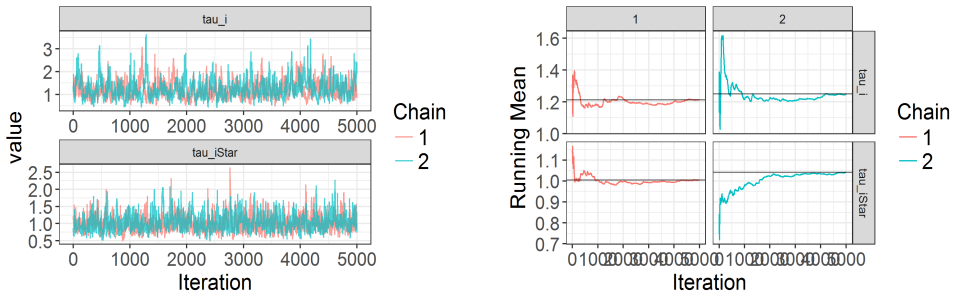

Figure A.9.6. Trace and running means plot for $\sigma^2_{u_i} = 1/\tau_i$ under the General Model

(D) Independent Correlated Model: Iterations: 400,000; Burn-in: 350,000; Thin: 10; Chains: 2

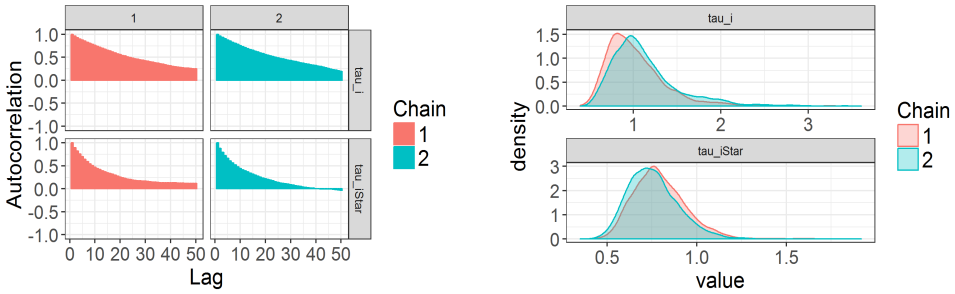

Figure A.9.7. Auto-correlation and density plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Independent Correlated Model

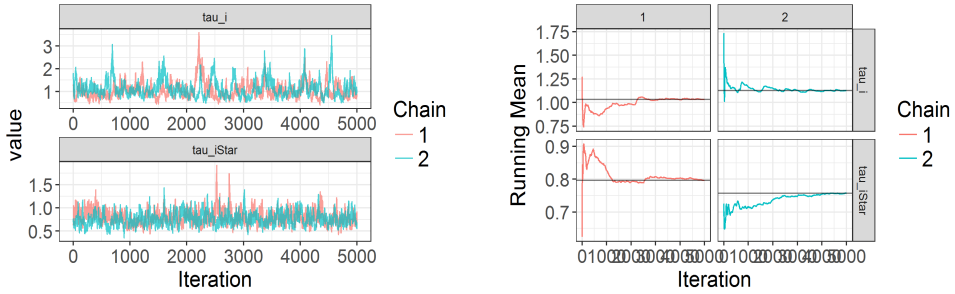

Figure A.9.8. Trace and running means plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Independent Correlated Model

(E) Un-Correlated Differential Model: Iterations: 375,000; Burn-in: 350,000; Thin: 5; Chains: 2

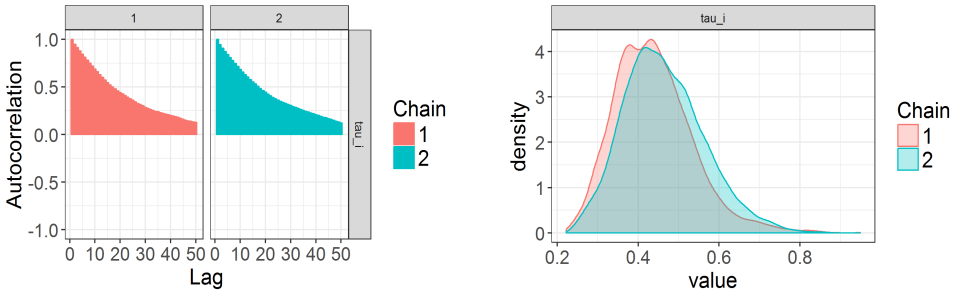

Figure A.9.9. Auto-correlation and density plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Un-Correlated Differential Model

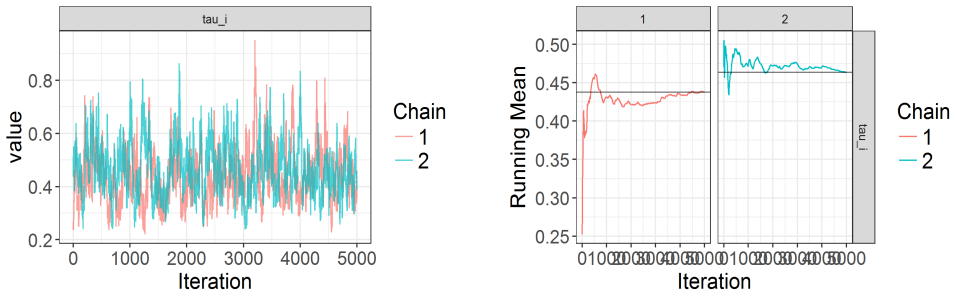

Figure A.9.10. Trace and running means plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Un-Correlated Differential Model

(F) Un-Correlated Non-Differential Model: Iterations: 350,000; Burn-in: 330,000; Thin: 4; Chains: 2

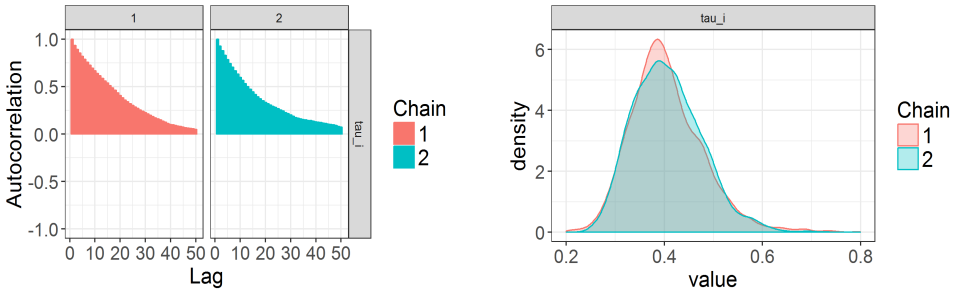

Figure A.9.11. Auto-correlation and density plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Un-Correlated Non-Differential Model

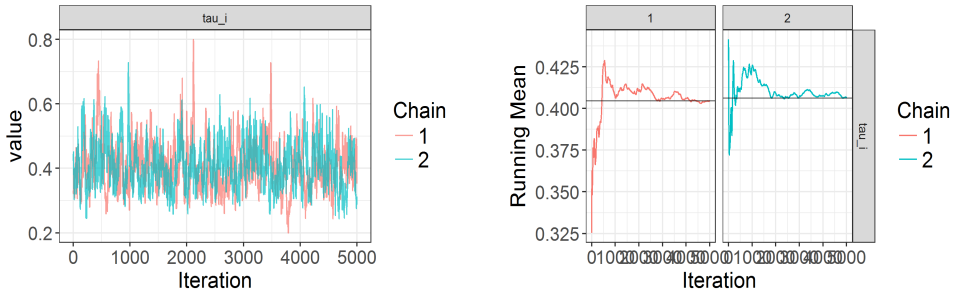

Figure A.9.12. Trace and running means plot for $\sigma^2_{u_i} = 1/\tau_i$ under the Un-Correlated Non-Differential Model

APPENDIX B

Models: JAGS and SAS

B.1 Analysis of Simulation Study based on work from Tang, et al[38] Discussed in Chapter Two

B.1.1 Frequentist (SAS) Model Code

```
Proc NLMIXED data=data;
parms theta0=-2 theta1=0.5 theta2=1 theta3=2
        theta4=0 theta5=0 theta6=0.8 theta7=-0.6
      delta0=-5 delta1=4 delta2=0.5 delta3=0
        delta4=0 delta5=1 delta6=0
      beta0=0.1 beta1=0.9 beta2=-0.04 beta3=0.8
        beta4=0.3 beta5=0.2
      gamma0=-2 gamma1=0 gamma2=2.5 gamma3=0 gamma4=0;

eta1_y1_x1=theta0 + theta1 + theta2*xstar + theta3 +
  theta4*age + theta5*black + theta6*riskchrt + theta7*hivpos;
p1_y1_x1=exp(eta1_y1_x1)/(1+exp(eta1_y1_x1));
eta1_y0_x1=theta0 + theta1 + theta2*xstar + theta4*age +
  theta5*black + theta6*riskchrt + theta7*hivpos;
p1_y0_x1=exp(eta1_y0_x1)/(1+exp(eta1_y0_x1));
eta1_y1_x0=theta0 + theta2*xstar + theta3 + theta4*age +
  theta5*black + theta6*riskchrt + theta7*hivpos;
p1_y1_x0=exp(eta1_y1_x0)/(1+exp(eta1_y1_x0));
eta1_y0_x0=theta0 + theta2*xstar + theta4*age +
  theta5*black + theta6*riskchrt + theta7*hivpos;
p1_y0_x0=exp(eta1_y0_x0)/(1+exp(eta1_y0_x0));

eta2_x1_y1=delta0 + delta1 + delta2 + delta3*age +
  delta4*black + delta5*riskchrt + delta6*hivpos;
p2_x1_y1=exp(eta2_x1_y1)/(1+exp(eta2_x1_y1));
eta2_x0_y1=delta0 + delta2 + delta3*age + delta4*black +
  delta5*riskchrt + delta6*hivpos;
p2_x0_y1=exp(eta2_x0_y1)/(1+exp(eta2_x0_y1));
eta2_x1_y0=delta0 + delta1 + delta3*age + delta4*black +
  delta5*riskchrt + delta6*hivpos;
p2_x1_y0=exp(eta2_x1_y0)/(1+exp(eta2_x1_y0));
eta2_x0_y0=delta0 + delta3*age + delta4*black +
```

```
  delta5*riskchrt + delta6*hivpos;
p2_x0_y0=exp(eta2_x0_y0)/(1+exp(eta2_x0_y0));

eta_x1=beta0 + beta1 + beta2*age + beta3*black +
  beta4*riskchrt + beta5*hivpos;
p_x1=exp(eta_x1)/(1+exp(eta_x1));
eta_x0=beta0 + beta2*age + beta3*black +
  beta4*riskchrt + beta5*hivpos;
p_x0=exp(eta_x0)/(1+exp(eta_x0));

eta_XC=gamma0 + gamma1*age + gamma2*black +
  gamma3*riskchrt + gamma4*hivpos;
p_xc=exp(eta_XC)/(1+exp(eta_XC));

likeM=((p1_y1_x1)**ystar*(1-p1_y1_x1)**(1-ystar)*
        (p2_x1_y1)**xstar*(1-p2_x1_y1)**(1-xstar)*
         p_x1*p_xc +
       (p1_y0_x1)**ystar*(1-p1_y0_x1)**(1-ystar)*
        (p2_x1_y0)**xstar*(1-p2_x1_y0)**(1-xstar)*
         (1-p_x1)*p_xc +
       (p1_y1_x0)**ystar*(1-p1_y1_x0)**(1-ystar)*
        (p2_x0_y1)**xstar*(1-p2_x0_y1)**(1-xstar)*
         (p_x0)*(1-p_xc) +
       (p1_y0_x0)**ystar*(1-p1_y0_x0)**(1-ystar)*
        (p2_x0_y0)**xstar*(1-p2_x0_y0)**(1-xstar)*
         (1-p_x0)*(1-p_xc))**(1-val);

eta1=theta0 + theta1*x + theta2*xstar +
  theta3*y + theta4*age + theta5*black +
  theta6*riskchrt + theta7*hivpos;
p1=exp(eta1)/(1+exp(eta1));

eta2=delta0 + delta1*x + delta2*y + delta3*age +
  delta4*black + delta5*riskchrt + delta6*hivpos;
p2=exp(eta2)/(1+exp(eta2));

eta3=beta0 + beta1*x + beta2*age + beta3*black +
  beta4*riskchrt + beta5*hivpos;
p3=exp(eta3)/(1+exp(eta3));

LikeV=(p1**ystar*(1-p1)**(1-ystar)*
        p2**xstar*(1-p2)**(1-xstar)*p3**y*
        (1-p3)**(1-y)*p_xc**x*
        (1-p_xc)**(1-x))**val;
```

```
Like = LikeM*LikeV;
loglik=log(Like);
model ystar~general(loglik);
run;
```

B.1.2 Bayesian OpenBUGS Model Code

```
model {
  for (i in 1:n) {
    BVStatusStar[i] ~ dbern(pBVStatusStar[i])
    TrichStar[i] ~ dbern(pTrichStar[i])
    logit(pBVStatusStar[i]) <- theta0 + theta1*Trich[i] +
      theta2*TrichStar[i] + theta3*BVStatus[i] +theta4*Age[i]+
      theta5*Black[i] + theta6*RiskChrt[i] + theta7*HIVPos[i]
    logit(pTrichStar[i]) <- delta0 + delta1*Trich[i] +
      delta2*BVStatus[i] + delta3*Age[i] + delta4*Black[i] +
      delta5*RiskChrt[i] + delta6*HIVPos[i]
    Trich[i] ~ dbern(PTrich[i])
    BVStatus[i] ~ dbern(PBVStatus[i])
    logit(PBVStatus[i]) <- beta0 + beta1*Trich[i] +
      beta2*Age[i] + beta3*Black[i] + beta4*RiskChrt[i] +
      beta5*HIVPos[i]
    logit(PTrich[i]) <- gamma0 + gamma1*Age[i] +
      gamma2*Black[i] + gamma3*RiskChrt[i] + gamma4*HIVPos[i]
  }
  for (j in 1:nv) {
    BVStatusStarv[j] ~ dbern(pBVStatusStarv[j])
    TrichStarv[j] ~ dbern(pTrichStarv[j])
    logit(pBVStatusStarv[j]) <- theta0 + theta1*Trichv[j] +
      theta2*TrichStarv[j] + theta3*BVStatusv[j] +
      theta4*AgeV[j] + theta5*BlackV[j] +
      theta6*RiskChrtV[j] + theta7*HIVPosV[j]
    logit(pTrichStarv[j]) <- delta0 + delta1*Trichv[j] +
      delta2*BVStatusv[j] + delta3*AgeV[j] +delta4*BlackV[j]+
      delta5*RiskChrtV[j] + delta6*HIVPosV[j]
    Trichv[j] ~ dbern(PTrichv[j])
    BVStatusv[j] ~ dbern(PBVStatusv[j])
    logit(PBVStatusv[j]) <- beta0 + beta1*Trichv[j] +
      beta2*AgeV[j] + beta3*BlackV[j] +
      beta4*RiskChrtV[j] + beta5*HIVPosV[j]
    logit(PTrichv[j]) <- gamma0 + gamma1*AgeV[j] +
      gamma2*BlackV[j] +gamma3*RiskChrtV[j] +gamma4*HIVPosV[j]
```

```
  }

  gamma0 ~ dnorm(0, 0.1)
  gamma1 ~ dnorm(0, 0.1)
  gamma2 ~ dnorm(0, 0.1)
  gamma3 ~ dnorm(0, 0.1)
  gamma4 ~ dnorm(0, 0.1)

  beta0 ~ dnorm(0, 0.1)
  beta1 ~ dnorm(0, 0.1)
  beta2 ~ dnorm(0, 0.1)
  beta3 ~ dnorm(0, 0.1)
  beta4 ~ dnorm(0, 0.1)
  beta5 ~ dnorm(0, 0.1)

  delta0 ~ dnorm(0, 0.1)
  delta1 ~ dnorm(0, 0.1)
  delta2 ~ dnorm(0, 0.1)
  delta3 ~ dnorm(0, 0.1)
  delta4 ~ dnorm(0, 0.1)
  delta5 ~ dnorm(0, 0.1)
  delta6 ~ dnorm(0, 0.1)

  theta0 ~ dnorm(0, 0.1)
  theta1 ~ dnorm(0, 0.1)
  theta2 ~ dnorm(0, 0.1)
  theta3 ~ dnorm(0, 0.1)
  theta4 ~ dnorm(0, 0.1)
  theta5 ~ dnorm(0, 0.1)
  theta6 ~ dnorm(0, 0.1)
  theta7 ~ dnorm(0, 0.1)
  }
```

B.2 Analysis of Simulation Study Discussed in Chapter Three

B.2.1 Frequentist (R) Model Code for the Baseline Model

```
FreqMLE <- function(input){
    beta0 <- input[1]
    beta1 <- input[2]
    beta2 <- input[3]
    delta0 <- input[4]
```

```
delta1 <- input[5]
delta2 <- input[6]
delta3 <- input[7]
gamma0 <- input[8]
gamma1 <- input[9]
theta0 <- input[10]
theta1 <- input[11]
theta2 <- input[12]
theta3 <- input[13]
theta4 <- input[14]

eta1_y1_x1 <- theta0 + theta1 + theta2*X_StarC +
  theta3 + theta4*C_1C
p1_y1_x1 <- exp(eta1_y1_x1)/(1+exp(eta1_y1_x1))
eta1_y0_x1 <- theta0 + theta1 + theta2*X_StarC +
  theta4*C_1C
p1_y0_x1 <- exp(eta1_y0_x1)/(1+exp(eta1_y0_x1))
eta1_y1_x0 <- theta0 + theta2*X_StarC + theta3 +
  theta4*C_1C
p1_y1_x0 <- exp(eta1_y1_x0)/(1+exp(eta1_y1_x0))
eta1_y0_x0 <- theta0 + theta2*X_StarC + theta4*C_1C
p1_y0_x0 <- exp(eta1_y0_x0)/(1+exp(eta1_y0_x0))

eta2_x1_y1 <- delta0 + delta1 + delta2 + delta3*C_1C
p2_x1_y1 <- exp(eta2_x1_y1)/(1+exp(eta2_x1_y1))
eta2_x0_y1 <- delta0 + delta2 + delta3*C_1C
p2_x0_y1 <- exp(eta2_x0_y1)/(1+exp(eta2_x0_y1))
eta2_x1_y0 <- delta0 + delta1 + delta3*C_1C
p2_x1_y0 <- exp(eta2_x1_y0)/(1+exp(eta2_x1_y0))
eta2_x0_y0 <- delta0 + delta3*C_1C
p2_x0_y0 <- exp(eta2_x0_y0)/(1+exp(eta2_x0_y0))

eta_x1 <- beta0 + beta1 + beta2*C_1C
p_x1 <- exp(eta_x1)/(1+exp(eta_x1))
eta_x0 <- beta0 + beta2*C_1C
p_x0 <- exp(eta_x0)/(1+exp(eta_x0))

eta_XC <- gamma0 + gamma1*C_1C
p_xc <- exp(eta_XC)/(1+exp(eta_XC))

LikeM <- (((p1_y1_x1)^Y_StarC)*((1-p1_y1_x1)^(1-Y_StarC))*
  ((p2_x1_y1)^X_StarC)*((1-p2_x1_y1)^(1-X_StarC))*
  (p_x1)*(p_xc) + ((p1_y0_x1)^Y_StarC)*
  ((1-p1_y0_x1)^(1-Y_StarC))*((p2_x1_y0)^X_StarC)*
```

```
      ((1-p2_x1_y0)^(1-X_StarC))*((1-p_x1))*(p_xc) +
      ((p1_y1_x0)^Y_StarC)*((1-p1_y1_x0)^(1-Y_StarC))*
      ((p2_x0_y1)^X_StarC)*((1-p2_x0_y1)^(1-X_StarC))*
      ((p_x0))*(1-p_xc) + ((p1_y0_x0)^Y_StarC)*
      ((1-p1_y0_x0)^(1-Y_StarC))*((p2_x0_y0)^X_StarC)*
      ((1-p2_x0_y0)^(1-X_StarC))*((1-p_x0))*(1-p_xc))^(1-val)

    eta1 <- theta0 + theta1*XC + theta2*X_StarC +
      theta3*YC + theta4*C_1C
    p1 <- exp(eta1)/(1+exp(eta1))
    eta2 <- delta0 + delta1*XC + delta2*YC + delta3*C_1C
    p2 <- exp(eta2)/(1+exp(eta2))
    eta3 <- beta0 + beta1*XC + beta2*C_1C
    p3 <- exp(eta3)/(1+exp(eta3))

    LikeV <- ((p1^Y_StarC)*((1-p1)^(1-Y_StarC))*(p2^X_StarC)*
                ((1-p2)^(1-X_StarC))*(p3^YC)*((1-p3)^(1-YC))*
                (p_xc^XC)*((1-p_xc)^(1-XC)))^val
    sum(log(LikeM*LikeV))
}
```

B.2.2 Bayesian (OpenBUGS/JAGS) Model Code for the Baseline Model

```
BUGSModel <- function(){
  for (i in 1:n) {
    Y_Star[i] ~ dbern(PY_Star[i])
    X_Star[i] ~ dbern(PX_Star[i])
    logit(PY_Star[i]) <- theta0 + theta1*X[i] +
      theta2*X_Star[i] + theta3*Y[i] + theta4*C_1[i]
    logit(PX_Star[i]) <- delta0 + delta1*X[i] +
      delta2*Y[i] + delta3*C_1[i]
    X[i] ~ dbern(PX[i])
    Y[i] ~ dbern(PY[i])
    logit(PY[i]) <- beta0 + beta1*X[i] + beta2*C_1[i]
    logit(PX[i]) <- gamma0 + gamma1*C_1[i]
  }
  for (j in 1:nv) {
    Y_StarV[j] ~ dbern(PY_StarV[j])
    X_StarV[j] ~ dbern(PX_StarV[j])
    logit(PY_StarV[j]) <- theta0 + theta1*X_V[j] +
      theta2*X_StarV[j] + theta3*Y_V[j] + theta4*C_1V[j]
    logit(PX_StarV[j]) <- delta0 + delta1*X_V[j] +
      delta2*Y_V[j] + delta3*C_1V[j]
```

```
    X_V[j] ~ dbern(PX_V[j])
    Y_V[j] ~ dbern(PY_V[j])
    logit(PY_V[j]) <- beta0 + beta1*X_V[j] + beta2*C_1V[j]
    logit(PX_V[j]) <- gamma0 + gamma1*C_1V[j]
  }

  gamma0 ~ dnorm(0, 0.1)
  gamma1 ~ dnorm(0, 0.1)

  beta0 ~ dnorm(0, 0.1)
  beta1 ~ dnorm(0, 0.1)
  beta2 ~ dnorm(0, 0.1)

  delta0 ~ dnorm(0, 0.1)
  delta1 ~ dnorm(0, 0.1)
  delta2 ~ dnorm(0, 0.1)
  delta3 ~ dnorm(0, 0.1)

  theta0 ~ dnorm(0, 0.1)
  theta1 ~ dnorm(0, 0.1)
  theta2 ~ dnorm(0, 0.1)
  theta3 ~ dnorm(0, 0.1)
  theta4 ~ dnorm(0, 0.1)
  }
```

B.3 Analysis of Simulation Study Discussed in Chapter Four

B.3.1 Frequentist (R) Model Code for General Misclassification Model

```
PROC NLMIXED data=Main cov tech=congra;
parms beta0=&beta0. beta1=&beta1. beta2=&beta2.
  gamma0=&gamma0. gamma1=&gamma1. gamma2=&gamma2.
  gamma3=&gamma3. gamma4=&gamma4. gamma5=&gamma5.
  sig2u_i=&sig2u_i. sig2u_iStar=&sig2u_iStar. psi=&psi.;
tau=beta0 + beta1*X_1 + beta2*X_2 + u1;
pY=exp(tau)/(1 + exp(tau));
etas=gamma0 + gamma1*Time2 + gamma2*Time3 + gamma3*Time4 +
    gamma4*X_1 + gamma5 + u2;
etaf=gamma0 + gamma1*Time2 + gamma2*Time3 + gamma3*Time4 +
    gamma4*X_1 + u2;
s=exp(etas)/(1 + exp(etas));
f=1/(1 + exp(etaf));
```

```
pY_Star=(1-f) + (s + f-1)*pY;
like1=((pY_Star**Y_Star)*(1-pY_Star)**(1-Y_Star))**(1-val);
like2=((s*pY)**(Y*Y_Star)*((1-s)*pY)**(Y*(1-Y_Star))*
         ((1-f)*(1-pY))**((1-Y)*Y_Star)*
         (f*(1-pY))**((1-Y)*(1-Y_Star)))**val;
like = like1*like2;
loglik=log(like);
model Y_Star ~ general(loglik);
random u1 u2~normal([0,0],[sig2u_i,psi,sig2u_iStar]) subject=id;
title 'General Misclassification';
run; quit;
```

B.3.2 Frequentist (R) Model Code for ICD Misclassification Model

```
PROC NLMIXED data=Main cov tech=trureg;
parms beta0=&beta0. beta1=&beta1. beta2=&beta2.
  gamma0=&gamma0. gamma1=&gamma1. gamma2=&gamma2.
  gamma3=&gamma3. gamma4=&gamma4. gamma5=&gamma5.
  sig2u_i=&sig2u_i. sig2u_iStar=&sig2u_iStar.;
tau=beta0 + beta1*X_1 + beta2*X_2 + u1;
pY=exp(tau)/(1 + exp(tau));
etas=gamma0 + gamma1*Time2 + gamma2*Time3 + gamma3*Time4 +
    gamma4*X_1 + gamma5 + u2;
etaf=gamma0 + gamma1*Time2 + gamma2*Time3 + gamma3*Time4 +
    gamma4*X_1 + u2;
s=exp(etas)/(1 + exp(etas));
f=1/(1 + exp(etaf));
pY_Star=(1-f) + (s + f-1)*pY;

like1=((pY_Star**Y_Star)*(1-pY_Star)**(1-Y_Star))**(1-val);
like2=((s*pY)**(Y*Y_Star)*((1-s)*pY)**(Y*(1-Y_Star))*
         ((1-f)*(1-pY))**((1-Y)*Y_Star)*
         (f*(1-pY))**((1-Y)*(1-Y_Star)))**val;
like = like1*like2;
loglik=log(like);
model Y_Star ~ general(loglik);
random u1 u2~normal([0,0],[sig2u_i,0,sig2u_iStar]) subject=id;
title 'Independent Correlated Differential Misclassification';
run;
```

B.3.3 Frequentist (R) Model Code for IUD Misclassification Model

```
PROC NLMIXED data=Main cov tech=congra;
parms beta0=&beta0. beta1=&beta1. beta2=&beta2.
  gamma0=&gamma0. gamma1=&gamma1. gamma2=&gamma2.
  gamma3=&gamma3. gamma4=&gamma4. gamma5=&gamma5.
  sig2u_i=&sig2u_i.;
pY=exp(tau)/(1 + exp(tau));
etas=gamma0 + gamma1*Time2 + gamma2*Time3 + gamma3*Time4 +
    gamma4*X_1 + gamma5;
etaf=gamma0 + gamma1*Time2 + gamma2*Time3 + gamma3*Time4 +
    gamma4*X_1;
s=exp(etas)/(1 + exp(etas));
f=1/(1 + exp(etaf));
pY_Star=(1-f) + (s + f-1)*pY;
like1=((pY_Star**Y_Star)*(1-pY_Star)**(1-Y_Star))**(1-val);
like2=((s*pY)**(Y*Y_Star)*((1-s)*pY)**(Y*(1-Y_Star))*
      ((1-f)*(1-pY))**((1-Y)*Y_Star)*
(f*(1-pY))**((1-Y)*(1-Y_Star)))**val;
like = like1*like2;
loglik=log(like);
model Y_Star ~ general(loglik);
random u1 ~ normal(0,sig2u_i) subject=id;
title 'Uncorrelated Differential Misclassification';
run;
```

B.3.4 Frequentist (R) Model Code for ND Misclassification Model

```
PROC NLMIXED data=Main cov tech=congra;
parms beta0=&beta0. beta1=&beta1. beta2=&beta2.
  gamma0=&gamma0. gamma5=&gamma5.
  sig2u_i=&sig2u_i.;
tau=beta0 + beta1*X_1 + beta2*X_2 + u1;
pY=exp(tau)/(1 + exp(tau));
etas=gamma0 + gamma5;
etaf=gamma0;
s=exp(etas)/(1 + exp(etas));
f=1/(1 + exp(etaf));
pY_Star=(1-f) + (s + f-1)*pY;

like1=((pY_Star**Y_Star)*(1-pY_Star)**(1-Y_Star))**(1-val);
like2=((s*pY)**(Y*Y_Star)*((1-s)*pY)**(Y*(1-Y_Star))*
      ((1-f)*(1-pY))**((1-Y)*Y_Star)*
```

```
(f*(1-pY))**((1-Y)*(1-Y_Star)))**val;
like = like1*like2;
loglik=log(like);

model Y_Star ~ general(loglik);
random u1 ~ normal(0,sig2u_i) subject=id;
title 'Non-Differential Misclassification';
run;
```

B.3.5 NIMBLE Model Code for the General Misclassification Model

```
general <- nimbleCode({
  tau_i ~ dgamma(2, 2)
  tau_iStar ~ dgamma(2, 2)
  sig2u_i <- 1/tau_i
  sig2u_iStar <- 1/tau_iStar

  rho ~ dunif(-1, 1)
  psi <- rho*sqrt(sig2u_i)*sqrt(sig2u_iStar)

  beta0 ~ dnorm(0, 0.1)
  beta1 ~ dnorm(0, 0.1)
  beta2 ~ dnorm(0, 0.1)
  gamma0 ~ dnorm(0, 0.1)
  gamma1 ~ dnorm(0, 0.1)
  gamma2 ~ dnorm(0, 0.1)
  gamma3 ~ dnorm(0, 0.1)
  gamma4 ~ dnorm(0, 0.1)
  gamma5 ~ dnorm(0, 0.1)
  mu[1] <- 0
  mu[2] <- 0
  covar[1,1] <- sig2u_i
  covar[1,2] <- psi
  covar[2,1] <- psi
  covar[2,2] <- sig2u_iStar
  for (j in 1:(n+nv)){
    u[j, 1:2] ~ dmnorm(mu[1:2], cov = covar[1:2,1:2])
  }
  for (j in 1:nv*J){
    Y_StarV[j] ~ dbern(pY_StarV[j])
    logit(pY_StarV[j]) <- gamma0 + gamma1*Time2V[j] +
      gamma2*Time3V[j] + gamma3*Time4V[j] +
      gamma4*X_1V[j] + gamma5*YV[j] + u[SubjV[j],2]
```

```
    YV[j] ~ dbern(pYV[j])
    logit(pYV[j])<-beta0 + beta1*X_1V[j] +
      beta2*X_2V[j] + u[SubjV[j], 1]
  }
  for (j in 1:n*J){
    Y_Star[j] ~ dbern(pY_Star[j])
    logit(pY_Star[j]) <- gamma0 + gamma1*Time2[j] +
      gamma2*Time3[j] + gamma3*Time4[j] +
      gamma4*X_1[j] + gamma5*Y[j] + u[Subj[j],2]
    Y[j] ~ dbern(pY[j])
    logit(pY[j])<-beta0 + beta1*X_1[j] +
      beta2*X_2[j] + u[Subj[j],1]
  }
})
```

B.3.6 NIMBLE Model Code for the ICD Misclassification Model

```
Ind_Corr <- nimbleCode({
  tau_i ~ dgamma(2, 2)
  tau_iStar ~ dgamma(2, 2)
  sig2u_i <- 1/tau_i
  sig2u_iStar <- 1/tau_iStar

  beta0 ~ dnorm(0, 0.1)
  beta1 ~ dnorm(0, 0.1)
  beta2 ~ dnorm(0, 0.1)
  gamma0 ~ dnorm(0, 0.1)
  gamma1 ~ dnorm(0, 0.1)
  gamma2 ~ dnorm(0, 0.1)
  gamma3 ~ dnorm(0, 0.1)
  gamma4 ~ dnorm(0, 0.1)
  gamma5 ~ dnorm(0, 0.1)
  mu[1] <- 0
  mu[2] <- 0
  covar[1,1] <- sig2u_i
  covar[1,2] <- 0
  covar[2,1] <- 0
  covar[2,2] <- sig2u_iStar
  for (j in 1:(n+nv)){
    u[j, 1:2] ~ dmnorm(mu[1:2], cov = covar[1:2,1:2])
  }
  for (j in 1:nv*J){
    Y_StarV[j] ~ dbern(pY_StarV[j])
```

```
    logit(pY_StarV[j]) <- gamma0 + gamma1*Time2V[j] +
      gamma2*Time3V[j] + gamma3*Time4V[j] +
      gamma4*X_1V[j] + gamma5*YV[j] + u[SubjV[j],2]
    YV[j] ~ dbern(pYV[j])
    logit(pYV[j])<-beta0 + beta1*X_1V[j] +
      beta2*X_2V[j] + u[SubjV[j], 1]
  }
  for (j in 1:n*J){
    Y_Star[j] ~ dbern(pY_Star[j])
    logit(pY_Star[j]) <- gamma0 + gamma1*Time2[j] +
      gamma2*Time3[j] + gamma3*Time4[j] + gamma4*X_1[j] +
      gamma5*Y[j] + u[Subj[j],2]
    Y[j] ~ dbern(pY[j])
    logit(pY[j])<-beta0 + beta1*X_1[j] +
      beta2*X_2[j] + u[Subj[j],1]
  }
})
```

B.3.7 NIMBLE Model Code for the IUD Misclassification Model

```
UnCorr_Diff <- nimbleCode({
  tau_i ~ dgamma(2, 2)
  sig2u_i <- 1/tau_i

  beta0 ~ dnorm(0, 0.1)
  beta1 ~ dnorm(0, 0.1)
  beta2 ~ dnorm(0, 0.1)
  gamma0 ~ dnorm(0, 0.1)
  gamma1 ~ dnorm(0, 0.1)
  gamma2 ~ dnorm(0, 0.1)
  gamma3 ~ dnorm(0, 0.1)
  gamma4 ~ dnorm(0, 0.1)
  gamma5 ~ dnorm(0, 0.1)
  for (j in 1:(n+nv)){
    u[j] ~ dnorm(0, var = sig2u_i)
  }
  for (j in 1:nv*J){
    Y_StarV[j] ~ dbern(pY_StarV[j])
    logit(pY_StarV[j]) <- gamma0 + gamma1*Time2V[j] +
      gamma2*Time3V[j] + gamma3*Time4V[j] +
      gamma4*X_1V[j] + gamma5*YV[j]
    YV[j] ~ dbern(pYV[j])
    logit(pYV[j])<-beta0 + beta1*X_1V[j] +
```

```
      beta2*X_2V[j] + u[SubjV[j]]
  }
  for (j in 1:n*J){
    Y_Star[j] ~ dbern(pY_Star[j])
    logit(pY_Star[j]) <- gamma0 + gamma1*Time2[j] +
      gamma2*Time3[j] + gamma3*Time4[j] +
      gamma4*X_1[j] + gamma5*Y[j]
    Y[j] ~ dbern(pY[j])
    logit(pY[j])<-beta0 + beta1*X_1[j] +
      beta2*X_2[j] + u[Subj[j]]
  }
})
```

B.3.8 NIMBLE Model Code for the ND Misclassification Model

```
Non_Diff <- nimbleCode({
  tau_i ~ dgamma(2, 2)
  sig2u_i <- 1/tau_i

  beta0 ~ dnorm(0, 0.1)
  beta1 ~ dnorm(0, 0.1)
  beta2 ~ dnorm(0, 0.1)
  gamma0 ~ dnorm(0, 0.1)
  gamma5 ~ dnorm(0, 0.1)
  for (j in 1:(n+nv)){
    u[j] ~ dnorm(0, var = sig2u_i)
  }
  for (j in 1:nv*J){
    Y_StarV[j] ~ dbern(pY_StarV[j])
    logit(pY_StarV[j]) <- gamma0 + gamma5*YV[j]
    YV[j] ~ dbern(pYV[j])
    logit(pYV[j])<-beta0 + beta1*X_1V[j] +
      beta2*X_2V[j] + u[SubjV[j]]
  }
  for (j in 1:n*J){
    Y_Star[j] ~ dbern(pY_Star[j])
    logit(pY_Star[j]) <- gamma0 + gamma5*Y[j]
    Y[j] ~ dbern(pY[j])
    logit(pY[j])<-beta0 + beta1*X_1[j] +
      beta2*X_2[j] + u[Subj[j]]
  }
})
```

APPENDIX C

ESS Settings and Code

C.1 ESS for Priors of Dependent Differential Misclassification Using Morita et al.[29] Provided ESS Calculator and Settings

Refer to Morita et al.[29] for a complete description of the formula sourced in the code below (`ESS_RegressionCalculator.R`); we used this calculator for each of the logistic regression models that make up dependent differential misclassification. The number of covariates varied between each of the 4 models and were appropriately set for use with the function inputs. We set $M = 50$ for each model since we assumed we would not see an ESS higher than that for any one particular model. We used 500 simulations ($NumSims = 500$) to replicate the number of replications in our simulation study. We did not investigate the parameter subsets since it was not within the scope of this investigation. We then replicated this work for each of the four priors used throughout Chapter Three: $N(0, \sigma^2 = 10)$, $N(\mu, \sigma^2 = 10)$, $N(\mu, \sigma^2 = 4)$, and $N(\mu, \sigma^2 = 1)$. Based upon the corresponding literature, the priors were set to $N(0, \sigma^2 = 1000)$ for any prior on a covariate that did not exist for the model at hand.

```
source("ESS_RegressionCalculator.R")
ESS_Function <- function(Guesses, Num_cov, Distribution_Type,
     Distribution_parm_1, Distribution_parm_2,
     M, NumSims, theta_sub1, theta_sub2){
    Reg_model <- 2
    Prior_0 <- c(Distribution_Type[1],
     Distribution_parm_1[1], Distribution_parm_2[1])
    Prior_1 <- c(Distribution_Type[2],
     Distribution_parm_1[2], Distribution_parm_2[2])
    Prior_2 <- c(Distribution_Type[3],
     Distribution_parm_1[3], Distribution_parm_2[3])
    Prior_3 <- c(Distribution_Type[4],
     Distribution_parm_1[4], Distribution_parm_2[4])
    Prior_4 <- c(Distribution_Type[5],
```

```
     Distribution_parm_1[5], Distribution_parm_2[5])
    Prior_5 <- c(Distribution_Type[6],
     Distribution_parm_1[6], Distribution_parm_2[6])
    Prior_6 <- c(Distribution_Type[7],
     Distribution_parm_1[7], Distribution_parm_2[7])
    Prior_7 <- c(Distribution_Type[8],
     Distribution_parm_1[8], Distribution_parm_2[8])
    Prior_8 <- c(Distribution_Type[9],
     Distribution_parm_1[9], Distribution_parm_2[9])
    Prior_9 <- c(Distribution_Type[10],
     Distribution_parm_1[10], Distribution_parm_2[10])
    Prior_10 <- c(Distribution_Type[11],
     Distribution_parm_1[11], Distribution_parm_2[11])
    Prior_11 <- c(Distribution_Type[12],
      Distribution_parm_1[12], Distribution_parm_2[12])
  resultList <-
     ESS_RegressionCalc ( Reg_model, Num_cov,
       Prior_0, Prior_1, Prior_2, Prior_3, Prior_4, Prior_5,
       Prior_6, Prior_7, Prior_8, Prior_9, Prior_10, Prior_11,
       M, NumSims, theta_sub1, theta_sub2)
  as.data.frame(cbind("Calculated ESS"=resultList,
                      "Estimated ESS"=Guesses),
    row.names = c("Whole \theta", "Sub-vector 1",
                  "Sub-vector 2"))
}
Theta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 4, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- rep(0, 12),
      Distribution_parm_2 <- c(rep(10, 5), rep(1000, 7)),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Centered_Theta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
    Num_cov <- 4, Distribution_Type <- rep(1, 12),
    Distribution_parm_1 <- c(-1, 1, 2, 1, -1, rep(0,7)),
    Distribution_parm_2 <- c(rep(10, 5), rep(1000, 7)),
    M <- 50, NumSims <- 500,
    theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
    theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Narrowed_Theta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
    Num_cov <- 4,
    Distribution_Type <- rep(1, 12),
    Distribution_parm_1 <- c(-1, 1, 2, 1, -1, rep(0,7)),
    Distribution_parm_2 <- c(rep(4, 5), rep(1000, 7)),
```

```
    M <- 50, NumSims <- 500,
    theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
    theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Informative_Theta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 4, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(-1, 1, 2, 1, -1, rep(0,7)),
      Distribution_parm_2 <- c(rep(1, 5), rep(1000, 7)),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Delta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 3, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- rep(0, 12),
      Distribution_parm_2 <- c(rep(10, Num_cov+1),
                               rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Centered_Delta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 3, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(-3, 1.5, 1.5, 1,
                               rep(0,11-Num_cov)),
      Distribution_parm_2 <- c(rep(10, Num_cov+1),
                               rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Narrowed_Delta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 3, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(-3, 1.5, 1.5, 1,
                               rep(0,11-Num_cov)),
      Distribution_parm_2 <- c(rep(4, Num_cov+1),
                               rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Informative_Delta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 3, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(-3, 1.5, 1.5, 1,
                               rep(0,11-Num_cov)),
      Distribution_parm_2 <- c(rep(1, Num_cov+1),
                               rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
```

```
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Beta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 2, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- rep(0, 12),
      Distribution_parm_2 <- c(rep(10, Num_cov+1),
                               rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Centered_Beta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 2, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(1, 1, -1, rep(0,11-Num_cov)),
      Distribution_parm_2 <- c(rep(10, Num_cov+1),
                               rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Narrowed_Beta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 2, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(1, 1, -1, rep(0,11-Num_cov)),
      Distribution_parm_2 <- c(rep(4, Num_cov+1),
                               rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Informative_Beta <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
     Num_cov <- 2, Distribution_Type <- rep(1, 12),
     Distribution_parm_1 <- c(1, 1, -1, rep(0,11-Num_cov)),
     Distribution_parm_2 <- c(rep(1, Num_cov+1),
                              rep(1000, 12-(Num_cov+1))),
     M <- 50, NumSims <- 500,
     theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
     theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Gamma <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 1, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- rep(0, 12),
      Distribution_parm_2 <- c(rep(10, Num_cov+1),
      rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Centered_Gamma <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 1, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(-1, 1, rep(0,11-Num_cov)),
```

```
      Distribution_parm_2 <- c(rep(10, Num_cov+1),
      rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Narrowed_Gamma <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 1, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(-1, 1, rep(0,11-Num_cov)),
      Distribution_parm_2 <- c(rep(4, Num_cov+1),
      rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
Informative_Gamma <- ESS_Function(Guesses <- c(0.8, 0.3, 0.5),
      Num_cov <- 1, Distribution_Type <- rep(1, 12),
      Distribution_parm_1 <- c(-1, 1, rep(0,11-Num_cov)),
      Distribution_parm_2 <- c(rep(1, Num_cov+1),
      rep(1000, 12-(Num_cov+1))),
      M <- 50, NumSims <- 500,
      theta_sub1 <- c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
      theta_sub2 <- c(0, rep(1, Num_cov),rep(0, 11-Num_cov)))
(ESS_Baseline <-
    as.numeric(Theta$`Calculated ESS`[1]) +
    as.numeric(Delta$`Calculated ESS`[1]) +
    as.numeric(Beta$`Calculated ESS`[1]) +
    as.numeric(Gamma$`Calculated ESS`[1]))
(ESS_Centered <-
    as.numeric(Centered_Theta$`Calculated ESS`[1]) +
    as.numeric(Centered_Delta$`Calculated ESS`[1]) +
    as.numeric(Centered_Beta$`Calculated ESS`[1]) +
    as.numeric(Centered_Gamma$`Calculated ESS`[1]))
(ESS_Narrowed <-
    as.numeric(Narrowed_Theta$`Calculated ESS`[1]) +
    as.numeric(Narrowed_Delta$`Calculated ESS`[1]) +
    as.numeric(Narrowed_Beta$`Calculated ESS`[1]) +
    as.numeric(Narrowed_Gamma$`Calculated ESS`[1]))
(ESS_Informative <-
    as.numeric(Informative_Theta$`Calculated ESS`[1]) +
    as.numeric(Informative_Delta$`Calculated ESS`[1]) +
    as.numeric(Informative_Beta$`Calculated ESS`[1]) +
    as.numeric(Informative_Gamma$`Calculated ESS`[1]))
```

REFERENCES

[1] Ando, T. and Tsay, R. (2010). Predictive likelihood for Bayesian model selection and averaging. *International Journal of Forecasting*, 26(4):744–763.

[2] Barron, B. A. (1977). The Effects of Misclassification on the Estimation of Relative Risk. *Biometrics*, 33(2):414.

[3] Bedrick, E., Johnson, W., and Christensen (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, 91(436):1450–1460.

[4] Brenner, H. and Gefeller, O. (1993). Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. *American Journal of Epidemiology*, 138(11):1007–1015.

[5] Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.

[6] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*, volume 105. Chapman & Hall/CRC, Boca Raton, FL, 2nd edition.

[7] Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series. Duxbury, 2nd edition.

[8] Christensen, R., Johnson, W., Branscum, A. J., and Hanson, T. E. (2011). *Bayesian Ideas and Data Analysis*. CRC press.

[9] Copeland, K. T., Checkoway, H., McMichael, A. J., and Holbrook, R. H. (1977). Bias due to misclassification in the estimation of relative risk. *American Journal of Epidemiology*, 105(5):488–495.

[10] Edwards, J. K., Cole, S. R., Troester, M. A., and Richardson, D. B. (2013). Accounting for Misclassified Outcomes in Binary Regression Models Using Multiple Imputation With Internal Validation Data. *American Journal of Epidemiology*, 177(9):904–912.

[11] Fox, M. P., Lash, T. L., and Greenland, S. (2005). A method to automate probabilistic sensitivity analyses of misclassified binary variables. *International Journal of Epidemiology*, 34(6):1370–1376.

[12] Gelman, A. (2004). Prior distributions for variance parameters in hierarchical models. Technical report, EERI Research Paper Series.

[13] Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, pages 1360–1383.

[14] Gerlach, R. and Stamey, J. (2007). Bayesian model selection for logistic regression with misclassified outcomes. *Statistical Modelling: An International Journal*, 7(3):255–273.

[15] Goldstein, N. D., Burstyn, I., Newbern, E. C., Tabb, L. P., Gutowski, J., and Welles, S. L. (2016). Bayesian Correction of Misclassification of Pertussis in Vaccine Effectiveness Studies: How Much Does Underreporting Matter? *American Journal of Epidemiology*, 183(11):1063–1070.

[16] Gosling, A. and Saloniki, E.-C. (2014). CORRECTION OF MISCLASSIFICATION ERROR IN DISABILITY RATES: CORRECTION OF MISCLASSIFICATION ERROR IN DISABILITY RATES. *Health Economics*, 23(9):1084–1097.

[17] Green, M. S. (1983). Use of predictive value to adjust relative risk estimates biased by misclassification of outcome status. *American Journal of Epidemiology*, 117(1):98–105.

[18] Greenland, S. (1988). Variance estimation for epidemiologic effect estimates under misclassification. *Statistics in Medicine*, 7(7):745–757.

[19] Greenland, S. (2008). Maximum-likelihood and closed-form estimators of epidemiologic measures under misclassification. *Journal of Statistical Planning and Inference*, 138(2):528–538.

[20] Holcroft, C. A., Rotnitzky, A., and Robins, J. M. (1997). Efficient estimation of regression parameters from multistage studies with validation of outcome and covariates. *Journal of Statistical Planning and Inference*, 65(2):349–374.

[21] Lash, T. L. and Fink, A. K. (2003). Semi-Automated Sensitivity Analysis to Assess Systematic Errors in Observational Data:. *Epidemiology*, 14(4):451–458.

[22] Lyles, R. H. (2002). A note on estimating crude odds ratios in case-control studies with differentially misclassified exposure. *Biometrics*, 58(4):1034–1036.

[23] Lyles, R. H. and Lin, J. (2010). Sensitivity analysis for misclassification in logistic regression via likelihood methods and predictive value weighting. *Statistics in Medicine*, 29(22):2297–2309.

[24] Lyles, R. H., Tang, L., Superak, H. M., King, C. C., Celentano, D. D., Lo, Y., and Sobel, J. D. (2011). Validation Data-based Adjustments for Outcome Misclassification in Logistic Regression: An Illustration. *Epidemiology*, 22(4):589–597.

[25] Lyles, R. H., Williamson, J. M., Lin, H.-M., and Heilig, C. M. (2005). Extending McNemar's Test: Estimation and Inference When Paired Binary Outcome Data Are Misclassified. *Biometrics*, 61(1):287–294.

[26] Magder, L. S. and Hughes, J. P. (1997). Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology*, 146(2):195–203.

[27] Marshall, R. J. (1990). Validation study methods for estimating exposure proportions and odds ratios with misclassified data. *Journal of Clinical Epidemiology*, 43(9):941–947.

[28] McInturff, P., Johnson, W. O., Cowling, D., and Gardner, I. A. (2004). Modelling risk when binary outcomes are subject to error. *Statistics in medicine*, 23(7):1095–1109.

[29] Morita, S., Thall, P. F., and Müller, P. (2008). Determining the Effective Sample Size of a Parametric Prior. *Biometrics*, 64(2):595–602.

[30] Morrissey, M. J. and Spiegelman, D. (1999). Matrix methods for estimating odds ratios with misclassified exposure data: extensions and comparisons. *Biometrics*, 55(2):338–344.

[31] Neuhaus, J. M. (1999). Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86(4):843–855.

[32] Neuhaus, J. M. (2002). Analysis of clustered and longitudinal binary data subject to response misclassification. *Biometrics*, 58(3):675–683.

[33] NIMBLE Development Team (2017). Nimble: An r package for programming with bugs models, version 0.6-8.

[34] Paulino, D. C., Soares, P., and Neuhaus, J. (2003). Binomial regression with misclassification. *Biometrics*, 59(3):670–675.

[35] Pepe, M. S. (1992). Inference Using Surrogate Outcome Data and a Validation Sample. *Biometrika*, 79(2):355.

[36] Richardson, S. and Gilks, W. R. (1993). A bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology*, 138(6):430–442.

[37] Smith, D. K., Warren, D. L., Vlahov, D., Schuman, P., Stein, M. D., Greenberg, B. L., Holmberg, S. D., and Group, H. I. V. E. R. S. (1997). Design and baseline participant characteristics of the Human Immunodeficiency Virus Epidemiology Research (HER) Study: a prospective cohort study of human immunodeficiency virus infection in US women. *American Journal of Epidemiology*, 146(6):459–469.

[38] Tang, L., Lyles, R. H., King, C. C., Celentano, D. D., and Lo, Y. (2015a). Binary regression with differentially misclassified response and exposure variables. *Statistics in Medicine*, 34(9):1605–1620.

[39] Tang, L., Lyles, R. H., King, C. C., Hogan, J. W., and Lo, Y. (2015b). Regression analysis for differentially misclassified correlated binary outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(3):433–449.

[40] Tang, L., Lyles, R. H., Ye, Y., Lo, Y., and King, C. C. (2013). Extended Matrix and Inverse Matrix Methods Utilizing Internal Validation Data When Both Disease and Exposure Status Are Misclassified. *Epidemiologic methods*, 2(1):49–66.

[41] Tomiyama, A. J., Hunger, J. M., Nguyen-Cuu, J., and Wells, C. (2016). Misclassification of cardiometabolic health when using body mass index categories in NHANES 2005-2012. *International Journal of Obesity*, 40(5):883–886.

[42] Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

[43] Wolfinger, R. D. (1999). Fitting nonlinear mixed models with the new NLMIXED procedure. In *Proceedings of the 24th Annual SAS Users Group International Conference (SUGI 24)*, pages 278–284.