ABSTRACT

Topics in Bayesian Models with Ordered Parameters: Response Misclassification, Covariate Misclassification, and Sample Size Determination

Kristen M. Tecson, Ph.D.

Chairperson: John W. Seaman, Jr.

Researchers often analyze data assuming models with constrained parameters. Order constrained parameters are of particular interest. In this dissertation, we examine Bayesian models which incorporate ordered parameters. We investigate *ordered differential response misclassification* in a logistic regression model and provide an adjustment for it using a conditional prior structure. We examine a parametric Bayesian Weibull proportional hazards model with *ordered covariate misclassification* and provide an adjustment for it. Finally, we consider *informative hypotheses* (Hoijtink, 2012) and perform sample size determination for this problem using the two priors approach of Brutti et al. (2008).

Topics in Bayesian Models with Ordered Parameters: Response Misclassification, Covariate Misclassification, and Sample Size Determination

by

Kristen M. Tecson, B.S., M.S.

A Dissertation

Approved by the Department of Statistical Science

_____

Jack D. Tubbs, Ph.D., Chairperson

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

Approved by the Dissertation Committee

_____

John W. Seaman, Jr., Ph.D., Chairperson

_____

David J. Kahle, Ph.D.

_____

Stephen T. McClain, Ph.D.

_____

James D. Stamey, Ph.D.

_____

Jack D. Tubbs, Ph.D.

Accepted by the Graduate School
August 2015

_____

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

Thank you to my committee members for your valuable insight on this research. I am forever grateful to you, Dr. Seaman, for choosing to work with me and for your guidance and wisdom throughout this process. Without you, I would not have learned the rules to the 'refrigerator magnet game,' let alone have a dissertation to show for it. Thank you for your interest and support in both my academic and personal endeavors.

I am so thankful for my professors and peers in the department. It has been an absolute privilege to spend the past few years learning from and with you. Thank you especially to Courtney, Joyce, Justin, and Michelle, for your much needed friendship and collaboration.

Thank you, Dr. Hill and Dr. Maddox, for convincing me to attend graduate school and for supporting me through my worst of days.

Thank you, Sunni, Kathleen, and all of my colleagues, for providing me the opportunity to gain professional experience alongside you.

Thank you to my parents and grandparents for making me value education from an early age. Thank you, Dad, for giving me 'hard' math problems upon my request as a child. Thank you, Mom, for teaching by example the love of reading. It is the coupling of your strengths that enables me to thrive.

Thank you, Logan, for agreeing to begin our marriage in graduate school and thank you always for your patience, love, and encouragement.

DEDICATION

To my parents, who always said they'd call me 'Doctor' someday.

To Logan, for our future.

CHAPTER ONE

Introduction

Order constrained models have interesting properties in both hypothesis testing and estimation, across a broad range of applications. These models have been thoroughly studied in both the frequentist and Bayesian literature. For an overview, see, for example, Robertson et al. (1988), Dunson and Neelon (2003), or Silvapulle and Sen (2004).

The Bayesian approach to models with constrained parameters affords the incorporation of prior information and ease of inference for functions of parameters. Importantly, the latter does not require recourse to large sample approximations typically needed in frequentist inference. In this dissertation, we utilize Bayesian methods to explore special topics in models with order restrictions on parameters. Specifically, we examine ordered differential response misclassification in logistic regression, a Weibull proportional hazards model with a covariate subject to ordered misclassification, and sample size determination under "informative hypotheses" (Hoijtink, 2012). The dissertation is organized as follows.

In Chapter Two, we explore ordered differential response misclassification in a logistic regression setting. We introduce this concept through an example involving racial differences in self-reported mammography use (Njai et al., 2011). This Bayesian logistic regression model has a misclassified response (self-reported mammography) with one perfectly recorded binary covariate (race). We then describe a method comprised of a conditional and marginal prior structure to adjust for this misclassification. We perform simulations to compare the results of this adjustment to that of the differential adjustment, which removes bias, but ignores order.

In Chapter Three, we explore covariate misclassification under constraints. Initially, we build a Weibull proportional hazards regression model with one binary covariate subject to ordered misclassification. We extend to a model with two binary

covariates, one of which is subject to ordered misclassification, depending on the levels of the perfectly recorded covariate. We use the conditional and marginal prior structure to adjust for the misclassification and compare the results to a naive model, an unordered misclassification adjusted model, and a perfectly classified model.

In Chapter Four, we introduce "informative hypotheses" (Hoijtink, 2012) through a oneway analysis of variance example. We perform Bayesian sample size determination while testing between an informative hypothesis and its complement using an empirical error rate criterion and the two-priors approach of Brutti et al. (2008). We compare results to those in Van Rossum et al. (2013) and extend the simulation to other dimensions, priors, and informative hypotheses.

In Chapter Five, we summarize the statistical problems and results presented in this dissertation. Additionally, we discuss plans for future work.

CHAPTER TWO

Bayesian Adjustments for Ordered Differential Response Misclassification

2.1 Introduction

2.1.1 Problem Overview

Statistical inference is invariably complicated when data are incomplete or imperfect. Subjects drop out of studies or fail to answer parts of questionnaires. Continuous responses like calorie intake may be mis-measured in diet records. Individuals may be misclassified as not having a disease, when in fact they do. In this chapter, we are concerned with the effect that ordered response misclassification may have in logistic regression.

Misclassified data can be problematic in statistical inference if not addressed. Apparent trends may be false, true trends may go undiscovered, and parameter estimates will be biased with underestimated standard deviations (Gustafson, 2003). For these reasons, there is a large literature on adjusting analyses for response misclassification. Before a correction can be performed, it is necessary to determine which type of misclassification is present in the data.

Suppose a test yields a binary response. Let $T+$ and $T-$ represent tests with positive and negative results, respectively. Let a subject's true status be $D+$ or $D-$, indicating positive and negative, respectively. Then *sensitivity* is $\eta = P(T+|D+)$ and *specificity* is $\theta = P(T-|D-)$. It is often convenient to consider outcomes in terms of false negatives and false positives. These quantities have the relationship

$$P(\text{False Negative}) = P(T-|D+) = 1-\eta$$

and

$$P(\text{False Positive}) = P(T+|D-) = 1-\theta.$$

When response misclassification is independent of covariate values, the misclassification is *non-differential*. For example, suppose we have data regarding employees on probation for violating their company's at-work alcohol consumption policy. We wish to determine the probability of a probational employee consuming alcohol based on his or her career field. Suppose the test used to determine alcohol consumption is fallible, yielding a false positive rate of 5%, regardless of career field. This misclassification is non-differential. Alternatively, if response misclassification depends on covariate values, it is *differential*. In this example, suppose the test yields a false positive rate 7% of the time for subjects in health care related fields and 2% of the time for subjects who are not in health care related fields.[1] This misclassification is differential. Many papers analyze data with these distinguishing characteristics. For an overview see, for example, Carroll et al. (2006) and Gustafson (2003).

Suppose it is known that the rates of misclassification have a distinct order which depends on covariate level. We refer to this as *ordered differential response misclassification*. Just as adjusting for non-differential misclassification is not sufficient when differential misclassification is present, incorporating an adjustment for differential misclassification in a model for data with ordered differential response misclassification may not provide the best solution. Additionally, incorporating the differential misclassification adjustment into a model with ordered differential misclassification may be inappropriate at times, a possibility that is explored in Section 2.4.3. To our knowledge, this issue has not been addressed in the literature.

Ordered differential response misclassification can occur in survey research. If the survey is administered orally, the respondent may intentionally lie about an answer in an effort to please or displease the survey administrator. Additionally, if the survey is taken in private, the respondent may unintentionally record an incorrect answer due to inaccurate memory recall or confusion of the question's wording or

[1] This example is adapted from information regarding an ethyl glucuronide test, which is affected by the use of alcohol-based hand sanitizers, a product that health care employees are exposed to at a much higher rate than employees in other career fields (Kirn, 2006).

response scale (Alwin, 2014). Misclassification may occur at different rates for various subpopulations of respondents, thus inducing the order.

2.1.2 Example: Self-Reported Mammography Use

The Behavioral Risk Factor Surveillance System (BRFSS) is a survey administered via telephone to women to determine who had a mammogram within the past two years. Njai et al. (2011) use BRFSS data from 1995 to 2006 to determine if response misclassification explains some of the racial differences in self-reported mammography use. Njai et al. (2011) determine the sensitivity, (η_i), and specificity, (θ_i), of self-reported mammography among African American $(i = A)$ and Caucasian $(i = C)$ women. They use sensitivity estimates $\hat{\eta}_A = \hat{\eta}_C = 0.97$ for both races. For specificity, they use the estimates $\hat{\theta}_A = 0.49$ and $\hat{\theta}_C = 0.62$ (Rauscher, 2008). In this example, the sensitivity is constant between both races; however, the specificity is ordered such that $\theta_A < \theta_C$. In our analyses, we treat this order as though $P(\theta_A < \theta_C) = 1$.

After adjusting for misclassification in a frequentist manner for the 2006 survey, Njai et al. (2011) estimate the true percentage of women who had a mammogram in the past two years to be $\hat{\pi}_A = 59\%$ for African American women and $\hat{\pi}_C = 65\%$ for Caucasian women. These numbers are lower than $\hat{p}_A = 78\%$ and $\hat{p}_C = 77\%$ obtained via self-reporting by African American and Caucasian women, respectively. This result makes sense intuitively due to the inverse relationship between false positive recordings and specificity. These summary statistics are displayed in Table 2.1.

Table 2.1: Mammography Use Statistics.

Race	$\hat{\eta}$	$\hat{\theta}$	$\hat{p}$	$\hat{\pi}$
African American	0.97	0.49	0.78	0.59
Caucasian	0.97	0.62	0.77	0.65

To find the probability of self-reported mammography use, we use the law of total probability and the information regarding sensitivity and specificity. We let Y be the true mammography use status and Y^* be the self-reported mammography use status (a surrogate for Y). Then for a respondent with race x,

$$
\begin{aligned}
p_x &= P(Y^* = 1|x) \\
&= P(Y^* = 1|Y = 1, x)P(Y = 1|x) \\
&\qquad + P(Y^* = 1|Y = 0, x)P(Y = 0|x) \\
&= \eta_x \pi_x + (1 - \theta_x)(1 - \pi_x),
\end{aligned}
$$

where $\pi_x = P(\text{Respondent had mammogram in past two years})$, η is the sensitivity, and θ is the specificity. Thus, the self-reported data for respondent j is given by

$$
Y_j^*|x_j = x \sim \text{Bernoulli}(\eta_x \pi_x + (1 - \theta_x)(1 - \pi_x)).
$$

Note that Y is reserved for perfectly measured responses and Y^* is used to indicate surrogate responses subject to misclassification. Njai et al. (2011) perform frequentist linear regression to examine the prevalence of mammography in the past two years. We perform Bayesian logistic regression to model the binary outcome of mammography use in the past two years using the single binary covariate, race.

In this chapter, we investigate features of logistic regression in the presence of ordered differential response misclassification. In Section 2.2, we briefly present frequentist and Bayesian methods to adjust models with differential and non-differential response misclassification. In Section 2.3, we propose an adjustment for ordered differential response misclassification via the general (four parameter) beta as a prior for selected misclassification parameters. This Bayesian approach benefits from the conditional properties of the ordered misclassification data. In Section 2.4, we illustrate what is lost when using current "naive" methods to analyze misclassified data with such ordering. In Section 2.5, we perform a small scale simulation study to compare the performance of the ordered differential response misclassification adjustment to the differential response misclassification adjustment. We provide concluding remarks in Section 2.6.

2.2.1 Frequentist Models

Although this dissertation is written from a Bayesian perspective, we briefly present methods used in the frequentist literature to overcome the problems associated with response misclassification. We focus on logistic regression models.

In the frequentist literature, it is common to build models with misclassification present and account for the resulting bias after the fact by making adjustments to the estimates. This also requires an adjustment to the standard deviation of the estimator, sometimes relying on the delta method. Hausman et al. (1998) propose the use of a modified maximum likelihood estimator to correct for misclassification. They also utilize a semiparametric approach to combine the maximum rank correlation estimator with isotonic regression.

An alternate method is to make adjustments to the data before building the model. This requires setting fixed values for the misclassification parameters, sensitivity and specificity. In the differential response misclassification adjustment, fixed values for sensitivity and specificity are set for each covariate level. For the non-differential response misclassification adjustment, only one value of sensitivity and one value of specificity need to be specified for the model due to the assumption that misclassification is independent of the covariate level (Thomas, Stram, and Dwyer, 1993).

Magder and Hughes (1997) use a version of the EM algorithm and Neuhaus (1999) uses maximum likelihood methods after incorporating fixed values for the misclassification parameters, sensitivity and specificity. Setting fixed values for sensitivity and specificity can be problematic because these parameters are typically unknown. Lyles et al. (2011) use main/external and main/internal validation studies to determine values for sensitivity and specificity. We will elaborate on incorporating sensitivity and specificity information into the model under the Bayesian paradigm.

2.2.2 Bayesian Models

Distinguishing features of a Bayesian analysis include the act of conditioning on the data and quantifying uncertainty about each unknown parameter with a prior distribution. A beneficial aspect of using Bayesian methods is the ability to incorporate historical or expert knowledge into the current analysis through these prior distributions. Prescott and Garthwaite (2002) consider a two-stage approach to analyze case-control studies based on a mis-measured binary covariate in which correct values are known for a subset of the data. Having high quality prior knowledge or expert opinion is important in misclassification problems due to the unknown sensitivity and specificity parameters (Gustafson and Greenland, 2014).

In any practical setting, the sum of specificity and sensitivity is greater than one; however, it is common to treat these parameters' priors as independent. This is reasonable because any test with $\eta + \theta < 1$ is impractical. Assuming independence of the misclassification parameters makes the joint prior easy to construct as it is simply the product of marginal priors.

For Bayesian misclassification problems, independent beta priors are typically used for sensitivity and specificity. It is common to interpret the sum of the beta distribution's shape parameters in terms of a prior equivalent sample size (PESS) of a binomial experiment (Morita et al., 2008). The beta distribution's first shape parameter is interpreted as the number of successes in such an experiment.

In the differential response misclassification setting, prior distributions are given for sensitivity and specificity at each covariate level. As in the frequentist case, adjusting for non-differential misclassification is simpler than the modifications needed to accommodate differential misclassification. In the non-differential response misclassification setting, only two priors for misclassification, one for sensitivity and one for specificity, need to be specified across all covariate levels.

2.3 Ordered Misclassification Adjustments

As we have seen, there are frequentist and Bayesian methods that adjust for response misclassification. In this section, we introduce adjustments for *ordered differential response misclassification*, that is, response misclassification in which covariate values impose an ordering on one or both misclassification parameters. The ordering may hold with certainty, or with probability less than one. We examine both possibilities, but choose to perform a Bayesian analysis of the mammography example assuming order with probability one.

2.3.1 Adjustment for Ordered Misclassification with Probability One

To make use of additional information regarding the order of the response misclassification rates, we propose utilizing a conditional prior structure. Specifically, we suggest the family of four parameter beta distributions, also referred to as general betas. The beta distribution on the interval (u, v) has probability density function

$$\text{Beta}_{[u,v]}(a,b) \equiv \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)(v-u)^{a+b-1}}(x-u)^{a-1}(v-x)^{b-1},\ u \le x \le v,$$

where $a > 0, b > 0$. We denote this distribution generically by $\text{Beta}_{[u,v]}(a,b)$. Note that if $Z \sim \text{Beta}_{[0,1]}(a,b)$, then $X = u + Z(v-u) \sim \text{Beta}_{[u,v]}(a,b)$. For more information regarding the general beta, see Appendix A.1.

2.3.1.1 *Two dimensional ordered misclassification adjustment.* Suppose we have an ordered differential response misclassification problem. Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$. Then we construct a joint prior distribution on the sensitivities and specificities as $\pi(\boldsymbol{\eta}, \boldsymbol{\theta})$. For simplicity, consider the two dimensional case where $\eta_1 < \eta_2$ and $\theta_1 < \theta_2$. As in Section 2.2.2, we assume the two prior distributions are independent, which we denote by

$$\pi(\eta_1, \eta_2) \perp \pi(\theta_1, \theta_2).$$

Here, however, we propose the joint prior

$$\pi(\boldsymbol{\eta}, \boldsymbol{\theta}) = \pi(\eta_2|\eta_1)\pi(\eta_1)\pi(\theta_2|\theta_1)\pi(\theta_1),$$

where

$$\begin{aligned}\pi(\eta_1) &= \text{Beta}(a_{\eta_1}, b_{\eta_1}),\\ \pi(\eta_2|\eta_1) &= \text{Beta}_{[\eta_1,1]}(a_{\eta_2}, b_{\eta_2}),\\ \pi(\theta_1) &= \text{Beta}(a_{\theta_1}, b_{\theta_2}),\end{aligned}$$

and

$$\pi(\theta_2|\theta_1) = \text{Beta}_{[\theta_1,1]}(a_{\theta_1}, b_{\theta_2}).$$

The hyperparameters $(a_{\eta_1}, \ldots, b_{\theta_2})$ are chosen to align the distributions with the prior knowledge of the sensitivities and specificities. We can select a beta distribution by specifying a mode, percentile, and bounds for its support.[2] Suppose from historical or expert information we have reason to believe that $\theta_1 = 60\%$, $\theta_2 = 75\%$, and the fifth percentile of the distribution of θ_2 is 70%. We use this information along with the conditional relationship among misclassification parameters and a prior effective sample size of 100 to arrive at

$$\theta_1 \sim \text{Beta}(60, 40), \quad \theta_2|\theta_1 \sim \text{Beta}_{[\theta_1,1]}(39, 61).$$

Doing so aligns the distributions with prior information about θ_1 and θ_2. It also ensures that for every value of $\pi(\theta_1)$, the corresponding value of $\pi(\theta_2|\theta_1)$ is larger. This example is shown in Figure 2.1 and the adjustment is utilized in an analysis in Section 2.4.4.

The strict orderings implied by this joint distribution constitute a very stringent assumption. If this assumption is incorrect, no amount of data will correct the imposed ordering. In Section 2.3.2, we relax this condition. As in any Bayesian analysis, the methods of this section are highly dependent on the quality of historical data or expert opinion obtained.

2.3.1.2 *Ordered misclassification adjustment for three or more dimensions.* Suppose we have an ordered differential response misclassification problem and we con-

[2] Care should be taken here since specifying a mode and a percentile need not uniquely identify a beta distribution.

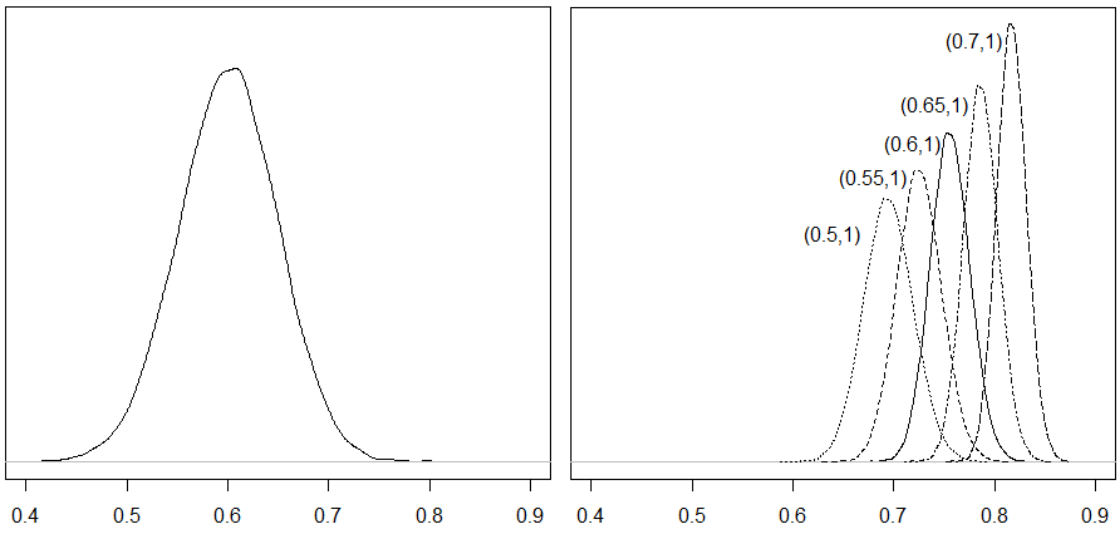

Figure 2.1: Beta(60,40) Distribution with Support (0,1) (left). Beta(39,61) distributions with support indicated in parentheses (right). In each case, the left endpoint of the support is the value of (θ_1) on which the prior for (θ_2) is conditioned.

struct a joint prior distribution on the sensitivities and specificities as $\pi(\boldsymbol{\eta}, \boldsymbol{\theta})$ for $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_p)$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. Consider the case where $\eta_1 < \eta_2 < \cdots < \eta_p$ and $\theta_1 < \theta_2 < \cdots < \theta_p$. We assume

$$\pi(\boldsymbol{\eta}) \perp \pi(\boldsymbol{\theta})$$

as before, but now the joint prior is given by

$$\pi(\boldsymbol{\eta}, \boldsymbol{\theta}) = \pi(\eta_p|\eta_{p-1}, \ldots, \eta_1) \cdots \pi(\eta_2|\eta_1)\pi(\eta_1)\pi(\theta_p|\theta_{p-1}, \ldots, \theta_1) \cdots \pi(\theta_2|\theta_1)\pi(\theta_1),$$

where

$$\begin{aligned}
\pi(\eta_1) &= \text{Beta}(a_{\eta_1}, b_{\eta_1}), \\
\pi(\eta_2|\eta_1) &= \text{Beta}_{[\eta_1,1]}(a_{\eta_2}, b_{\eta_2}), \\
\pi(\eta_k|\eta_{k-1}, \ldots, \eta_1) &= \text{Beta}_{[\eta_{k-1},1]}(a_{\eta_k}, b_{\eta_k}), k = 3, \ldots, p, \\
\pi(\theta_1) &= \text{Beta}(a_{\theta_1}, b_{\theta_1}), \\
\pi(\theta_2|\theta_1) &= \text{Beta}_{[\theta_1,1]}(a_{\theta_2}, b_{\theta_2}),
\end{aligned}$$

and

$$\pi(\theta_k|\theta_{k-1}, \ldots, 1) = \text{Beta}_{[\theta_{k-1},1]}(a_{\theta_k}, b_{\theta_k}), k = 3, \ldots, p.$$

Again, the hyperparameters are chosen to reflect prior information of the sensitivities and specificities.

2.3.2 Stochastically Ordered Misclassification Adjustment

We briefly discuss possible adjustments for stochastically ordered misclassification rates. The following papers propose methods to incorporate stochastic orderings among parameters in models. None of the papers discuss misclassification of any kind; however, stochastically ordered misclassification could be an area of application for these methods.

Madi et al. (2000) form hierarchical priors and incorporate order restrictions in the hyperparameter stage. Dunson and Peddada (2008) use a class of restricted dependent Dirichlet process priors. The priors "have full support in the space of stochastically ordered distributions, and can be used for collections of unknown mixture distributions to obtain a flexible class of restricted dependent Dirichlet process prior mixture models." Additionally, Evans et al. (1997) perform analyses using a Dirichlet prior on cell probabilities and proceed as if they are independent.

Another possible way to incorporate information regarding stochastic ordering among misclassification parameters is to make the shape parameters of one of the priors dependent on the shape parameters of the other prior. Doing so maintains the full support of the distribution, (0,1) for a beta distribution, but allows the prior to be centered over the desired value, which depends on the other misclassification parameter.

For example, suppose we obtain information regarding θ_2 conditional on information regarding θ_1. The shape parameters of the underlying beta distribution for θ_2 could be made to depend on θ_1. For example, we might require that

$$\text{mode}(\theta_2|\theta_1) = \alpha + \beta\theta_1$$

and add a percentile requirement, also dependent on θ_1. Or we might have

$$E(\theta_2|\theta_1) = \alpha + \beta\theta_1$$

and, for $\gamma > 0$,

$$V(\theta_2|\theta_1) = \gamma V(\theta_1).$$

The assumption of stochastically ordered misclassification is less stringent than the assumption of order with probability one. Adjusting for stochastically ordered misclassification as described above does not guarantee the order of the misclassification rates will be preserved; this is appropriate when order with probability one cannot be assumed. As before, this method is dependent on the quality of historical data or expert opinion. The stochastically ordered adjustment may be an appropriate topic for future exploration and development.

2.4 Analysis of Mammography Data

2.4.1 Naive Analysis

Recall the mammography use example in Section 2.1. We begin by developing a conditional means prior (Bedrick, Christensen, and Johnson, 1996) for a logistic regression model while assuming no misclassification. For more information on conditional means priors, see Appendix A.2. Without misclassification, we have

$$Y_j|\pi_j \sim \text{ Bernoulli}(\pi_j),$$

where

$$\text{logit}(\pi_j) = \beta_0 + \beta_1 x_{1j},$$

and x_1 is race (0 = African American, 1 = Caucasian). The configuration matrix for the CMP elicitation is

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{x}}_A' \\ \tilde{\mathbf{x}}_C' \end{bmatrix},$$

where the covariate configurations are

$$\tilde{\mathbf{x}}_A = \begin{bmatrix} \text{Baseline} \\ \text{African American} \end{bmatrix}, \tilde{\mathbf{x}}_C = \begin{bmatrix} \text{Baseline} \\ \text{Caucasian} \end{bmatrix}.$$

Suppose we have a single expert from whom we elicit beta priors for the two $\tilde{\mathbf{x}}_k$ configuration vectors above. For $\tilde{\mathbf{x}}_A$, knowing only that race is African American,

suppose the expert believes the most likely value for the probability of having a mammogram in the past two years is 0.59 and and thinks it very unlikely this probability is less than 0.44. If we interpret the latter as the 5th percentile of a beta density, then the resulting prior for $\tilde{\pi}_A = P(Y = 1|\tilde{x}_A)$ is a Beta$(19, 13)$ distribution. In terms of prior equivalent sample size, this represents 19 prior successes out of a total number of 32 binomial trials. This produces an expected value of 59.4%.

Similarly, for $\tilde{\mathbf{x}}_C$, knowing only that race is Caucasian, suppose the expert believes the most likely value for the probability of having a mammogram in the past two years is 0.65 and thinks it unlikely that this probability is lower than 0.50. Then the resulting beta prior for $\tilde{\pi}_C = P(Y = 1|\tilde{x}_C)$is a Beta$(21, 12)$ distribution. In terms of prior equivalent sample size, this represents 21 prior successes out of a total number of 33 binomial trials. This produces an expected value of 63.6%. See Figure 2.2 for an illustration of the conditional means priors produced from this information.

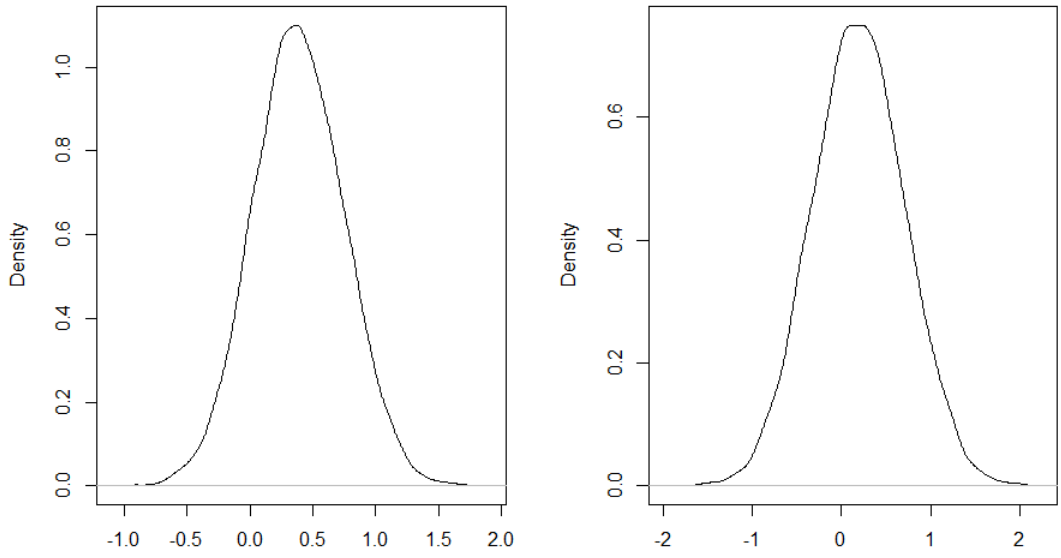

Figure 2.2: Conditional Means Priors Induced on β_0 (left) and β_1 (right) for Mammography Example.

We perform 10,000 burn-ins and 50,000 posterior samples to produce the results in Table 2.2 and Figure 2.3. The posterior estimates for the probabilities of mammography use are equal to the self-reported mammography use probabilities. This occurred because we naively assumed that the self-reported mammography use

data were recorded without error. Additionally, even though we incorporated expert opinion into the model via conditional means priors, the data overwhelmed the priors. This is not surprising because the sample size is very large ($n_A > 30,000$ and $n_C > 100,000$) and the prior equivalent sample sizes for the CMP for β_0 and β_1 are only in the thirties (we examine smaller sample sizes in Section 2.5). Note that the posterior means differ considerably from the adjusted probability estimates from Njai et al. (2011). This is because we have ignored misclassification.

Also of interest in this logistic regression model is the odds ratio of having a mammogram in the past two years for Caucasian women compared to African American women; that is

$$OR_{CA} = \exp(\beta_1). \tag{2.1}$$

Table 2.2: Posterior Summary Statistics for Naive Analysis (The adjusted estimates in the second column are from Njai et al., 2011.).

Parameter	Adjusted Estimate	Mean	SD	2.50%	Median	97.50%
β_0	–	1.258	0.012	1.233	1.258	1.282
$\beta 1$	–	−0.050	0.014	−0.078	−0.050	−0.023
OR_{CA}	–	0.951	0.014	0.925	0.951	0.978
π_A	0.59	0.779	0.002	0.774	0.779	0.783
π_C	0.65	0.770	0.001	0.767	0.770	0.772

One of the benefits of a Bayesian analysis is the ease of assessing transformations of parameters, such as (2.1). Using the posterior mean in Table 2.2, we estimate the odds ratio as 0.951. Thus, the estimated odds of Caucasian women having a mammogram in the past two years is 0.95 times less likely than those of African American women.

2.4.2 Non-Differential Response Misclassification Analysis

We conduct this analysis with the same conditional means prior distributions for the regression parameters as in Section 2.4.1, but now acknowledge the presence of misclassification in the model. In this section, we perform a correction for non-differential response misclassification. As was briefly discussed in Section 2.2, we

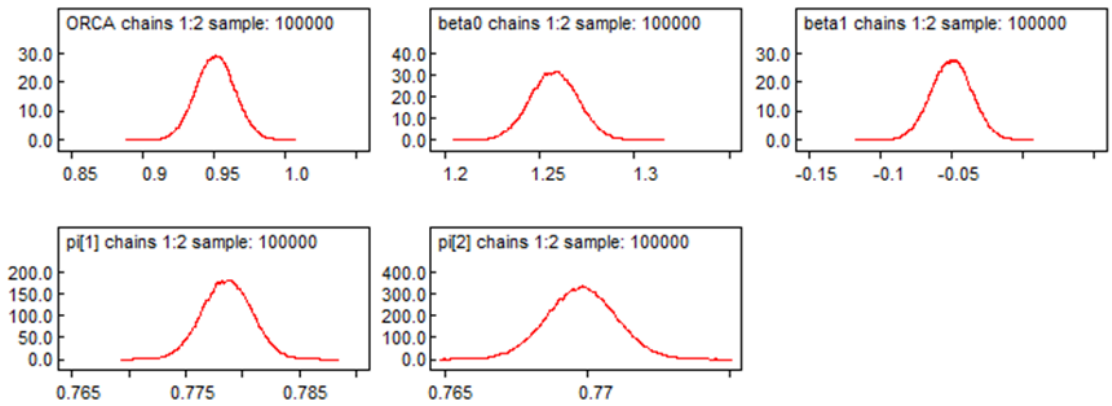

Figure 2.3: Posterior Densities from Naive Analysis.

adjust for this type of misclassification by specifying one prior for the sensitivity and one prior for the specificity. These priors remain unchanged regardless of covariate level. This is a reasonable restriction for the sensitivity because information suggests both races have sensitivity of 0.97. Using a prior equivalent sample size of 100 yields the prior $\pi(\eta) \sim \text{Beta}(97, 3)$. Because we are illustrating the results of assuming non-differential misclassification in the model, we somewhat arbitrarily set the common prior on specificity to be $\pi(\theta) \sim \text{Beta}(55, 45)$, which has a mode between the specificity estimates due to Njai et al. (2011), 0.49 and 0.62.

Mild autocorrelation in preliminary MCMC runs indicated a need to thin the chains. We used a burn-in of 100,000 values, retaining every 10th iteration. We followed this with 500,000 iterations, again retaining every 10th value. No other convergence issues were observed. The posterior densities are sufficiently smooth and are provided in Figure 2.4.

For this non-differential response misclassification adjusted model, the estimate for OR_{CA} is 0.93. Thus, the estimated odds of Caucasian women having a mammogram in the past two years is 0.93 times less likely than those of African American women.

Although the resulting posterior standard deviations are larger than those of the naive analysis, this method is not ideal, as evidenced by the poor posterior estimates of the mammography use probabilities. This is because the non-differential misclas-

Table 2.3: Posterior Summary Statistics for Non-Differential Misclassification Analysis.

Parameter	Adjusted Estimate	Mean	SD	2.50%	Median	97.50%
β_0	–	0.521	0.148	0.225	0.522	0.811
β_1	–	−0.074	0.022	−0.119	−0.074	−0.033
η	0.97	0.970	0.016	0.931	0.972	0.993
OR_{CA}	–	0.929	0.020	0.888	0.929	0.968
π_A	0.59	0.627	0.035	0.556	0.628	0.692
π_C	0.65	0.610	0.035	0.537	0.610	0.676
θ	0.55	0.544	0.043	0.462	0.544	0.630

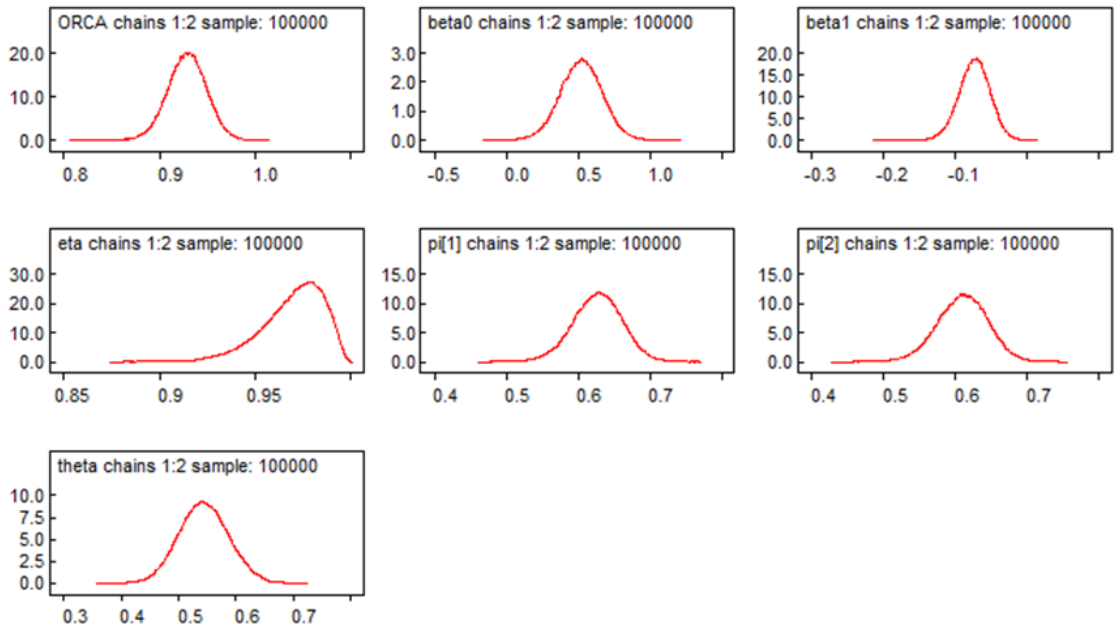

Figure 2.4: Posterior Densities from Non-Differential Misclassification Analysis.

sification correction is too simplistic for this problem in that it only incorporates one prior for specificity. Doing so cannot adequately adjust estimates.

2.4.3 Differential Response Misclassification Analysis

We continue to use the conditional means priors for the regression parameters as in the two previous sections; however, we now acknowledge the presence of differential response misclassification. For the survey response problem, the sensitivity was estimated by Njai et al. (2011) to be 0.97. Again using a prior equivalent sample size of 100, we specify the priors as $\pi(\eta_A) = \pi(\eta_C) = \text{Beta}(97, 3)$. The specificities for African Americans and Caucasians are 0.49 and 0.62, respectively, again as estimated

by Njai et al. (2011). We construct two priors, each with a prior equivalent sample size of 100. These priors are $\pi(\theta_A) = \text{Beta}(49, 51)$ and $\pi(\theta_C) = \text{Beta}(62, 38)$.

We used a burn-in of 100,000 values, retaining every 10th iteration. We followed this with 500,000 iterations, again retaining every 10th value. The results are provided in Table 2.4 and the posterior densities are plotted in Figure 2.5.

Table 2.4: Posterior Summary Statistics for Differential Misclassification Analysis.

Parameter	Adjusted Estimate	Mean	SD	2.50%	Median	97.50%
β_0	–	0.339	0.185	−0.042	0.344	0.691
β_1	–	0.303	0.236	−0.149	0.298	0.781
η_A	0.97	0.969	0.017	0.928	0.972	0.993
η_C	0.97	0.971	0.016	0.932	0.974	0.994
OR_{CA}	–	1.392	0.338	0.862	1.347	2.183
π_A	0.59	0.583	0.045	0.490	0.585	0.666
π_C	0.65	0.654	0.033	0.588	0.655	0.718
θ_A	0.49	0.491	0.045	0.406	0.490	0.580
θ_C	0.62	0.614	0.047	0.522	0.614	0.705

For this differential response misclassification adjusted model, the estimate for OR_{CA} is 1.39. Thus, the estimated odds of Caucasian women having a mammogram in the past two years is 1.39 times greater than those of African American women.

The parameter estimates are all near the adjusted estimates and the standard deviations are larger than those obtained via the naive model. In fact, the estimates for the probabilities of mammography use by race are within a few one-thousandths of the adjusted estimates. This method adequately corrects the misclassification; however, it does not use or benefit from the *a priori* knowledge of the order amongst the misclassification parameters.

Since this method does not explicitly take order into account, there are times that it may be violated. For this example, violating the order means $P(\theta_A < \theta_C) < 1$. To examine this more closely, we conduct a simulation with 10,000 replications to determine the probability of order violation when using the differential response misclassification adjustment. Table 2.5 displays the probability of order violations as

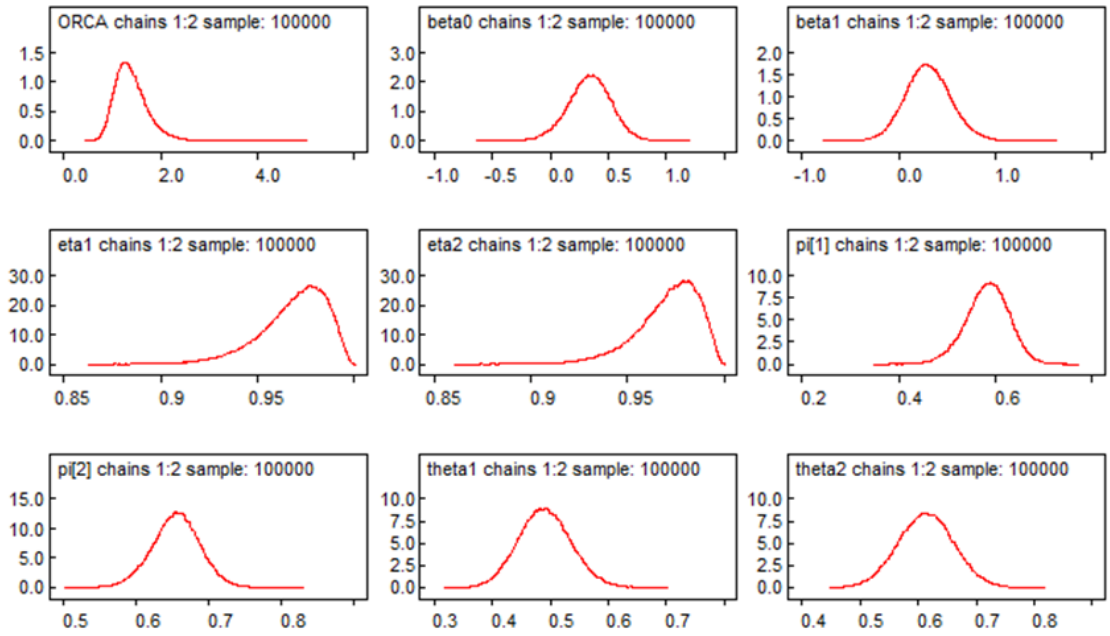

Figure 2.5: Posterior Densities from Differential Misclassification Analysis.

$\delta \equiv \theta_C - \theta_A$ increases while using priors based on an equivalent sample size of 100. As expected, the probability of violation decreases as δ increases.

Table 2.5: Extent of Order Violations using Independent Prior Distributions.

δ	$\pi(\theta_A)$	$\pi(\theta_C)$	P(Order Violated)
0.05	Beta(50, 50)	Beta(55, 45)	0.2416
0.10	Beta(50, 50)	Beta(60, 40)	0.0777
0.15	Beta(50, 50)	Beta(65, 35)	0.0160
0.20	Beta(50, 50)	Beta(70, 30)	0.0023

For this example, $\theta_A = 0.49$, $\theta_C = 0.62$, and the resulting probability of order violation is 0.0341.

2.4.4 Ordered Differential Response Misclassification Analysis

We continue to use the conditional means priors for the regression parameters as in the three previous sections; however, we make an adjustment for the ordered differential response misclassification. The estimated sensitivity is the same for African Americans and Caucasians, so we specify $\pi(\eta_A) = \pi(\eta_C) \sim \text{Beta}(97, 3)$, as before. Since the specificities are ordered, we specify the marginal prior distribution for

African American specificity as $\pi(\theta_A) \sim \text{Beta}(49, 51)$ and condition the values of the prior for Caucasian specificity on the values of the prior for African American specificity. Provided a mode of 0.62 and a ninety fifth percentile of 0.82, we solve for the shape parameters and find $\pi(\theta_C|\theta_A) \sim \text{Beta}_{[\theta_A,1]}(24, 76)$. Note that it is recommended to specify the fifth percentile if the mode is larger than 0.50, but complications arise due to the shifted support. We have incorporated the known order of the response misclassification into the model with probability one.

The likelihood is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\theta}) = \prod_{j=1}^{n} [\pi_j \eta_j + (1 - \pi_j)(1 - \theta_j)]^{y_j^*} [\pi_j (1 - \eta_j)(1 - \pi_j)\theta_j]^{1-y_j^*}.$$

The joint prior for the misclassification parameters is

$$\pi(\boldsymbol{\eta}, \boldsymbol{\theta}) = \text{Beta}(\eta_A|97, 3)\text{Beta}(\eta_C|97, 3)\text{Beta}(\theta_A|49, 51)\text{Beta}_{[\theta_A,1]}(\theta_C|24, 76)$$

and the prior induced on the regression parameters is

$$\pi(\boldsymbol{\beta}) = \tilde{\mathbf{X}}^{-1}(\tilde{\boldsymbol{\pi}}),$$

where

$$\tilde{\boldsymbol{\pi}} = \text{Beta}(\tilde{\pi}_A|19, 13)\text{Beta}(\tilde{\pi}_C|21, 12).$$

Thus, the posterior is

$$\pi(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{x}, \mathbf{y}^*) \propto L(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\theta})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\eta}, \boldsymbol{\theta}).$$

This model is diagramed in Figure 2.6.

We perform a burn-in of 100,000 values, retaining every 10th iteration. We follow this with 500,000 iterations, again retaining every 10th value. The results are in Table 2.6 and the posterior densities are in Figure 2.7. The parameter estimates are all near the adjusted estimates and the standard deviations are larger than those obtained from the naive model.

For this ordered differential misclassification adjusted model, the estimate for OR_{CA} is 1.38. Thus, the estimated odds of Caucasian women having a mammogram in the past two years is 1.38 times greater than those of African American women.

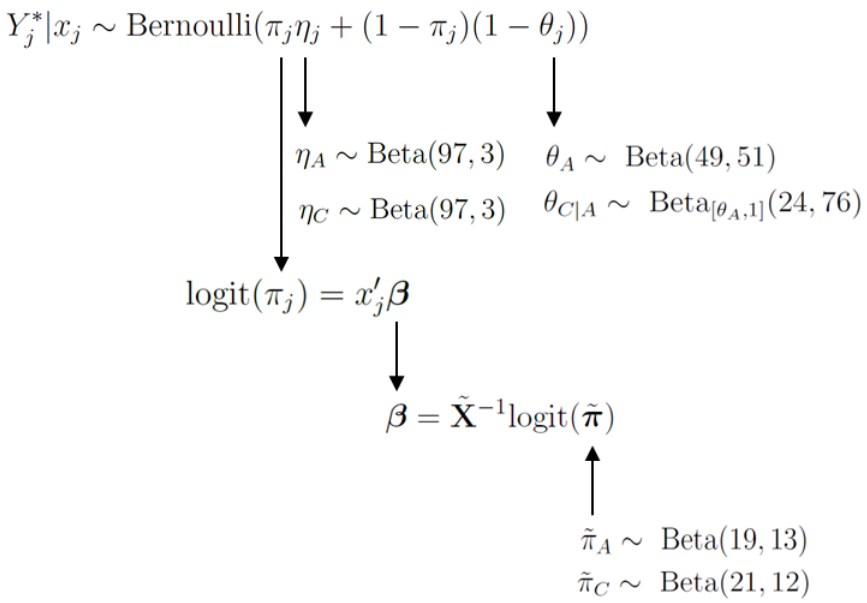

Figure 2.6: Ordered Differential Response Misclassification Model Diagram.

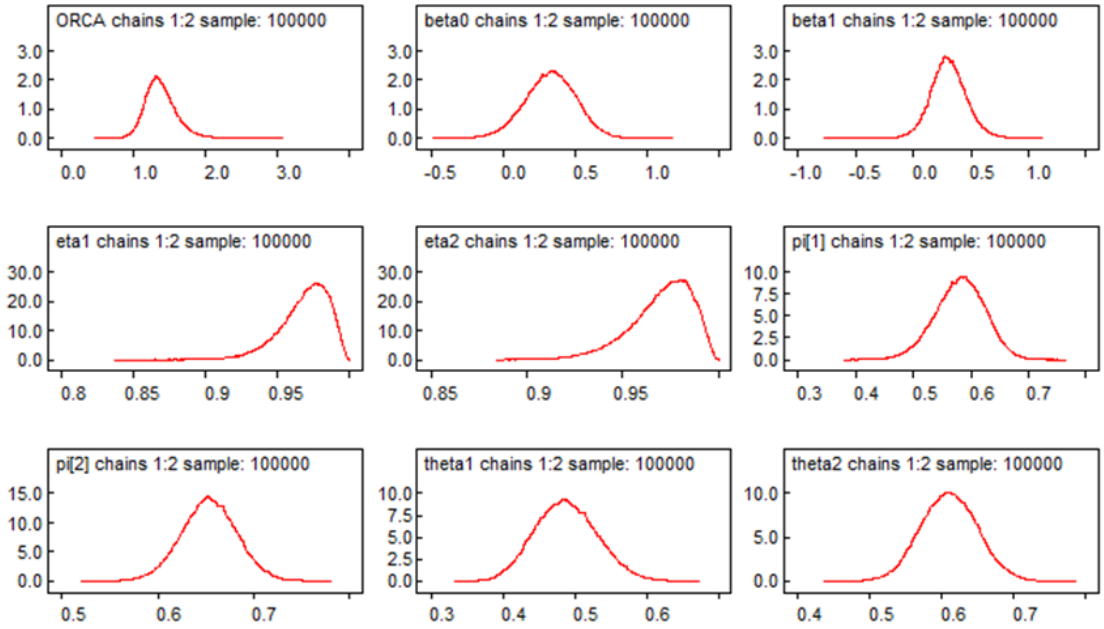

Figure 2.7: Posterior Densities from Ordered Differential Misclassification Analysis.

2.4.5 Comparison of Results Under Different Assumptions of Misclassification

The naive approach yields poor parameter estimates. The non-differential response misclassification adjustment allows only one prior for the specificity, which is not suitable for the problem. The standard deviations for the non-differential re-

Table 2.6: Posterior Summary Statistics for Ordered Differential Misclassification Analysis.

Parameter	Adjusted Estimate	Mean	SD	2.50%	Median	97.50%
β_0	–	0.330	0.178	−0.029	0.334	0.671
β_1	–	0.308	0.156	0.007	0.303	0.630
η_A	0.97	0.969	0.017	0.929	0.972	0.993
η_C	0.97	0.970	0.016	0.932	0.973	0.994
OR_{CA}	–	1.377	0.219	1.007	1.354	1.877
π_A	0.59	0.581	0.043	0.493	0.583	0.662
π_C	0.65	0.654	0.029	0.597	0.654	0.710
θ_A	0.49	0.488	0.043	0.408	0.487	0.575
θ_C	0.62	0.610	0.039	0.535	0.610	0.687

sponse misclassification adjusted model are larger than the naive model's, but the parameter estimates from this adjustment are still far from the adjusted estimates.

The odds ratios under these two models are both less that one, which indicate that African American women are more likely than Caucasian women to have had a mammogram in the past two years. The differential and ordered adjusted models yield odds ratios that are greater than one, which indicate that Caucasian women are more likely than African American women to have had a mammogram in the past two years. These conflicting inferences are a common problem in models with misclassified data. This illustrates one of the reasons it is important to adjust for misclassification if it is present.

The differential response misclassification adjustment produces parameter estimates that reflect those found in Njai et al. (2011) with larger standard deviations than the naive model, but fails to incorporate the known ordering in the data roughly 3% of the time in this example. Using the ordered differential response misclassification adjustment produces parameter estimates that reflect those found in Njai et al. (2011). Additionally, it does not require substantial additional work and is preferred because it ensures that the order in the misclassification rates is preserved. Interestingly, the standard deviations observed are smaller than those produced by the differential response misclassification adjustment, which may be another advan-

tage of the method. In Section 2.5, we conduct a small scale simulation to compare the properties of these two adjustments more closely.

2.5 Simulation

In the previous section, we illustrated shortcomings of analyses of data with ordered differential response misclassification under a naive assumption, a non-differential misclassification assumption, and a differential misclassification assumption. The naive and non-differential misclassification methods were shown to be inadequate and will not be included in the simulation. The differential misclassification correction adequately adjusted the estimates in the previous section, so we compare simulation results of this method to the ordered differential response misclassification adjustment's.

Both methods performed similarly in the mammogram use example, but the ordered differential response misclassification adjustment yielded smaller posterior standard deviations than the differential response misclassification adjustment. For this reason, we examine coverage and credible interval width of the slope parameter, β_1. In the mammography example, $\delta \equiv \theta_C - \theta_A = 0.13$ and $\delta_\pi \equiv \pi_C - \pi_A = 0.06$. The range of plausible values of δ is limited because it is desirable for both η and θ to be relatively high. We choose a high and low value for δ and δ_π and consider sample sizes of 500 and 1000. The design points are provided in Table 2.7 and the corresponding priors are in Table 2.8. Note that $\pi(\theta_C)$ in Table 2.8 is for the differential response misclassification adjustment and $\pi(\theta_C|\theta_A)$ is for the ordered differential response misclassification adjustment.

Table 2.7: Design Points for Response Misclassification Simulation.

Sim.	δ	δ_π	θ_A	θ_C	$\eta_A = \eta_C$	β_0	β_1	π_A	π_C	p_A	p_C
1	0.05	0.05	0.70	0.75	0.97	0.00	0.20	0.50	0.55	0.64	0.65
2	0.05	0.20	0.70	0.75	0.97	0.00	0.85	0.50	0.70	0.64	0.75
3	0.20	0.05	0.70	0.90	0.97	0.00	0.20	0.50	0.55	0.64	0.58
4	0.20	0.20	0.70	0.90	0.97	0.00	0.85	0.50	0.70	0.64	0.71

To complete this simulation, we use R2WinBUGS to perform 200,000 burn-ins and 500,000 posterior samples with a thinning factor of 10. Using a thinning factor of 10 keeps only every 10th iteration, leaving 20,000 burn-ins and 50,000 posterior samples. Doing so eliminates autocorrelation and speeds convergence. We perform 100 replications for each of the 16 design points. The simulation results for β_1 are in Table 2.9.

Table 2.8: Shape Parameters of Beta Priors for Simulation. The support of $\pi(\theta_C|\theta_A)$ is $(\theta_A, 1)$; all other priors have support (0,1).

θ_A	θ_C	$\theta_C\|\theta_A$	$\eta_A = \eta_C$	π_A	π_C
$(35, 15)$	$(37.5, 12.5)$	$(9, 41)$	$(48.5, 1.5)$	$(25, 25)$	$(27.5, 22.5)$
$(35, 15)$	$(37.5, 12.5)$	$(9, 41)$	$(48.5, 1.5)$	$(25, 25)$	$(35, 15)$
$(35, 15)$	$(45, 5)$	$(33, 17)$	$(48.5, 1.5)$	$(25, 25)$	$(27.5, 22.5)$
$(35, 15)$	$(45, 5)$	$(33, 17)$	$(48.5, 1.5)$	$(25, 25)$	$(35, 15)$

The conditional means priors induced on the regression parameters are plotted in Figure 2.8 and Figure 2.9. To observe the updating from the likelihood, selected parameters' priors and posteriors for the design points used in Simulation 1 of Table 2.7 are plotted in Figure 2.10. The posteriors are less variable than the priors and the distributions are centered near true parameter values.

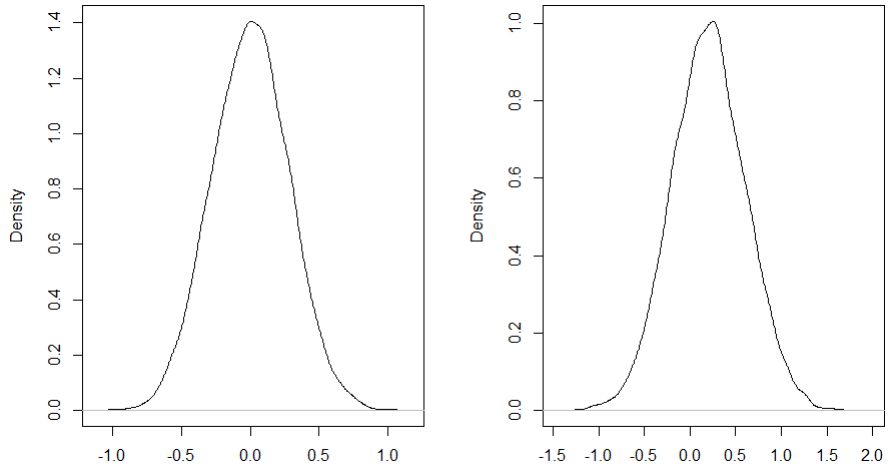

Figure 2.8: Priors Induced on β_0 (left) and β_1 (right) for Simulations 1 and 3.

After obtaining the 100 replications for each simulation design point in Table 2.7, we investigate the simulation's variability by examining posterior credible

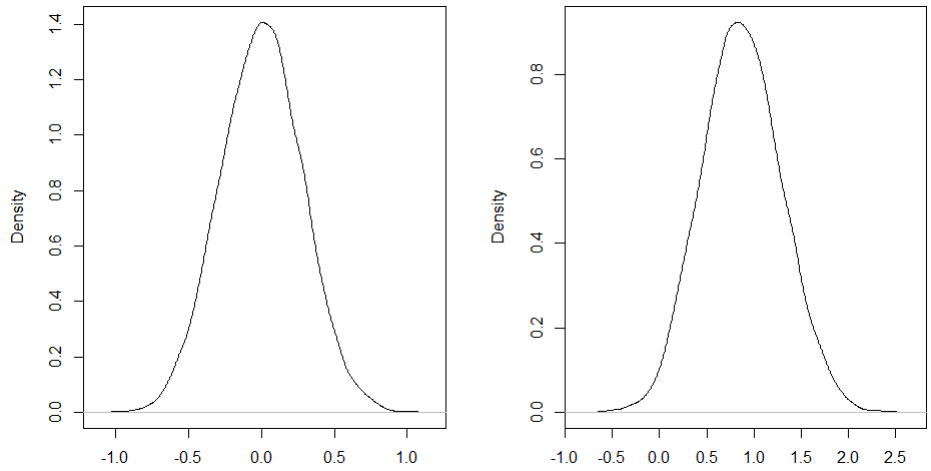

Figure 2.9: Priors Induced on β_0 (left) and β_1 (right) for Simulations 2 and 4.

intervals. Figures 2.11-2.14 display the simulation variability for β_1 at each design point. The line across the middle of each plot corresponds to the true parameter value for the simulation (Table 2.7) and the black circle is the median of the 100 posterior means. The two short lines above and below the middle line are the medians of the 100 95% credible interval upper and lower bounds, respectively. Additionally, the grey vertical bar represents $\pm$ 1 simulation standard deviation from the median of 100 posterior means and 95% credible interval bounds.

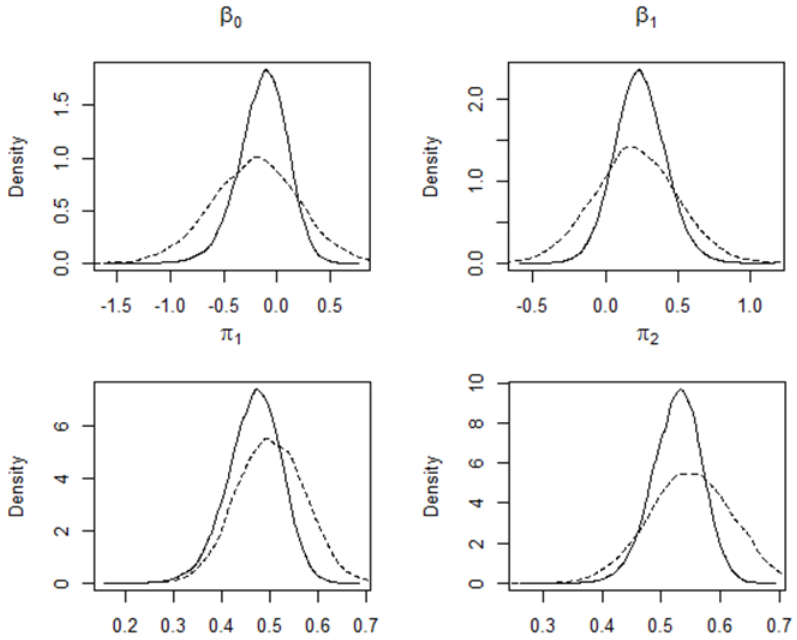

Figure 2.10: Prior (Dashed) and Posterior (Solid) Distributions for Simulation 1 (ordered adjustment, $n = 500$).

Table 2.9: Simulation Results for β_1. Mean (width). All simulations had 100% coverage.

Adjustment (Sample Size)	Simulation 1	Simulation 2	Simulation 3	Simulation 4
True β_1 Value	0.200	0.850	0.200	0.850
Ordered ($n = 500$)	0.214	0.836	0.197	0.814
	(0.691)	(0.798)	(0.743)	(0.813)
Differential ($n = 500$)	0.181	0.722	0.185	0.765
	(0.827)	(0.888)	(0.796)	(0.850)
Ordered ($n = 1000$)	0.210	0.853	0.203	0.839
	(0.611)	(0.724)	(0.691)	(0.758)
Differential ($n = 1000$)	0.183	0.719	0.189	0.780
	(0.794)	(0.857)	(0.762)	(0.810)

In each simulation, the coverage for the 95% credible interval is 100%. Additionally, the ordered differential response misclassification adjustment produces a narrower credible interval than those found using the differential response misclassification adjustment. Thus, the analysis benefits from the dependent prior structure by preserving order and yielding more precise estimates for β_1 than those from the independent prior structure.

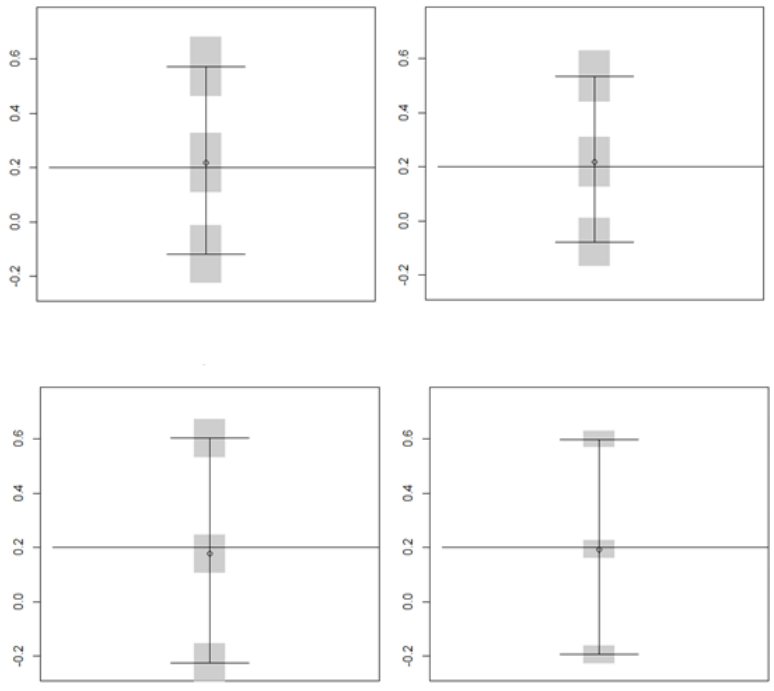

Figure 2.11: Simulation Summary for β_1 in Simulation 1. Row 1: Ordered adjustment, Row 2: Differential adjustment (left: $n = 500$, right: $n = 1000$).

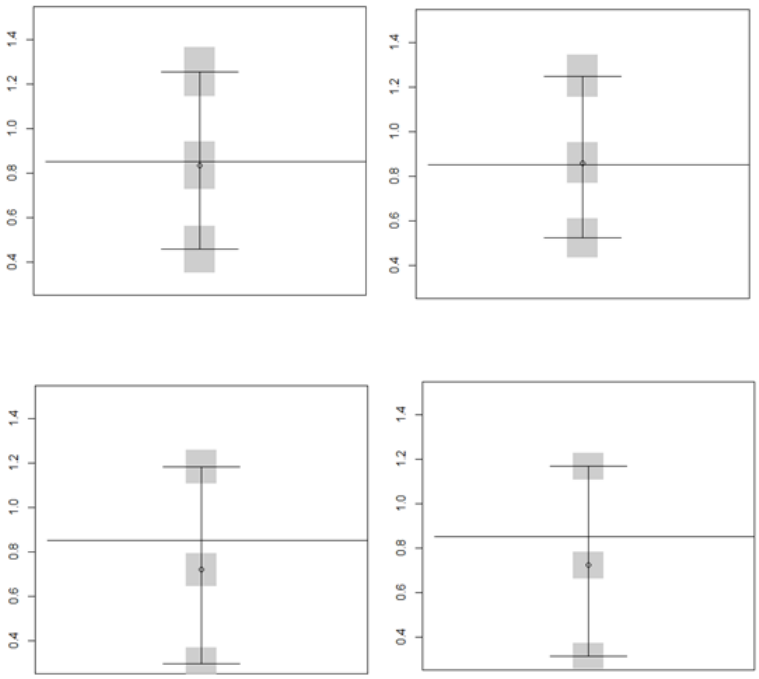

Figure 2.12: Simulation Summary for β_1 in Simulation 2. Row 1: Ordered adjustment, Row 2: Differential adjustment (left: $n = 500$, right: $n = 1000$).

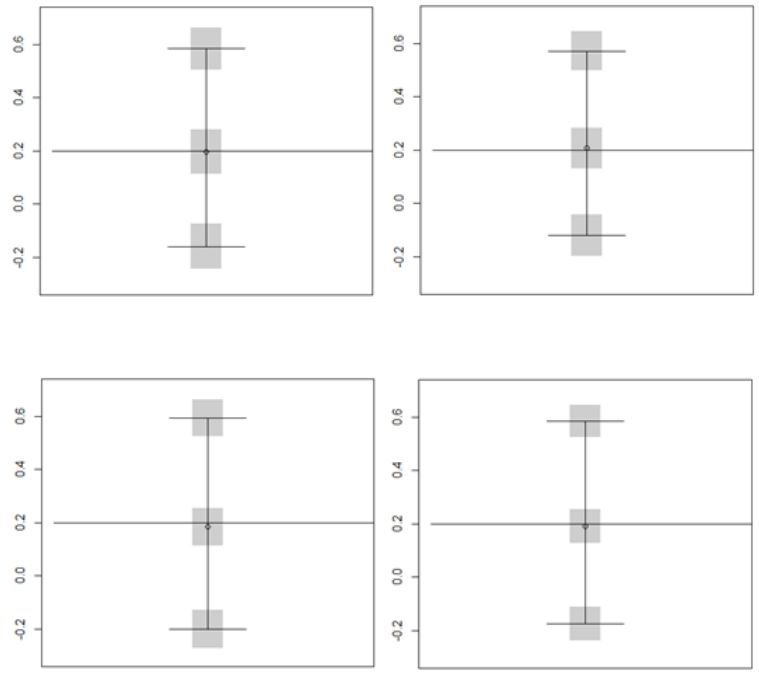

Figure 2.13: Simulation Summary for β_1 in Simulation 3. Row 1: Ordered adjustment, Row 2: Differential adjustment (left: $n = 500$, right: $n = 1000$).

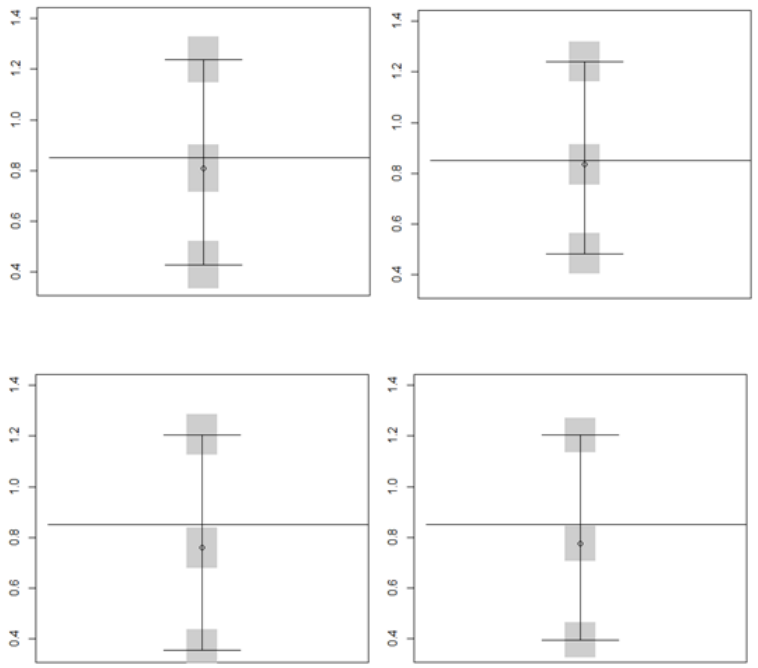

Figure 2.14: Simulation Summary for β_1 in Simulation 4. Row 1: Ordered adjustment, Row 2: Differential adjustment. (left: $n = 500$, right: $n = 1000$).

2.6 Concluding Remarks

In this chapter, we discussed ordered differential response misclassification, a problem that exists, but is currently ignored in the literature. We proposed a method to adjust the ordered differential response misclassification in a logistic regression model. This method incorporates information regarding sensitivity and specificity through a series of marginal and conditional prior distributions. We recognize that the proposed method is largely dependent on expert opinion and is only as good as the quality of information obtained during the prior elicitation process. We saw improvements in modeling mammography use by race after correcting for the ordered response misclassification. This method requires minimal additional work compared to the differential misclassification correction. Additionally, in small scale simulations, the estimates had adequate removal of bias, complete coverage, and narrower credible intervals compared to the results from the differential adjustment. For these reasons, the ordered differential misclassification adjustment is preferred over the differential adjustment when order with probability one may be assumed.

CHAPTER THREE

Weibull Proportional Hazards Regression with Ordered Covariate Misclassification

3.1 Introduction

Suppose we wish to study the potential impact of race/ethnicity on progression-free survival for patients with a certain type of cancer (CDC, 2014). We also want to consider the size of hospital as a potential predictor. To this end, suppose we obtain data from a source such as the Greater Bay Area Cancer Registry (Cancer Prevention Institute of California, 2015). Specifically, define

$$x = \begin{cases} 0 & \text{if the subject is White and Hispanic;} \\ 1 & \text{if the subject is White and non-Hispanic,} \end{cases}$$

and

$$z = \begin{cases} 0 & \text{if the hospital is large and public;} \\ 1 & \text{if the hospital is small and private.} \end{cases}$$

Unfortunately, race/ethnicity is often misclassified in such records, and differentially so with respect to hospital size. Gomez et al. (2003) and Gomez and Glaser (2006) show that large, public hospitals have more error in racial/ethnic classification than small, private hospitals. The misclassified surrogate is described by

$$x^* = \begin{cases} 0 & \text{if the subject is listed as White and Hispanic in the cancer registry;} \\ 1 & \text{if the subject is listed as White and non-Hispanic in the cancer registry.} \end{cases}$$

In this example of *differential misclassification*, the racial/ethnic classification is more prone to error in large hospitals. We are particularly interested in orderings imposed on the misclassification rates for categorical covariates, focusing primarily on binary covariates in this chapter. We refer to this as *ordered covariate misclassification*. The order may occur from a dependence on levels of another covariate, or alternatively, the dependence may occur between the sensitivity and specificity directly.

For the examples considered in this chapter, we assume that the rates of misclassification are ordered with probability one. To avoid potentially misleading inferences from models with misclassified data, we must provide an adjustment for the misclassification (Gustafson, 2003). To do this, we incorporate prior information regarding the misclassification parameters in a Bayesian model.

We illustrate the versatility of the ordered misclassification adjustment from the last chapter by first building a parametric Bayesian survival model with one misclassified binary covariate and then extend the model to include the addition of a perfectly recorded binary covariate. We incorporate prior information of misclassification using four-parameter beta distributions within a proportional hazards regression model. For the latter, we assume a Weibull baseline hazard (Klein and Moeschberger ,2005, Luo et al., 2012).

In Section 3.2, we provide an overview of parametric proportional hazard regression. In Section 3.3, we discuss complications of and adjustments for covariate misclassification, including ordered covariate misclassification. In Section 3.4 and Section 3.5, we provide examples which compare results obtained from models under different assumptions regarding the misclassified data. In Section 3.6 and Section 3.7, we conduct small-scale simulations to study the performance of our proposed correction. We make concluding remarks in Section 3.8.

3.2 Survival Analysis

3.2.1 Basic Functions

Suppose T is a non-negative random variable with CDF F and continuous PDF, f. We take f to be a distribution of lifetimes for subjects in a population and define the survival function as

$$S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(u)du.$$

It follows that

$$f(t) = -\frac{dS(t)}{dt}.$$

Another quantity of interest is the hazard function, the instantaneous rate of death:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

Since we have assumed that T is continuous,

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \ln[S(t)]}{dt}.$$

The hazard function may be preferred over the survival function because it more clearly depicts the failure pattern. Intuitively, as the hazard rate increases, the probability of failure increases. Similarly, the probability of failure decreases as the hazard rate decreases. Compare the survival and hazard functions in Figure 3.1 for an illustration.

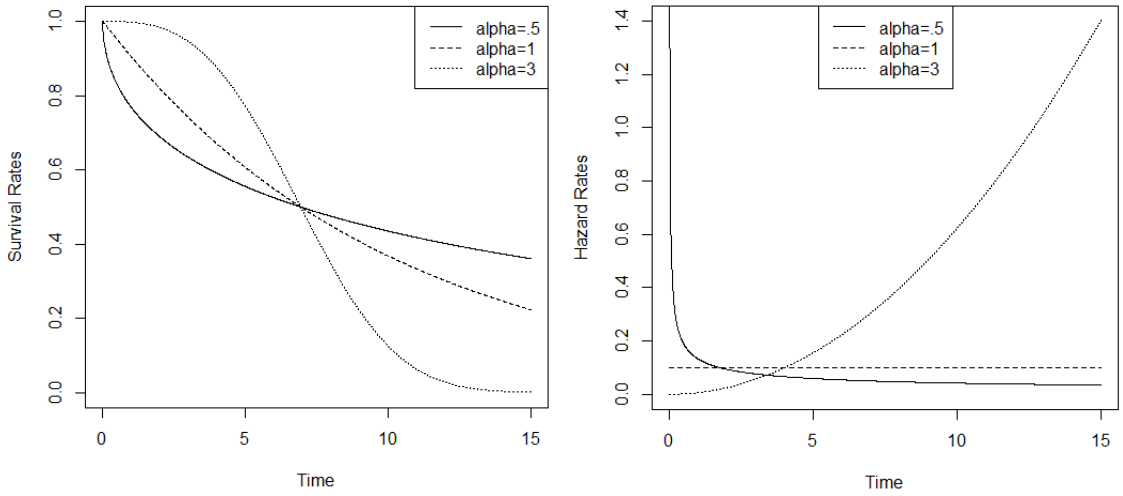

Figure 3.1: Weibull Survival Functions (left) and Hazard Functions (right), a Recreation from Page 29 of Klein and Moeschberger (2005).

The Weibull distribution is often used in survival analysis. A random variable, T, follows a Weibull distribution if

$$f(t|\alpha, \lambda) = \alpha \lambda t^{\alpha - 1} e^{-\lambda t^{\alpha}}, \alpha > 0, \lambda > 0.$$

The survival function for the Weibull is

$$\begin{aligned} S(t) &= P(T > t) \\ &= \int_t^\infty \alpha\lambda u^{\alpha-1} \exp(-\lambda u^\alpha) du \\ &= \exp(-\lambda t^\alpha) \end{aligned}$$

and the hazard function is

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{\alpha\lambda t^{\alpha-1} \exp(-\lambda t^\alpha)}{\exp(-\lambda t^\alpha)} \\ &= \alpha\lambda t^{\alpha-1}. \end{aligned}$$

The Weibull distribution can mimic characteristic shapes of many different distributions and is flexible enough to model a variety of data sets. The Weibull distribution can also model hazard functions that are increasing, decreasing or constant, as shown in Figure 3.1.

3.2.2 Proportional Hazards Model

Regression models can be used to relate survival functions or hazard functions to covariates. A widely used approach is to fit a *proportional hazards model*, defined as

$$h(t|\mathbf{x}) = h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta}),$$

where $h_0(t)$ is a *baseline hazard function*, $\mathbf{x}$ is a $p \times 1$ vector of covariates, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients. For an introduction to such models see, for example, Klein and Moeshberger (2005). The baseline hazard function is chosen so as to render the hazard positive. For any such function, the ratio of hazards lends the model its name: Suppose two subjects have covariate values $\mathbf{x}_a$ and $\mathbf{x}_b$. Then the hazard ratio is

$$\begin{aligned} \frac{h(t)|\mathbf{x}_a}{h(t)|\mathbf{x}_b} &= \frac{h_0(t) \exp(\mathbf{x}_a'\boldsymbol{\beta})}{h_0(t) \exp(\mathbf{x}_b'\boldsymbol{\beta})} \\ &= \exp(\mathbf{x}_a'\boldsymbol{\beta} - \mathbf{x}_b'\boldsymbol{\beta}) \\ &= \exp[(\mathbf{x}_a - \mathbf{x}_b)']\boldsymbol{\beta}. \end{aligned}$$

Thus, the hazard rates are proportional. This is a strong assumption and may not always be reasonable.

If a parametric distribution for $h_0(t)$ is not specified, then partial likelihood methods can be used to construct a *Cox proportional hazards model*. In our development, we specify a Weibull baseline hazard function, using Bayesian methods to fit the model. That is,

$$h_0(t) = \alpha\lambda t^{\alpha-1}$$

and

$$h(t|\mathbf{x}) = \alpha t^{\alpha-1}\exp(\mathbf{x}'\boldsymbol{\beta}),$$

where λ is absorbed into the baseline hazard without loss of generality.

The likelihood function for this data model, with perfect classification and no censoring is given by

$$L(\alpha, \boldsymbol{\beta}|\mathbf{x}) \;=\; \prod_{i=1}^{n} \alpha t_i^{\alpha-1}\exp(\mathbf{x}_i'\boldsymbol{\beta}). \tag{3.1}$$

The censored regression model can also be used, but we choose to assume no censoring in order to focus on ordered misclassification.

For the proportional hazards model, an intuitive interpretation for a coefficient, β_j, of a binary variable, x_j, is the hazard ratio. That is, $HR = \exp(\beta_j)$ is the hazard ratio for being in the group where $x_j = 1$ versus the group where $x_j = 0$. If $\beta_j = 0$, this indicates that x_j has no association with survival time; $\beta_j > 0$ indicates that $x_j = 1$ has a higher hazard of death, and $\beta_j < 0$ indicates that $x_j = 1$ has a lower hazard of death. This relationship may be obfuscated when x_j is misclassified and inferences must be made using its surrogate, x_j^*.

3.3 Covariate Misclassification and Adjustments

Suppose a test yields a binary response: a subject is positive or negative for some condition. Let $T+$ and $T-$ represent tests with positive and negative results, respectively. Let a subject's true status be $D+$ or $D-$, indicating positive or negative, respectively. Then *sensitivity* is $\eta = P(T+|D+)$ and *specificity* is $\theta = P(T-$

$|D-)$. Suppose we construct a model assuming perfectly classified data, utilizing the likelihood function described in (3.1). If misclassification is present, such that the likelihood function is truly described by

$$L(\alpha, \boldsymbol{\beta}|\mathbf{x}^*) \;=\; \prod_{i=1}^{n} \alpha t_i^{\alpha-1} \exp(\mathbf{x}_i^{*'}\boldsymbol{\beta}), \tag{3.2}$$

where $P(x^* = 1|x) = \eta x + (1-\theta)(1-x)$, then we have constructed a *naive model* because no correction for misclassification was performed.

On average, parameter estimates for models with misclassified covariates may be biased when no adjustments for misclassification are utilized (Greenland, 1980). Several methods used to adjust for misclassification in proportional hazards models are presented in papers such as Zucker and Spiegleman (2004), Wang and Song (2013), and Bang et al. (2013). Methods to adjust for covariate misclassification require information about the misclassification parameters, but sensitivity and specificity are rarely known exactly. In Bayesian modeling, beta distributions are ideal for representing uncertainty about probabilities. They are easily elicited using a variety of methods and their contribution to the analysis can be assessed with concepts like prior equivalent sample size (PESS). For an overview see, for example, Morita et al. (2008).

To make use of the additional information regarding the order of the covariate misclassification rates, we use a four parameter beta distribution as a prior for the misclassification parameter with the larger value conditioned on the misclassification parameter with the smaller value. The beta distribution on the interval $[u, v]$ has probability density function

$$\text{Beta}_{[u,v]}(x|a,b) \equiv \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)(v-u)^{a+b-1}}(x-u)^{a-1}(v-x)^{b-1},\ u \le x \le v,$$

where $a > 0, b > 0$. We use the notation $\text{Beta}_{[u,v]}(x|a,b)$ to denote a generic four-parameter beta distribution. We omit the subscript $[u, v]$ when $u = 0$ and $v = 1$. For more information on the general beta distribution, see Appendix A.1.

Suppose a covariate is subject to ordered misclassification depending on the misclassified covariate. For example, suppose a diagnostic screening for a life threat-

ening condition must be interpreted by a clinician. Since the consequence of failing to diagnose a patient is more severe than incorrectly diagnosing a patient as positive, it is ideal to yield a higher probability of false positive results than false negative results. Thus, it is desirable for the sensitivity to be higher than the specificity. Define the joint prior

$$\pi(\eta, \theta) = \pi(\eta|\theta)\pi(\theta),$$

where

$$\pi(\theta) = \text{Beta}(a_\theta, b_\theta),$$

and

$$\pi(\eta|\theta) = \text{Beta}_{[\theta,1]}(a_\eta, b_\eta).$$

The hyperparameters $(a_\theta, \ldots, b_\eta)$ are chosen to align the means of the prior distributions with the elicited sensitivities and specificities. See Garthwaite, Kadane, and O'Hagan (2005) or Kinnersley and Day (2013) for additional details on the prior elicitation process.

Suppose the covariate's ordered misclassification is attributable to another covariate, as in the race/ethnicty and hospital example. We assume

$$\pi(\eta_1, \eta_2) \perp \pi(\theta_1, \theta_2).$$

Let $\boldsymbol{\eta} = \eta_1, \eta_2$ and $\boldsymbol{\theta} = \theta_1, \theta_2$. For $\eta_1 < \eta_2$ and $\theta_1 < \theta_2$, we propose the dependent joint prior

$$\pi(\boldsymbol{\eta}, \boldsymbol{\theta}) = \pi(\eta_2|\eta_1)\pi(\eta_1)\pi(\theta_2|\theta_1)\pi(\theta_1),$$

where

$$\begin{aligned} \pi(\eta_1) &= \text{Beta}(a_{\eta_1}, b_{\eta_1}), \\ \pi(\eta_2|\eta_1) &= \text{Beta}_{[\eta_1,1]}(a_{\eta_2}, b_{\eta_2}), \\ \pi(\theta_1) &= \text{Beta}(a_{\theta_1}, b_{\theta_1}), \end{aligned}$$

and

$$\pi(\theta_2|\theta_1) = \text{Beta}_{[\theta_1,1]}(a_{\theta_2}, b_{\theta_2}).$$

The hyperparameters $(a_{\eta_1}, \ldots, b_{\theta_2})$ are chosen to align the means of the prior distributions with prior information about the misclassification. This can be generalized to categorical covariates with three or more levels, as described in Section 2.3.1.2.

We do not wish to construct a prior so informative that it dominates the posterior. However, an informative prior is warranted if *a priori* expert knowledge or historical data is available. One way to quantify this prior information is to do so with the concept of prior effective sample size, as discussed by Morita et al. (2008). In the context of this problem, it is appropriate to consider the sum of a beta distribution's shape parameters as a number of binomial trials, in which the first shape parameter is interpreted as the number of successes out of such trials.

3.4 Examples with One Covariate

The misclassification parameters' priors, $\pi(\eta)$ and $\pi(\theta)$, are often treated independently even though it is assumed that $P(\eta + \theta > 1) = 1$ for any test of interest. Although common, we should not ignore dependence due to the level of the misclassified covariate, as in the fallible test example where false positive results are more desirable than false negative results.

The model for this problem is diagramed in Figure 3.2 and is appropriate for $P(\theta < \eta) = 1$. For $P(\eta < \theta) = 1$, the joint prior for the misclassification parameters is

$$\pi(\eta, \theta) = \text{ Beta}(\eta|a_\eta, b_\eta)\ \text{Beta}_{[\eta,1]}(\theta|a_\theta, b_\theta).$$

In the following sections, we let the prior(s) for the regression parameter(s) be diffuse normal distributions given by

$$\beta_i \sim \text{N}(0, \tau),$$

where the precision, τ, is taken to be small. The likelihood for this data is given by

$$L(\alpha, \beta|\mathbf{x}^*) = \prod_{j=1}^{n} \alpha t_j^{\alpha-1} \exp(x_j^{*'}\beta)$$

and the joint prior is

$$\pi(\alpha, \beta, \theta, \eta) = \text{Exp}(\alpha|1.711)\text{N}(\beta|0, 0.01)\text{Beta}(\theta|7, 3)\ \text{Beta}_{[\theta,1]}(\eta|6.6, 3.4).$$

Thus, the posterior is

$$\pi(\alpha, \beta, \eta, \theta | \mathbf{x}^*) \propto L(\alpha, \beta | \mathbf{x}^*) \pi(\alpha, \beta, \eta, \theta).$$

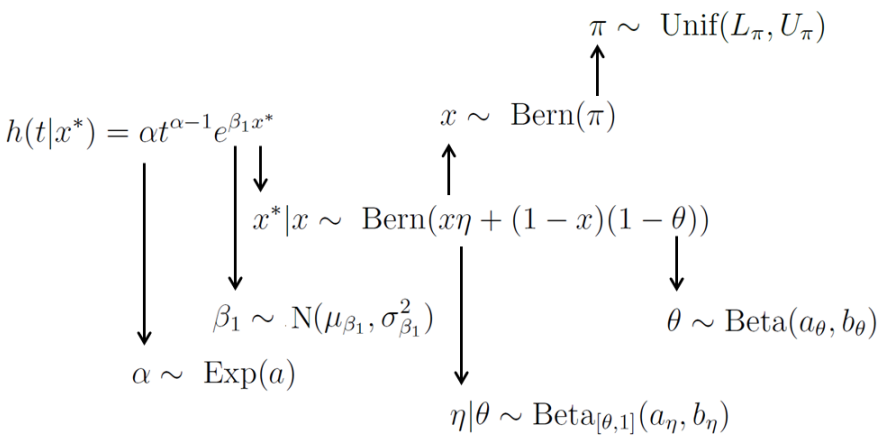

Figure 3.2: Proportional Hazards Model with a Covariate Subject to Ordered Misclassification.

Again consider the fallible test example in which a clinician must subjectively interpret diagnostic screenings. Suppose the screening is for a particular factor, x, and it is of interest to determine if the survival times for patients affected by the same type and stage of cancer depend on that particular factor. Further, the test used to detect x is fallible with $P(0 < \theta < \eta < 1) = 1$ and we only observe its surrogate, x^*. The patients can be categorized as group 1 if $x = 0$ and group 2 if $x = 1$. We assign the surrogate groups 1^* and 2^* based on the value of x^*. Suppose from historical evidence that $\theta = 0.7$. Using a prior equivalent sample size of 10, we specify a beta prior with mean 0.7:

$$\pi(\theta) = \text{Beta}(7, 3).$$

Additionally, suppose an expert in the field believes that the most likely value of η is 0.9 and that a value less than 0.85 is unlikely, which we interpret as the 5th percentile of the prior. Using this information, as well as a prior equivalent sample size of 10, yields the prior

$$\pi(\eta|\theta) = \text{Beta}_{[\theta,1]}(6.6, 3.4).$$

This prior has a range dependent on θ and has a mean of 0.9 when $\theta = 0.7$.

We generate Weibull proportional hazards data by substituting the values from Table 3.1 into the process detailed in Appendix B.2. Let ζ_1 and ζ_2 be the hypothetical medians of group 1 and group 2, respectively. We generate $n = 50$ survival times for each of the 2 groups for a total of $N = 100$ observations with $\eta = 0.9$ and $\theta = 0.7$, yielding an overall misclassification rate of 14%.

Table 3.1: Summary of Parameters for Fallible Test Example.

ζ_1	ζ_2	β	θ	η	α
90 Days	498 Days	-1.0	0.70	0.90	0.5843

We use two chains and a burn-in of 2000 values. We follow this with 10,000 posterior samples. No convergence issues were observed. The posterior densities are shown in Figure 3.3. The posterior summary statistics are provided in Table 3.2. Adjusting for the ordered misclassification yields a posterior estimate of β_1 that is nearly equal to the true parameter value. Additionally, the estimated hazard rate for those in group 1 compared to group 2 is 0.3844, which is close to the true parameter value, 0.3679. To gain insight on this model, which utilizes the ordered adjustment, we consider models utilizing different assumptions about misclassification.

Table 3.2: Posterior Results for the Fallible Test Example, Assuming Ordered Misclassification.

Parameter	Truth	Mean	SD	2.5%	Median	97.5%
β_1	-1.0000	-0.9918	0.3094	-1.6410	-0.9799	-0.4240
η	0.9000	0.9188	0.0477	0.8039	0.9282	0.9843
HR	0.3679	0.3844	0.1183	0.1868	0.3723	0.6499
θ	0.7000	0.7378	0.0918	0.5498	0.7403	0.9070

Recall the naive model which assumes the likelihood given in (3.1), when the data truly follow (3.2). This model typically yields inaccurate point estimates and artificially narrow interval estimates.

Incorporating independent beta prior distributions in Bayesian models can improve estimates when *a priori* information is available regarding the misclassification rates. This unordered misclassification adjustment is common practice, but it ignores

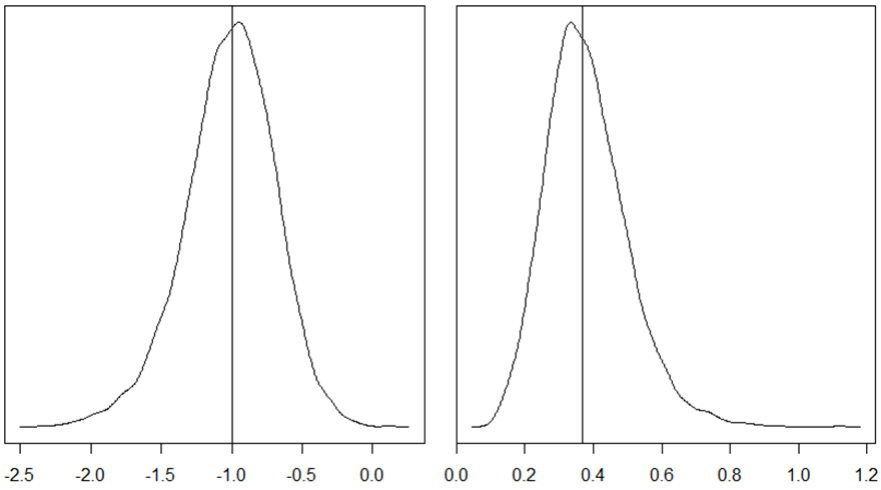

Figure 3.3: Posterior Densities of β_1 (left) and $HR = \exp(\beta_1)$ (right) for the Fallible Test Example, Assuming Ordered Misclassification. The vertical bar in each plot represents the true parameter value.

potentially beneficial information regarding order. In this example, the unordered joint prior on the misclassification parameters with a prior equivalent sample size of 10 is

$$\pi(\theta, \eta) = \text{Beta}(\theta|7, 3)\text{Beta}(\eta|9, 1).$$

We compare the results from the model assuming ordered misclassification to those obtained under the assumptions of the naive model, unordered adjustment model, and a perfectly classified model. We perform the analysis with the perfectly recorded data simply to reiterate that the generated data reflect our expectations.

As before, we use two chains and a burn-in of 2000 values. Then, we follow this with 10,000 posterior samples. No convergence issues were observed. We compare the posterior summary statistics for β_1 from these four models in Table 3.3. Additionally, we compare the posterior summary statistics for $HR = \exp(\beta_1)$ from these models in Table 3.4.

The naive model yields an estimate for β_1 that is far from the true parameter value, whereas the three remaining models' estimates for β_1 are all near the true value. The models acknowledging the presence of misclassification have posterior standard deviations that are larger than the models making no such assumption, as is to be

Table 3.3: Comparing Posterior Results of β_1 for the Fallible Test Example Under Different Misclassification Assumptions.

Model	Mean	SD	2.5%	Median	97.5%
Truth	−1.0000	–	–	–	–
Naive	−0.7892	0.2071	−1.1970	−0.7907	−0.3813
Unordered Adjustment	−0.9973	0.3107	−1.625	−0.9921	−0.4129
Ordered Adjustment	−0.9918	0.3094	−1.6410	−0.9799	−0.4240
Perfectly Classified	−0.9917	0.2058	−1.3890	−0.9917	−0.5834

expected. The potential advantage of modeling ordered misclassification parameters will be considered in Section 3.6.

Table 3.4: Comparing Posterior Results of $HR = \exp(\beta_1)$ for the Fallible Test Example Under Different Misclassification Assumptions.

Model	Mean	SD	2.5%	Median	97.5%
Truth	0.3679	–	–	–	–
Naive	0.4649	0.0961	0.3053	0.4554	0.6796
Unordered Adjustment	0.3920	0.1147	0.1993	0.3815	0.6508
Ordered Adjustment	0.3844	0.1183	0.1868	0.3723	0.6499
Perfectly Classified	0.3798	0.0787	0.2485	0.3714	0.5585

The posterior densities for β_1 under these four assumptions are graphed in Figure 3.4. Additionally, the posterior densities for HR under these four assumptions are graphed in Figure 3.5. Both figures reiterate the findings from Table 3.3 in that the two adjusted models yield estimates near the true parameter value and have more variability than the other two models.

The prior distributions for the misclassification parameters of this example are in in Figure 3.6. For this analysis, we choose a prior equivalent sample size of 10 because it is one-tenth of the total sample size. Also shown in Figure 3.6 are other possible prior distributions which reflect the same prior information regarding misclassification, but do so at different prior equivalent sample sizes. Distributions with larger prior equivalent sample sizes have less variability than those with smaller prior equivalent sample size.

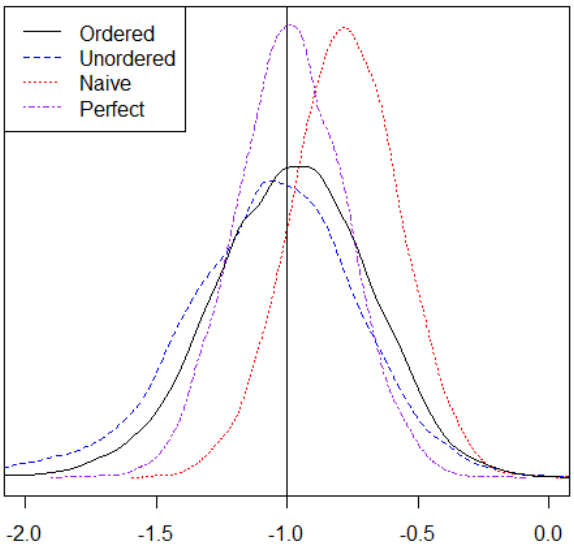

Figure 3.4: Posterior Densities for β_1 in the Fallible Test Example. The vertical bar represents the true value.

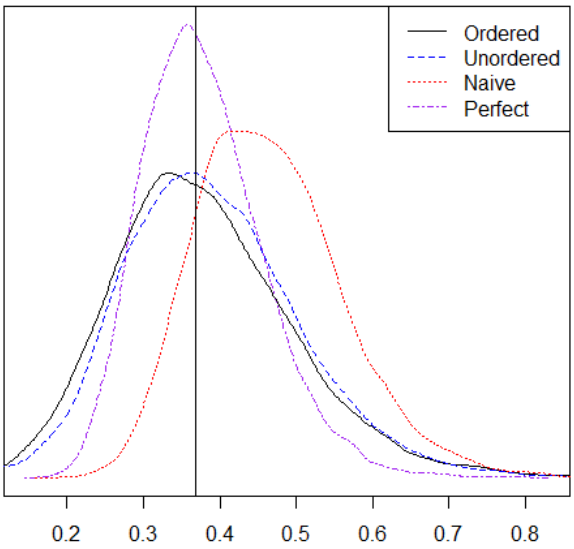

Figure 3.5: Posterior Densities for $HR = \exp(\beta_1)$ in the Fallible Test Example. The vertical bar represents the true value.

3.5 Examples with Two Covariates

Cancer mortality rates differ by race (CDC, 2014). Recall that Gomez and Glaser (2006) show that race/ethnicity is misclassified in the Greater Bay Area Cancer Registry and that the rates of misclassification differ among the races/ethnicities and hospitals in the registry. Further, Gomez et al. (2003) found that of the hospitals

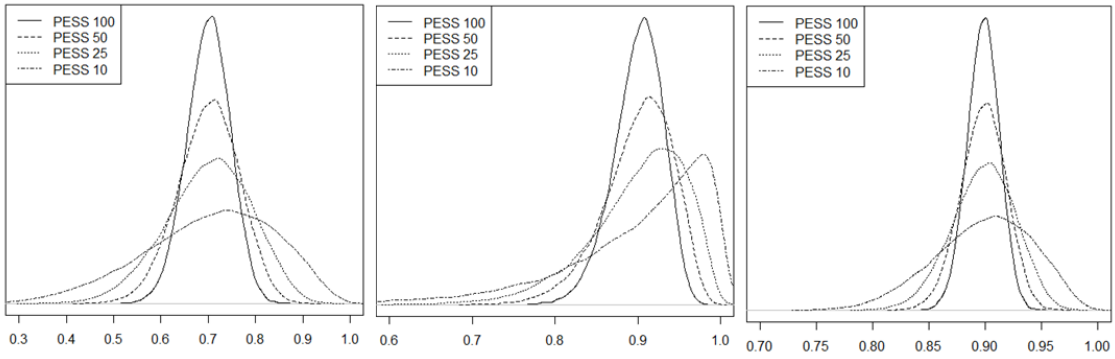

Figure 3.6: Possible Beta Priors for Misclassification Parameters with Varying Prior Equivalent Sample Sizes. Beta priors for θ with mean 0.7 (left). Beta priors for η with mean 0.9 (middle). Beta priors for $\eta|\theta$ with support (0.7,1) and mean 0.9 (right).

examined, roughly 20% reported assigning patients' race/ethnicity based only on surname. This gives rise to potential racial/ethnic misclassification, especially for women who take their husband's surname in an interracial/interethnic marriage. This scenario, with its potential misclassification, motivates the following hypothetical example.

Models incorporating knowledge of misclassification must utilize estimated values of misclassification parameters. Bayesian models that do so require priors on those parameters. These can be constructed with a combination of expert opinion and either internal or external validation data. See Prescott and Garthwaitte (2002) for an overview of obtaining prior distributions via validation substudy. In our case, suppose that we have a separate study of classification performance at hospitals contributing to the cancer registry utilized for the main survival data set. If that data set is available and can be itself analyzed with a Bayesian model, then the resulting posteriors on the misclassification parameters can be employed in the misclassified proportional hazards model. Suppose we have the slightly less convenient situation in which the separate data are not available, but the results of a frequentist analysis are available. In particular, suppose we have maximum likelihood estimators (MLEs) and 95% interval estimates for sensitivity and specificity. For example, suppose that the MLE of the sensitivity of racial/ethnic classification at a large hospital is $\hat{\eta}_1 =$

0.96 and the MLE of the sensitivity of racial/ethnic classification at a small hospital is $\hat{\eta}_2 = 0.96$. Further, the estimated positive predictive value at the large hospital is $\widehat{PPV}_1 = 0.78$ and the estimated positive predictive value at the cancer center is $\widehat{PPV}_2 = 0.92$. The MLE, $\hat{\theta}_i$, for $i = 1, 2$ is

$$\hat{\theta}_i = 1 - \frac{(\hat{\eta}_i\hat{\pi}_i/\widehat{PPV}_i) - \hat{\eta}_i\hat{\pi}_i}{1 - \hat{\pi}_i}.$$

Thus, the MLE of the specificity of racial/ethnic classification at the large hospital is $\hat{\theta}_1 = 0.73$ and the MLE of the specificity of racial/ethnic classification at the small hospital is $\hat{\theta}_2 = 0.92$.[1] Assuming that $\boldsymbol{\eta} \perp \boldsymbol{\theta}$ for $\boldsymbol{\eta} = (\eta_1, \eta_2)$ and $\boldsymbol{\theta} = (\theta_1, \theta_2)$, the Bayesian model for this general problem is diagramed in Figure 3.7.

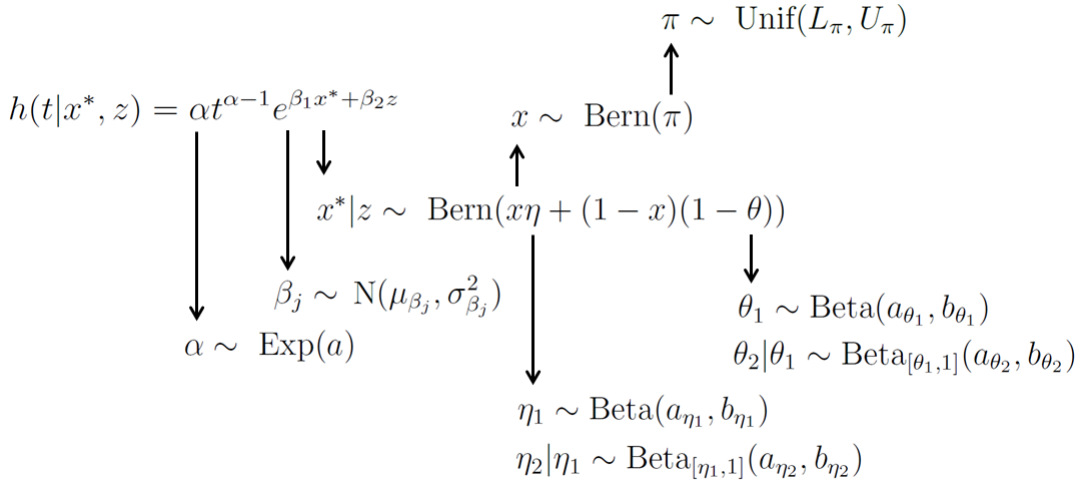

Figure 3.7: Proportional Hazards Model with a Perfectly Recorded Binary Covariate and a Covariate Subject to Ordered Misclassification.

Suppose that, given $\hat{\theta}_2 = 0.92$ and a 95% confidence interval, we set the 5th percentile of the prior for θ_2 at 0.82. Then, using a prior equivalent sample size of 20, we specify the joint prior for the misclassification parameters as

$$\pi(\eta_1, \eta_2, \theta_2|\theta_1) = \text{Beta}(\eta_1|19.2, 0.8)\text{Beta}(\eta_2|19.2, 0.8)\text{Beta}(\theta_1|14.6, 5.4)\text{Beta}_{[\theta_1,1]}(\theta_2|14, 6).$$

Note that because the estimates of η_1 and η_2 are the same, we do not employ the general beta correction for the sensitivities. The prior $\pi(\theta_2|\theta_1)$ ranges from θ_1 to 1 and has a mean of 0.92 when $\theta_1 = 0.73$.

[1] This is an adaptation from Gomez and Glaser (2006).

Define a subject's group as

$$\text{Group } = \begin{cases} 1 & \text{if } x = 0, z = 0; \\ 2 & \text{if } x = 1, z = 0; \\ 3 & \text{if } x = 0, z = 1; \\ 4 & \text{if } x = 1, z = 1. \end{cases}$$

We generate $n = 50$ survival times for each of the 4 groups for a total of $N = 200$ survival times based on the hypothetical median (ζ) survival times for each group displayed in Table 3.5.

Table 3.5: Summary of Parameters for Race/Ethnicity and Hospital Example.

ζ_1	ζ_2	ζ_3	ζ_4	β_1	β_2	η_1	η_2	θ_1	θ_2	α
90	498	45	251	-1.0	0.3	0.96	0.96	0.73	0.92	0.5843

The likelihood is

$$L(\alpha, \boldsymbol{\beta}|\mathbf{x}^*, \mathbf{z}) = \prod_{j=1}^{n} \alpha t_j^{\alpha-1} \exp(x_j^{*'}\beta_1) \exp(z_j'\beta_2)$$

and the joint prior is

$$\begin{aligned} \pi(\alpha, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\theta}) &= \text{Exp}(\alpha|1.711)\text{N}(\beta_1|0, 0.01)\text{N}(\beta_2|0, 0.01)\text{Beta}(\eta_1|19.2, 0.8) \\ &\quad \times \text{Beta}(\eta_2|19.2, 0.8)\ \text{Beta}(\theta_1|14.6, 5.4)\ \text{Beta}_{[\theta_1,1]}(\theta_2|14, 6). \end{aligned}$$

Thus, the posterior is

$$\pi(\alpha, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{x}^*, \mathbf{z}) \propto L(\alpha, \boldsymbol{\beta}|\mathbf{x}^*, \mathbf{z})\pi(\alpha, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\theta}).$$

To perform the analysis, we use two chains and a burn-in of 2000 values. We follow this with 10,000 posterior samples. No convergence issues were observed. The posterior summary statistics under the ordered misclassification assumption are provided in Table 3.6. The posterior estimate of $\beta_1, -1.036$, is close to the true parameter value, -1.000. Additionally, the estimated hazard rate of death for White, non-Hispanic patients compared to White, Hispanic patients is 0.3634, which is close to the true parameter value of 0.3679.

Table 3.6: Posterior Results for the Race/Ethnicity and Hospital Example, Assuming Ordered Misclassification.

Parameter	Truth	Mean	SD	2.5%	Median	97.5%
β_1	-1.0000	-1.0360	0.2215	-1.5100	-1.0220	-0.6395
β_2	0.3000	0.3239	0.1423	0.0402	0.3253	0.6019
η_1	0.9600	0.9523	0.0425	0.8469	0.9636	0.9990
η_2	0.9600	0.9703	0.0291	0.8944	0.9790	0.9996
HR	0.3679	0.3634	0.0774	0.2190	0.3611	0.5249
θ_1	0.7300	0.7296	0.0698	0.5867	0.7320	0.8593
θ_2	0.9200	0.9177	0.0338	0.8407	0.9217	0.9707

We compare the results obtained using the ordered misclassification assumption to those obtained with a naive model, a model using the unordered adjustment, and a perfectly classified model. The unordered joint prior is given by

$$\pi(\eta_1, \eta_2, \theta_1, \theta_2) = \text{Beta}(\eta_1|19.2, 0.8)\text{Beta}(\eta_2|19.2, 0.8)\text{Beta}(\theta_1|14.6, 5.4)\text{Beta}(\theta_2|18.4, 1.6).$$

To perform the analyses, we use two chains and a burn-in of 2000 values. We follow this with 10,000 posterior samples. The posterior summary statistics for β_1 from these four models are in Table 3.7. The naive model yields an estimate that is pulled toward zero, away from the true parameter value. Both of the adjusted models, as well as the perfectly classified model produce estimates for β_1 that are near the true parameter value. Additionally, the posterior summary statistics for $HR = \exp(\beta_1)$ using these four models are in Table 3.8. The posterior densities of β_1 for the four models are in Figure 3.8 and the posterior densities for the hazard rates are in Figure 3.9.

Table 3.7: Comparing Posterior Results for β_1 in the Race/Ethnicity and Hospital Example Under Different Misclassification Assumptions.

Model	Mean	SD	2.5%	Median	97.5%
Truth	-1.0000	–	–	–	–
Naive	-0.8368	0.1460	-1.1170	-0.8388	-0.5460
Unordered	-1.0350	0.2234	-1.5200	-1.0180	-0.6400
Ordered	-1.0360	0.2215	-1.5100	-1.0220	-0.6395
Perfectly Classified	-1.0440	0.1487	-1.3350	-1.0450	-0.7497

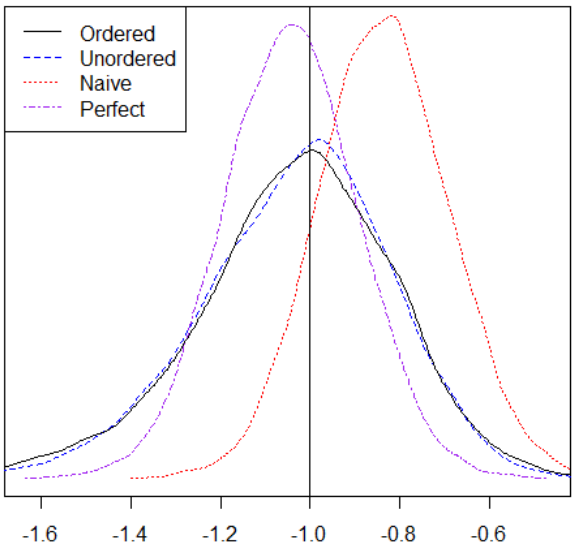

Figure 3.8: A Comparison of β_1 Densities for the Race/Ethnicity and Hospital Example. The vertical line represents the true value.

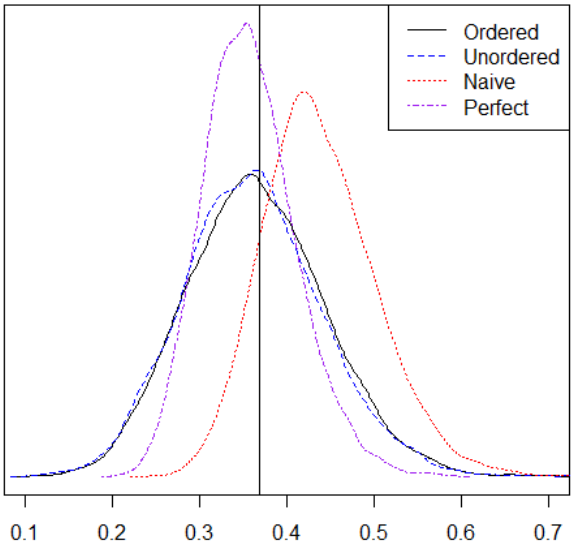

Figure 3.9: A Comparison of $HR = \exp(\beta_1)$ for the Race/Ethnicity and Hospital Example. The vertical line represents the true value.

3.6 Simulation for Single Covariate

To investigate the performance of and relationship between the models, we perform small-scale simulations assuming different levels of separation between ordered parameters. We quantify this difference as $\delta = [(\eta - \theta)/\theta] \times 100$ and examine $\delta = 10, 20$, and 30. We consider $n = 50$ for the 2 groups for a total sample size of $N = 100$, as before. The simulation design points are presented in Table 3.9. We

Table 3.8: Comparing Posterior Results for $HR = \exp(\beta_1)$ in the Race/Ethnicity and Hospital Example Under Different Misclassification Assumptions.

Model	Mean	SD	2.5%	Median	97.5%
Truth	0.3679	–	–	–	–
Naive	0.4365	0.0643	0.3246	0.4310	0.5741
Unordered	0.3638	0.0783	0.2188	0.3613	0.5273
Ordered	0.3634	0.0774	0.2190	0.3611	0.5249
Perfectly Classified	0.3564	0.0530	0.2637	0.3527	0.4703

compare the models assuming PESS = 10. To complete this simulation, we perform 5,000 burn-ins, 10,000 posterior samples, and 100 replications. Convergence diagnostics were examined for a random sample of replications and no problems were found. The simulation results for β_1 are in Table 3.10 and the simulation results for the hazard rate, $HR = \exp(\beta_1)$, are in Table 3.11.

Table 3.9: Design Points for Single Covariate Simulation.

δ	β_1	θ	η
10%	-1.00	0.70	0.77
20%	-1.00	0.70	0.84
30%	-1.00	0.70	0.91

Table 3.10: Single Covariate Simulation Results for β_1. Mean of posterior means (coverage) [width].

Model	$\delta = 10$	$\delta = 20$	$\delta = 30$
Truth	-1.0000	-1.0000	-1.0000
Naive	-0.5604	-0.6481	-0.7658
	(0.3600)	(0.5800)	(0.7900)
	[0.7559]	[0.7616]	[0.7789]
Unordered Adjustment	-1.0171	-1.0083	-1.005
	(1.0000)	(1.0000)	(1.0000)
	[1.6743]	[1.5060]	[1.3262]
Ordered Adjustment	-0.9926	-0.9892	-0.9959
	(1.0000)	(1.0000)	(1.0000)
	[1.6462]	[1.4834]	[1.2717]

The simulation summary graphics are displayed in Figure 3.10. The line across the middle of each plot represents the true value of β_1 and the black circle represents the median of the 100 posterior means. Additionally, the horizontal black lines below

and above the circle are the medians of the 100 posterior 2.5th and 97.5th percentiles, respectively. The grey boxes surrounding the lines represent ±1 simulation standard deviation of the posterior means and percentiles. The naive model's median is biased toward zero, while both of the adjusted models are practically unbiased. The model using the ordered adjustment yields slightly narrower bands than the model using the unordered adjustment.

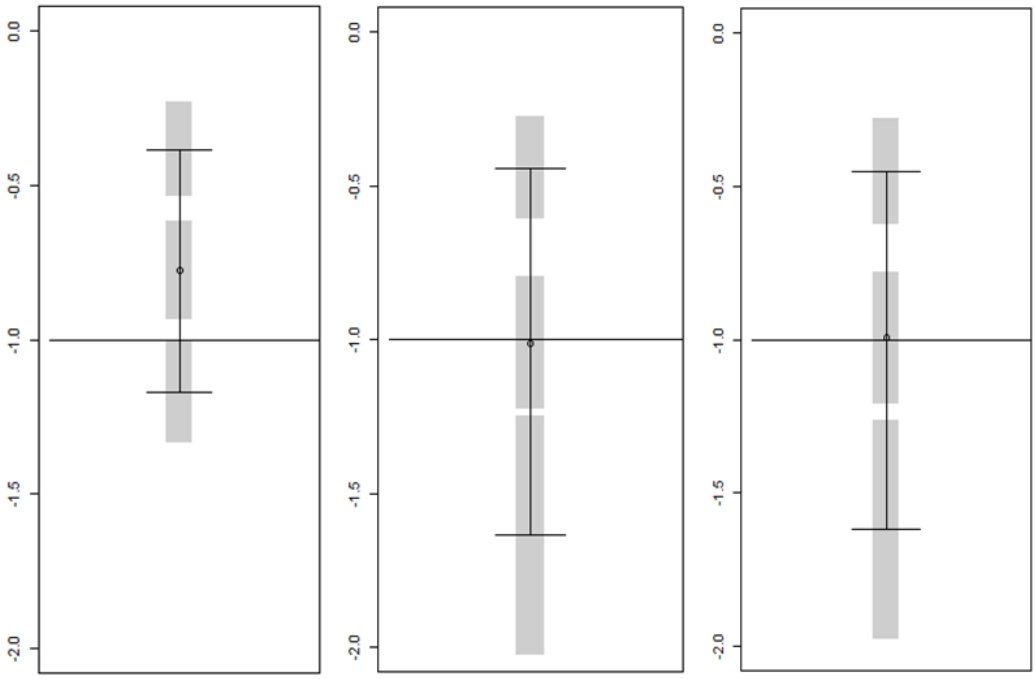

Figure 3.10: Simulation Summary Graphics for Single Covariate Simulation at $\delta = 30$. From left to right: naive model, unordered adjusted model, ordered adjusted model.

3.7 Simulation for Two Covariates

As in the previous section, primary interest for this problem is in the separation between dependent misclassification parameters. We again quantify this difference as $\delta = [(\theta_2 - \theta_1)/\theta_1] \times 100$ and examine $\delta = 10, 20$, and 30. We consider $n = 50$ for the 4 groups for a total sample size of $N = 200$. The simulation design points are in Table 3.12. We perform the misclassification adjusted analyses assuming PESS = 20, which is one-tenth of the total sample size. To complete the simulations, we perform 5,000 burn-ins, 10,000 posterior samples, and 100 replications. Convergence diagnostics were examined for a random sample of replications and no problems were

Table 3.11: Single Covariate Simulation Results for HR. Mean of posterior means (coverage) [width].

Model	$\delta = 10$	$\delta = 20$	$\delta = 30$
Truth	0.3679	0.3679	0.3679
Naive	0.5886	0.5401	0.4802
	(0.3600)	(0.5800)	(0.7900)
	[0.4457]	[0.4119]	[0.3752]
Unordered Adjustment	0.4133	0.4060	0.3962
	(1.0000)	(1.000)	(1.0000)
	[0.6116]	[0.5511]	[0.4804]
Ordered Adjustment	0.4228	0.4108	0.3965
	(1.0000)	(1.0000)	(1.0000)
	[0.6174]	[0.5481]	[0.4709]

found. The results are for β_1 are in Table 3.13 and the results for $HR = \exp(\beta_1)$ are in Table 3.14.

Table 3.12: Design Points for Two Covariate Simulation.

δ	β_1	β_2	η_1	η_2	θ_1	θ_2
10%	-1.00	0.30	0.96	0.96	0.70	0.77
20%	-1.00	0.30	0.96	0.96	0.70	0.84
30%	-1.00	0.30	0.96	0.96	0.70	0.91

Table 3.13: Two Covariates Simulation Results for β_1. Mean of posterior means (coverage) [width].

Model	$\delta = 10$	$\delta = 20$	$\delta = 30$
Truth	-1.0000	-1.0000	-1.0000
Naive	-0.8006	-0.8229	-0.8512
	(0.7400)	(0.7800)	(0.8300)
	[0.5770]	[0.5759]	[0.5732]
Unordered Adjustment	-1.1447	-1.1515	-1.1544
	(0.9400)	(0.9500)	(0.9600)
	[0.8481]	[0.8218]	[0.7772]
Ordered Adjustment	-1.1394	-1.1392	-1.1460
	(0.9400)	(0.9500)	(0.9600)
	[0.8463]	[0.7969]	[0.7658]

The simulation summary graphics are displayed in Figure 3.11. The line across the middle of each plot represents the true value of β_1 and the black circle represents the median of the 100 posterior means. Additionally, the horizontal black lines below

and above the circle are the medians of the 100 posterior 2.5th and 97.5th percentiles, respectively. The grey boxes surrounding the lines represent ±1 simulation standard deviation of the posterior means and percentiles. The naive model yields estimates that are biased toward zero. The misclassification adjusted models have reduced bias, but slightly overcorrect the effect of misclassification bias. The model using the ordered adjustment yields slightly narrower bands than the model using the unordered adjustment.

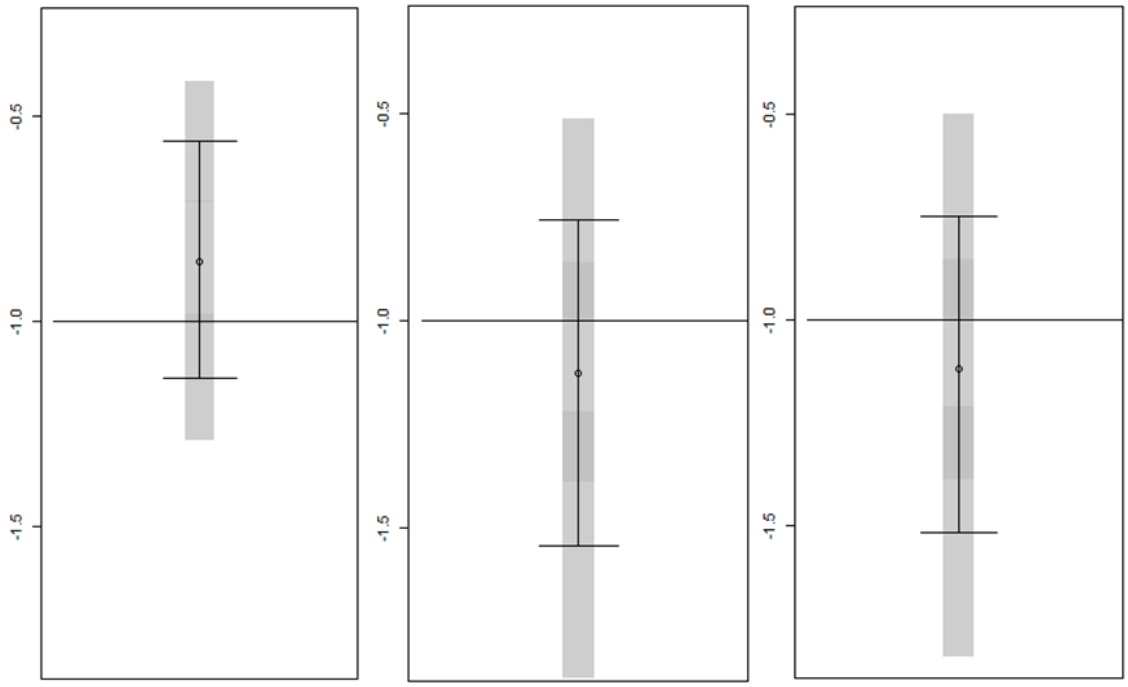

Figure 3.11: Simulation Summary Graphics for Two Covariate Simulation at $\delta = 20$. From left to right: naive model, unordered adjusted model, ordered adjusted model.

3.8 Concluding Remarks

Proportional hazards models enable researchers to evaluate the relationship between time-to-event data and covariates. When one or more of the covariates is misclassified, the inferences drawn from the model become unreliable, as the parameter estimates do not likely reflect the true parameter values. There are scenarios which may cause the covariate's misclassification rates to be ordered. In this chapter, we presented an example in which a fallible test yields more false positive than false negative results. We also explored an example derived from Gomez and Glaser (2006)

Table 3.14: Two Covariates Simulation Results for HR. Mean of posterior means (coverage) [width].

Model	$\delta = 10$	$\delta = 20$	$\delta = 30$
Truth	0.3679	0.3679	0.3679
Naive	0.4600	0.4493	0.4368
	(0.7400)	(0.7800)	(0.8300)
	[0.2661]	[0.2592]	[0.2510]
Unordered Adjustment	0.3335	0.3306	0.3288
	(0.9400)	(0.9500)	(0.9600)
	[0.2730]	[0.2625]	[0.2497]
Ordered Adjustment	0.3355	0.3342	0.3315
	(0.9400)	(0.9500)	(0.9600)
	[0.2738]	[0.2599]	[0.2482]

in which racial/ethnic misclassification of patients depends on the type of hospital in which the patient is treated. In the first example, η and θ are ordered amongst themselves. In the second, the misclassification order occurs from the levels of another covariate.

When order exists with probability one, an independent prior structure is inappropriate, as the order may not be preserved. We proposed an adjustment for ordered misclassification based on a conditional prior structure, using the general beta distribution. This adjustment preserves the order of misclassification with probability one. Adjusting for ordered covariate misclassification with the proposed method requires minimal additional work compared to the unordered misclassification adjustment. In addition to preserving the misclassification order, the method also yields improved estimation precision in some cases.

We recognize that it is not always appropriate to assume order with probability one. If this assumption cannot be made, then this adjustment is inappropriate. This adjustment, as well as any Bayesian method, is limited by the quality of historical data and/or expert opinion obtained.

CHAPTER FOUR

Bayesian Sample Size Determination for Informative Hypotheses

4.1 Introduction

Suppose we design an experiment to produce continuous responses with group means $\mu_1, \ldots, \mu_J$ which share a common variance, σ^2. Furthermore, we believe the means follow the order given by

$$H_I : \mu_1 < \mu_2 < \cdots < \mu_J. \tag{4.1}$$

This is commonly referred to as an *informative hypothesis* in the Bayesian literature (Hoijtink, 2012). Suppose we wish to construct a test of this informative hypothesis against its complement. Then we have H_I vs. H_C: $\mu_i \geq \mu_j$ for some $i < j$. Define the effect size as

$$\delta_i \equiv \left| \frac{\mu_i - \mu_{i+1}}{\sigma} \right|, \ i = 1, \ldots, J-1.$$

Additionally, define the error probabilities

$$\alpha_C = P(\text{select } H_I | H_C) \tag{4.2}$$

and

$$\alpha_I = P(\text{select } H_C | H_I). \tag{4.3}$$

Suppose we wish to distinguish between H_I and H_C at pre-specified effect sizes and error probabilities. We must find an appropriate sample size to satisfy these conditions. In this chapter, we do so by implementing a Bayesian sample size determination technique based on empirical selection error probabilities and the two-priors approach. As we shall see, this new approach affords advantages over the traditional sample size methods employed in Hoijtink (2012).

4.2.1 Frequentist Testing Procedures

In the frequentist setting, the most commonly utilized procedure to analyze three or more group means with equal variance is a oneway analysis of variance (ANOVA). The hypotheses used in testing $j = 1, \ldots, J$ normal means in this setting are

$$H_0 : \mu_i = \mu_j \text{ for all } i, j$$

and

$$H_A : \mu_i \neq \mu_j \text{ for some } i \neq j,$$

where H_0 denotes the null hypothesis and H_A denotes the alternative. In the usual approach (Stoline, 1981), if the overall F-test rejects the null hypothesis, further tests are required to distinguish the relationships between the groups' means. These tests include pairwise comparisons and orthogonal contrasts. Performing tests additional to the overall F-test increases the probability of incorrectly rejecting a null hypothesis one or more times, the family-wise error rate. This family-wise error rate inflation is directly related to a decrease in power. Even utilizing adjustments such as Bonferonni's method may not provide an optimal solution to the multiple test requirement.

Due to family-wise error rate inflation and other drawbacks associated with ad hoc multiple tests, frequentists developed methods specifically to test hypotheses with inequality constrained parameters. These methods include likelihood ratio tests based on restricted maximum likelihood estimators and planned contrast tests (Robertson et al., 1988). The likelihood ratio tests are evaluated with a χ^2 or F distribution and are formed using restricted maximum likelihood estimates. Planned contrast tests are evaluated using a t or F distribution. If the one-sided p-value is significantly small, the researcher rejects the null hypothesis and automatically concludes that the means follow the order specified by the alternative hypothesis. Incorporating inequality constraints in the alternative hypothesis can increase power and decrease

the time required for testing by eliminating the need for ad hoc multiple tests. The literature on order restricted frequentist analysis is rich (see Silvapulle et al., 2004 or Kopylev, 2012, for example). However, we consider a Bayesian formulation of this problem and focus ultimately on sample size determination.

4.2.2 Bayesian Hypothesis Testing

The Bayesian approach to hypothesis testing relies on posterior probabilities, information criteria (for model selection), or on Bayes factors, which contrast prior and posterior odds for the hypotheses of interest (Robert, 2007). Bayes factors provide the support of one hypothesis relative to another. A value larger than 1 indicates preference for the hypothesis in the numerator, a value smaller than 1 indicates preference for the hypothesis in the denominator, and a value of 1 indicates no preferential model. There exists much debate on cutoff values for Bayes factors in an effort to avoid the arbitrariness of the 0.05 'rule' of p-value significance in frequentist analysis. In the context of our problem, we contrast the prior and posterior odds for H_I and H_C. Here, the posterior probability of H_I refers to the probability of the set

$$\mathcal{H}_I = \{\boldsymbol{\mu} \in \mathbb{R}^J : \mu_1 < \cdots < \mu_J\}, \tag{4.4}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J)$. Additionally, the set corresponding to H_C is simply its complement, $\mathcal{H}_C \equiv \overline{\mathcal{H}_I}$.

The ability to use prior information is an advantage of the Bayesian approach. Another is that, given the posterior, any number of hypotheses can be tested, including pairwise tests. The multilevel Bayesian model requires neither ad hoc tests to determine the order of parameters, nor a correction for the family wise error rate. Gelman, Hill, and Yajima (2012) note that "rather than correcting for a perceived problem, we just build the multiplicity into the model from the start."

In the Bayesian evaluation of informative hypotheses, all hypotheses with inequality constraints are nested in an unconstrained (encompassing) hypothesis de-

noted by

$$H_E : \mu_1, \mu_2, \ldots, \mu_J, \tag{4.5}$$

corresponding to the set

$$\mathcal{H}_E = \{\boldsymbol{\mu} \in \mathbb{R}^J\}. \tag{4.6}$$

A prior for this hypothesis is called an *encompassing prior* (Klugkist et al., 2005). Through use of indicator functions, all priors corresponding to informative hypotheses nested in the encompassing hypothesis may be created. An advantage to this approach is that these priors only require the hyperparameters corresponding to the unconstrained model, H_E, to be specified. This set of priors does not favor any particular model. Additionally, the corresponding marginal priors are designed to be vague. This effectively creates a class of priors that are informative solely through the specified order. For example, in a J group ANOVA model with

$$y_i = \sum_{j=1}^{J} \mu_j d_{ji} + \varepsilon_i, \varepsilon_i \sim \mathrm{N}(0, \sigma^2),$$

one possible encompassing prior is given by

$$\pi(\boldsymbol{\mu}, \sigma^2 | H_E) = \prod_{j=1}^{J} \mathrm{N}(\mu_j | \mu_0, \tau_0^2) \Gamma^{-1}(\sigma^2 | a, b),$$

where $\mathrm{N}(\cdot|\cdot)$ and $\Gamma^{-1}(\cdot|\cdot)$ denote normal and inverse gamma densities, respectively. Here, μ_0, τ_0^2, a, and b are all hyperparameters chosen to make $\pi(\boldsymbol{\mu}, \sigma^2 | H_E)$ relatively noninformative.[1] We can create a prior for the informative hypothesis, H_I, by limiting the domain of the encompassing prior to match the order specified by the informative hypothesis via an indicator function. We have

$$\pi(\boldsymbol{\mu}, \sigma^2 | H_I) = \frac{\prod_{j=1}^{J} \mathrm{N}(\mu_j | \mu_0, \tau_0^2) \Gamma^{-1}(\sigma^2 | a, b) I_{H_I}}{\int\int \prod_{j=1}^{J} \mathrm{N}(\mu_j | \mu_0, \tau_0^2) \Gamma^{-1}(\sigma^2 | a, b) I_{H_I} d\boldsymbol{\mu} d\sigma^2}, \tag{4.7}$$

where

$$I_{H_I} = \begin{cases} 1 & \text{If the order follows that of } H_I; \\ 0 & \text{Otherwise.} \end{cases}$$

[1] A prior which is noninformative relative to the likelihood function. This prior has little impact on the posterior.

Note that the normal and inverse-gamma are conditionally independent given the ordering. One could provide a more general covariance structure, but like Klugkist et al., 2005, we retain the conditional independence structure.

As an example, Figure 4.1 displays 95% contours for three bivariate normal priors for (4.4) with $J = 2$. Clearly, all three priors distribute probability equally on both inequality constrained spaces, i.e. $P(\mu_1 < \mu_2) = P(\mu_2 < \mu_1) = 0.5$. The two circular contours are from encompassing priors, in that they satisfy the covariance structure implied by (4.7). The prior represented by the more eccentric elliptical contour is not considered encompassing because it requires a more general covariance structure.

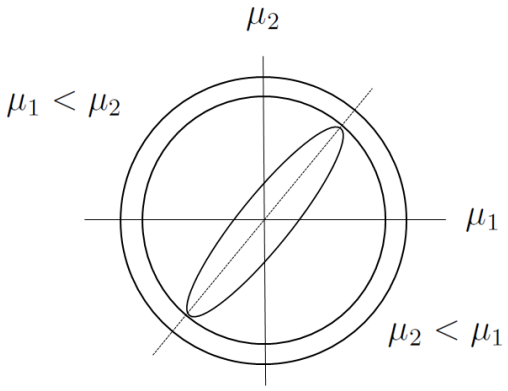

Figure 4.1: 95% Contours of Bivariate Priors. The circular contours satisfy the requirements of encompassing priors. The elliptical contour does not satisfy the encompassing prior covariance structure requirement.

As previously mentioned, the unconstrained (encompassing) prior is specifically designed not to favor an hypothesis over another. Generally, for models with strict inequalities and J means, there are $J!$ possible configurations of the means, including that in (4.1). To model a lack of a prior preference of one configuration over another, we can require that each has prior probability $1/J! \equiv c_I$.

We suggest the following formal definitions for the continuous case. Let $\mathcal{P}_J \equiv \mathcal{P}\{1, 2, \ldots, J\}$ denote the set of permutations of the integers $\{1, 2, \ldots, J\}$. Let $\mathbf{p} \equiv$

$(p_1, \ldots, p_J) \in \mathcal{P}_J$. Define the set

$$\mathcal{H}_{\mathbf{p}} = \{\boldsymbol{\mu} \in \mathbb{R}^J : \mu_{p_1} < \mu_{p_2} < \cdots < \mu_{p_J}\}.$$

For any $\mathbf{p} \in \mathcal{P}_J$, the *complexity* (Klugkist, Laudy, and Hoijtink, 2005) of the informative hypothesis is the prior probability[2]

$$c_I = \int_{\mathcal{H}_{\mathbf{p}}} \pi(\boldsymbol{\mu}) d\boldsymbol{\mu}, \text{ for any } \mathbf{p} \in \mathcal{P}_J.$$

Further, $c_I = 1/J!$ for any $\mathbf{p} \in \mathcal{P}_J$ with only strict inequalities. Similarly, the *fit* of the informative hypothesis is the posterior probability

$$f_I = \int_{\mathcal{H}_{\mathbf{p}}} \pi(\boldsymbol{\mu}|\text{data}) d\boldsymbol{\mu}, \text{ for any } \mathbf{p} \in \mathcal{P}_J.$$

Klugkist et al. (2005) showed that the Bayes factor for comparing an informative hypothesis against an unconstrained alternative hypothesis, (4.5), is given by

$$BF_{IE} = f_I / c_I.$$

Since $H_C = \overline{\mathcal{H}_I}$, the complexity and fit of the complementary collection of hypotheses are $c_C = 1 - c_I$ and $f_C = 1 - f_I$, respectively. Thus, the Bayes factor comparing H_C to H_E is

$$BF_{CE} = \frac{f_C}{c_C} = \frac{(1 - f_I)}{(1 - c_I)}.$$

Finally, Van Rossum et al. (2013) conclude that the Bayes factor for testing an informative hypothesis against its complement is

$$BF_{IC} = \frac{BF_{IE}}{BF_{CE}} = \frac{f_I / c_I}{(1 - f_I)/(1 - c_I)}.$$

They select H_I when $BF_{IC} > 1$ and select H_C when $BF_{IC} < 1$. Thus,

$$\alpha_C = P(\text{select } H_I | H_C) = P(BF_{IC} > 1 | H_C)$$

and

$$\alpha_I = P(\text{select } H_C | H_I) = P(BF_{IC} < 1 | H_I).$$

[2] Our notation and formal definition of these concepts is new, but the ideas of complexity, encompassing priors, etc., are due to Klugkist, Hoijtink, and their colleagues.

Although there is no null hypothesis in the Bayesian analysis outlined above, the resulting errors are analogous to that of Type I and Type II errors in the frequentist paradigm. The errors in both paradigms may be intentionally altered through the choice of sample size.

4.3 Bayesian Sample Size Determination

Due to cost, time, and, especially in biomedical applications, ethical constraints,[3] sample size determination is a crucial component of the experimental design process. Because of its importance, there exist many methods to determine an appropriate sample size. In fields such as drug development, operating characteristics (error probability, coverage, average credible interval width, etc.) must be examined, even in a Bayesian analysis, as detailed in the *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials* (FDA 2010). The sample size determination problem of interest arises from choosing between H_I and H_C. The challenge is to choose a sample size that satisfies some criterion regarding the hypothesis test performance.

4.3.1 Fixed Parameters

Choosing between H_I and H_C can result in two types of errors, either by incorrectly selecting H_I or incorrectly selecting H_C, as defined in (4.2) and (4.3), respectively. As previously discussed, Van Rossum et al. (2013) perform their analysis with encompassing priors and then use the Bayes factor, BF_{IC}, to make selection decisions. As suggested by Hoijtink (2012), Van Rossum et al. (2013) perform a Bayesian simulation at three fixed sample sizes in a repeated sampling framework to obtain the empirical error probabilities associated with their testing procedure. We consider this in more detail in Section 4.4.2. They subsequently find the appropriate sample size for the desired effect size based on pre-defined error probabilities.

[3] We want to avoid treating patients with an inferior drug when there is a more efficacious alternative. The latter might be the existing standard of treatment or the new drug; either way we do not want to expose more patients than necessary to the inferior treatment.

4.3.2 Two-Priors Approach

As in Van Rossum et al. (2013), we use a Bayesian sample size determination method based on empirical error rates associated with the hypothesis test of H_I v. H_C. However, we utilize the two-priors approach as discussed in Brutti et al. (2008). Doing so increases flexibility in simulation studies by replacing fixed parameters with probability distributions. These distributions are called *design priors* and are typically very informative. They are used specifically for sampling purposes. An example of incorporating design priors in a simulation study as opposed to specifying fixed parameters is illustrated in Figure 4.2.

In the figure, two models are specified, one with common effect size 0.5 and the other with 0.2. The number line represents a typical simulation relying on fixed parameters. The two sets of density plots illustrate the variability incorporated during the data generation process. Initially, we use the same standard deviation in the design priors for the models at both effect sizes, which explains the variation in the distributional overlap in the two sets of density plots. Selecting different values for the design prior standard deviation allows manipulation of the distributional overlap.

After using design priors to find parameter values, data are generated, and *analysis priors* are incorporated as in a typical Bayesian analysis. This two-priors method generally requires the use of Markov chain Monte Carlo methods. To our knowledge, the two-priors approach has not been used in sample size determination for testing informative hypotheses.

4.4 Simulation

4.4.1 Motivating Example

As discussed in Section 4.2.2, Van Rossum et al. (2013) developed a method using BF_{IC} to test one informative hypothesis, H_I, against its complement, H_C. In the article, they apply their selection method to data in Van de Schoot et al. (2010) on the association between popularity and antisocial behavior.

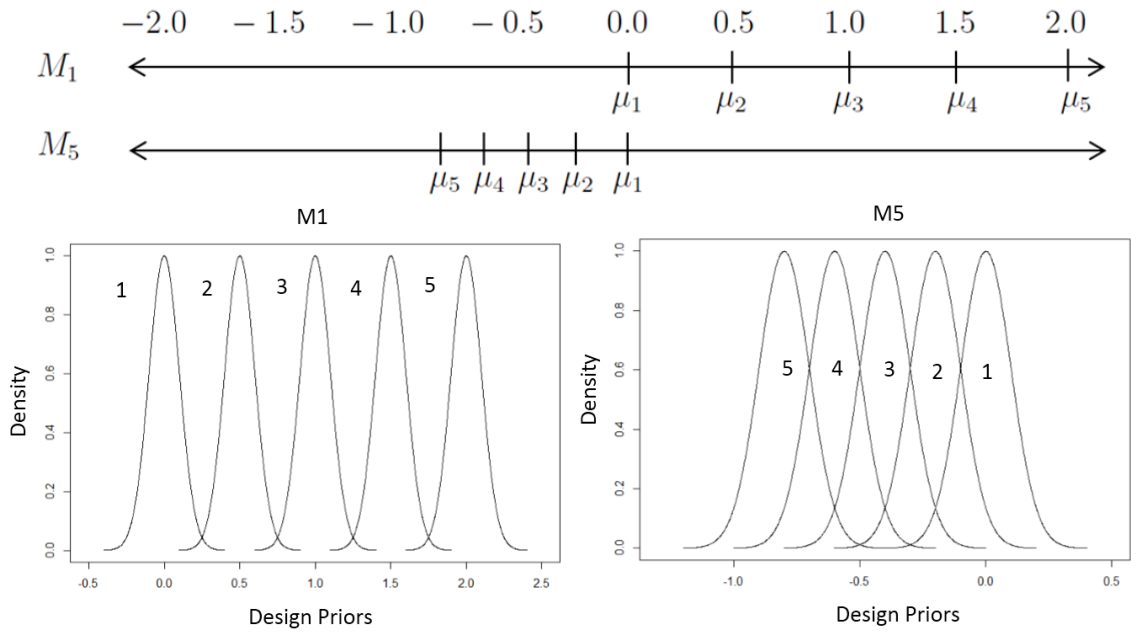

Figure 4.2: Design Prior Flexibility for Selected Models from Section 4.4.1. M_1 has $\delta_i = 0.5\ \forall i$ and M_5 has $\delta_i = 0.2\ \forall i$.

In that article, subjects from preparatory vocational schools are first categorized into five sociometric status groups using a peer assessment tool. The five sociometric groups are Group 1: neglected, Group 2: popular, Group 3: average, Group 4: rejected, Group 5: controversial. Once the groups are established, a slightly modified version of the Anti-Social Behavior Questionnaire (Host et al., 1998) is administered to the students. The questionnaire is comprised of the question: 'did you conduct this behavior' for twelve items, such as 'stealing money from home.' Each item is measured with a four point frequency scale (no, once, sometimes, often). Van de Schoot et al. (2010) expect the average frequencies of antisocial behavior to be smallest for the neglected group, followed by the popular, average, rejected, and controversial groups.

Van Rossum et al. (2013) use the example from Van de Schoot et al. (2010) to develop an informative hypothesis; however, they change the range of possible values for their simulation to be between −0.8 and 2.0. The researchers' informative hypothesis is

$$H_I : \mu_1 < \mu_2 < \mu_3 < \mu_4 < \mu_5. \tag{4.8}$$

The task is to find a sample size that enables the researchers to distinguish between H_I and H_C while keeping the α_I and α_C errors 'small.'

4.4.2 Fixed Parameters

Van Rossum et al. (2013) examine five models for the antisocial behavior problem at three preset sample sizes. Figure 4.3 is one of many possible representations of those five models. Table 4.1 specifies the actual values assigned to the means used in the models. Additionally, they set $\sigma^2 = 1$ for all groups across all models. Van Rossum et al. (2013) design M_1 and M_2 to follow the order constraints in H_I at effect sizes 0.5 and 0.2, respectively. Additionally, they choose M_3, M_4, and M_5 in H_C with an increasing number of H_I order violations at effect sizes 0.5, 0.5, and 0.2, respectively.

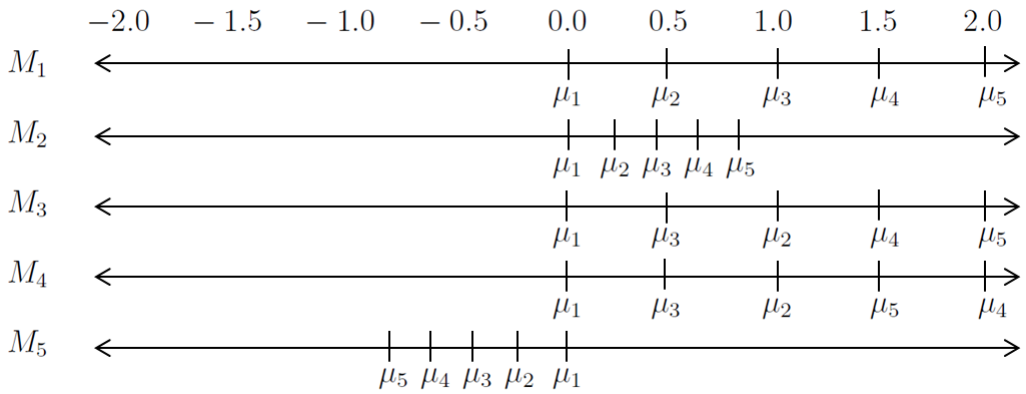

Figure 4.3: Possible Mean Scores of the Anti-social Behavior Questionnaire by Group.

Table 4.1: Parameter Specification for Antisocial Behavior Example.

Model	μ_1	μ_2	μ_3	μ_4	μ_5
M_1	0.0	0.5	1.0	1.5	2.0
M_2	0.0	0.2	0.4	0.6	0.8
M_3	0.0	1.0	0.5	1.5	2.0
M_4	0.0	1.0	0.5	2.0	1.5
M_5	0.0	−0.2	−0.4	−0.6	−0.8

To complete their sample size simulation, they use a Gibbs sampler and perform 1000 burn-ins and 10,000 posterior iterations in WinBUGS for 1000 replications at each M_k, $k = 1, \ldots, 5$, and each $n = 10, 20, 40$. They calculate BF_{IC} for each

replication and record the corresponding choice of hypothesis based on the predefined cutoff value of 1. Then they calculate the empirical error probabilities (α_I and α_C) and make sample size decisions based on those error rates. The results are compared to the two-priors simulation results in Figure 4.8.

4.4.3 Two-Priors

The simulation algorithm for the two priors approach applied to the sample size problem for informative hypotheses at a fixed sample size is represented by the diagram in Figure 4.4. The algorithm is as follows:

(1) Select an hypothesis and specify the corresponding design and analysis priors, maximum tolerated empirical error rate, and maximum sample size.

(2) Generate parameters using the design priors.

(3) Conditional on the generated parameters, generate data from a K-dimensional (K = number of groups) multivariate normal distribution with a constant variance and no correlation between groups.

(4) Combine the generated data with the analysis priors in WinBUGS to obtain posterior distributions.

(5) Calculate BF_{IC} and record the model selections for all N replications.

(6) Calculate the empirical error probabilities.

(7) Repeat with different sample sizes until the maximum tolerated empirical error rate is achieved with the smallest sample size.

For the antisocial behavior example, we generate data from the K-dimensional multivariate normal distribution using the design prior values in Table 4.1 as the means. To be consistent with the simulation setup in Van Rossum et al. (2013), we specify a constant variance ($\sigma^2 = 1$) of antisocial behavior, and no correlation between the sociometric groups. To obtain results comparable to Van Rossum et al. (2013),

we use the five models summarized in Table 4.1 at sample sizes (10, 20, 40), (analysis) priors, burn-ins (1000), posterior samples (10,000), and replications ($N = 1000$).

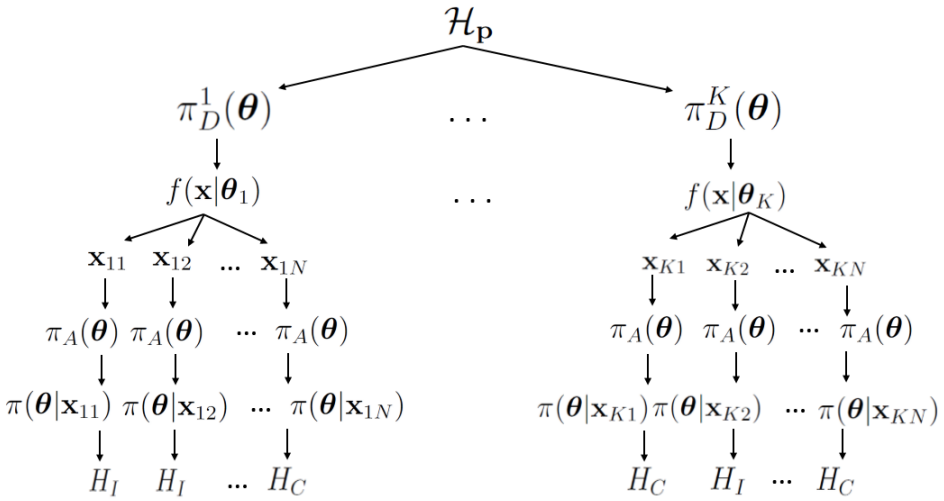

Figure 4.4: Diagram of Simulation Algorithm.

The joint design priors can be written as

$$\pi(\tilde{\mu_j}, \tilde{\sigma}) = \pi(\tilde{\mu_j}|\tilde{\sigma})\pi(\tilde{\sigma}).$$

We take $\pi(\tilde{\mu_j}|\tilde{\sigma})$ as $\mathrm{N}(\mu_j, \phi^2)$ and $\pi(\tilde{\sigma})$ as $\mathrm{U}(a, b)$. We choose $\phi = 0.1$, and set $a = 0.5$, and $b = 1.5$ to center the uniform distribution at 1 because this is the value that Van Rossum et al. (2013) use as the standard deviation between each μ_j. Note that use of a uniform prior on the standard deviation is preferred over the once common inverse-gamma prior because of its superior MCMC convergence properties (see Spiegelhalter et al., 2004 and Gelman, 2006). We set $\phi = 0.1$ to allow some possibility of overlap in the distributions of the means. Figure 4.2 depicts the overlap for models 1 and 5, which represent both effect sizes examined in this simulation. There is a noticeable difference in overlap for the two effect sizes. For this reason, we perform an additional simulation in Section 4.5.2 to investigate the effect of changing the value of ϕ on sample size.

To facilitate comparison to the results in Van Rossum et al. (2013), we use the same analysis priors, chosen to be relatively noninformative and provide equal support for all possible ordered models. The priors are given below as joint independent

normal-gamma (gamma on precision, inverse-gamma on variance) distributions. Since this example problem has a maximum range of 2 units across group means, specifying a precision of 0.001 in the prior may seem alarming; however, there are no problems with convergence.

$$Y_i \sim N\left(\sum_{j=1}^{5} \mu_j d_{ji}, \tau\right)$$
$$\downarrow \qquad \longrightarrow \ \tau \sim \Gamma(0.01, 0.01)$$
$$\mu_j \sim N(0.0, 0.001)$$

Figure 4.5: Analysis Priors for 5 Group ANOVA Model.

We randomly sample simulation replications to examine data generation and convergence. We include one plot for each assumption investigated in this chapter. To verify the generated data follow the order we expect, we create box plots for samples of simulation replications. An example plot is shown for the data using Model 1 in Figure 4.6. The data clearly follow the requisite ordering.

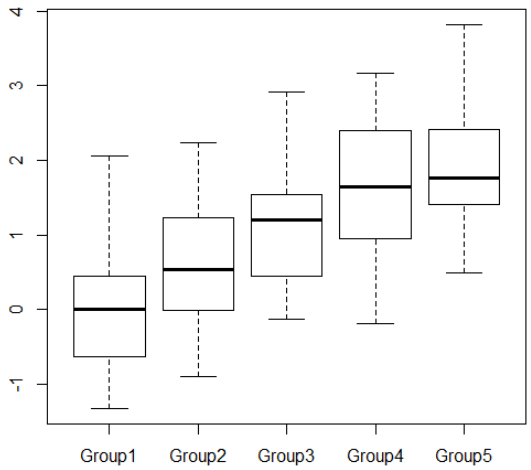

Figure 4.6: Boxplots for Model 1 Sample Size 40.

We investigate MCMC convergence for a randomly chosen sample of simulation replications, using parallel chains. Gelman-Rubin plots and other diagnostics reveal no convergence problems for the sampled replications.

After obtaining the 1000 replications for each model, we investigate the simulation's variability by examining posterior credibility intervals. Figure 4.7 displays groups 1 and 5 of Model 1. The line across the middle of each plot corresponds to the true parameter value for the group (from Table 4.1). The two short lines above and below the middle line are the medians of the 1000 95% credible interval upper and lower bounds, respectively. Additionally, the black circle is the median of the 1000 posterior means. The grey vertical bar represents ± 1 simulation standard deviation from the median of 1000 posterior means and 95% credible interval bounds. The two dark grey bands indicate overlapping simulation standard deviations. As expected, the simulation variability decreases as the sample size increases and looks similar for both groups. Because the data are normal and symmetric, we could have elected to use the less robust mean summary statistic instead of the median.

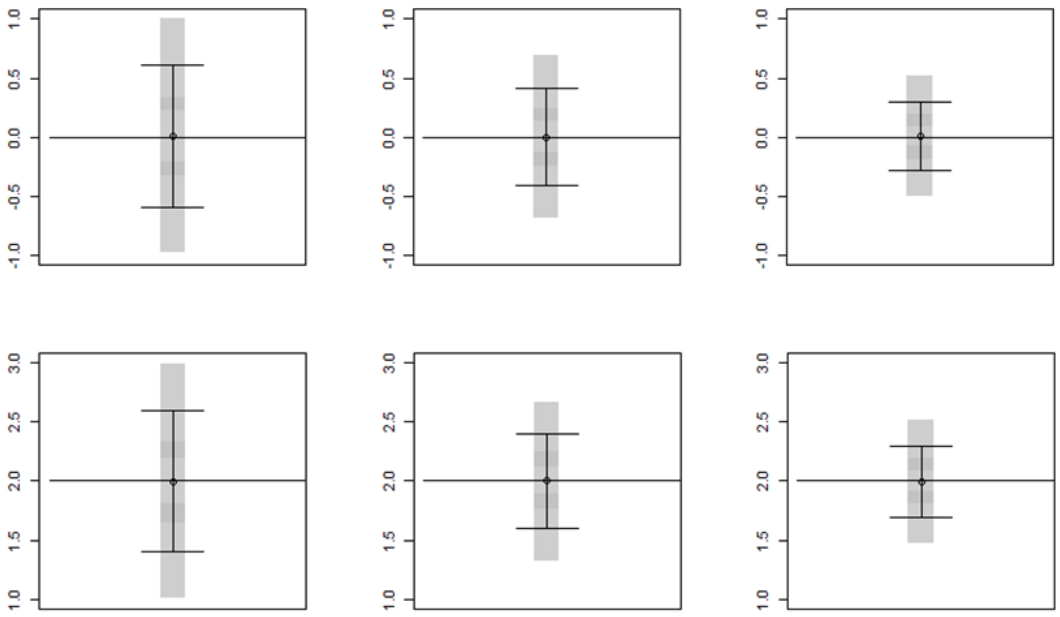

Figure 4.7: Simulation Summary for Model 1: Group 1 (top). Group 5 (bottom). N = 10, 20, 40 (left to right).

The empirical error probabilities displayed on the left side of Figure 4.8 correspond to Models 1 and 2 (Table 4.1), which both follow the order specified in H_I (4.8). Therefore, the error probabilities (α_I) are the percents of times out of 1000 that H_C is selected. An example of how to use this plot follows. Suppose the client is

certain his model follows M_2 and wants to bound α_I at 0.075. Then we suggest the method of sample size determination relying on fixed parameters, which requires a sample size of 20. However, if the client is less certain his model follows M_2, but still wants to bound α_I at 0.075, we suggest the two priors approach, which incorporates his uncertainty regarding the parameters' order, in the simulation through the use of design priors. Doing so requires a sample size of 40.

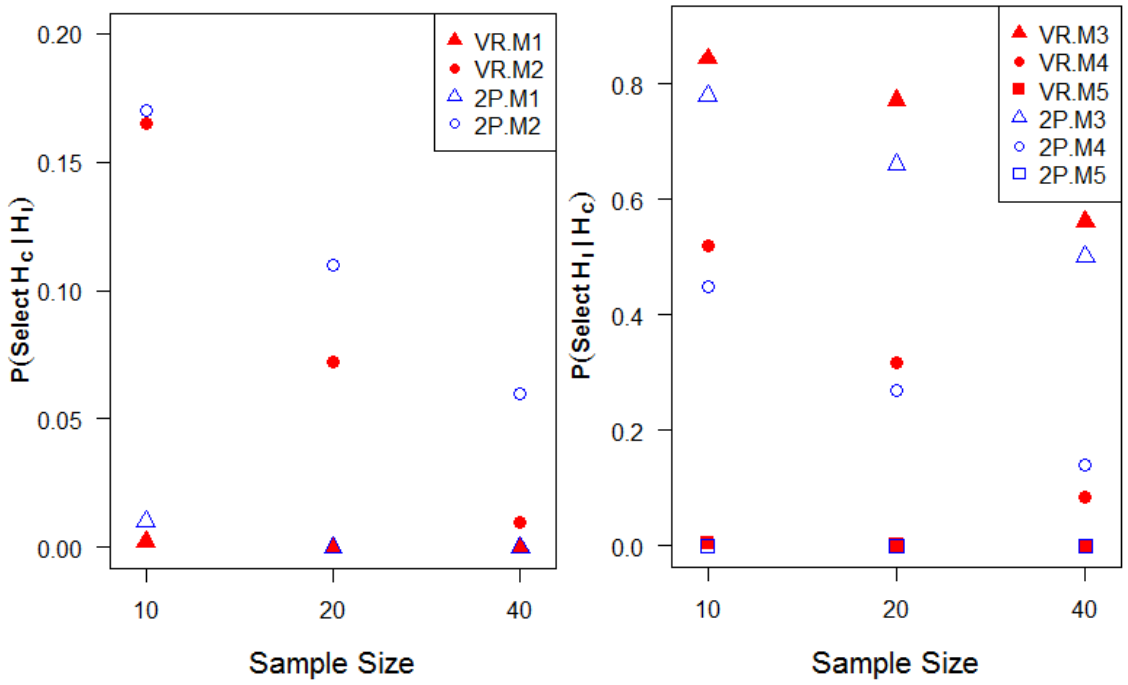

Figure 4.8: Empirical Errors from the Simulation for Models 1 and 2 (left) and 3, 4, and 5 (right).

The error rates displayed on the right side of Figure 4.8 correspond to Models 3, 4, and 5 (Table 4.1), which all violate the order given in H_I (4.8). Model 3 has one paired violation, Model 4 has two paired violations, and Model 5 completely violates the order in H_I. Since BF_{IC} distinguishes between H_I and H_C, the errors (α_C) represent the percents of times out of 1000 that H_I is selected. The Bayes factor easily determines that Model 5 follows the order given by H_C, even with a sample as small as 10. Model 4 requires a sample size of at least 40 to provide a reasonable error probability for both the fixed parameter and two-priors simulation methods. Even at a sample size of 40, the error rate for Model 3 is high using both simulation methods

($\alpha_C > 0.5$). Because these error rates are so high, it may not be appropriate to use BF_{IC} to distinguish between an informative hypothesis and its complement when there is one paired order violation present in the data for a five parameter model.

4.5 Additional Simulation Experiments

4.5.1 Dimensionality

When few order violations are present relative to the number of groups in a model (i.e. Model 3), BF_{IC} has difficulty correctly distinguishing H_I from H_C. For this reason, the sample size required to obtain the maximum tolerated error rate is large. In the 5-dimensional case examined in Section 4.4.3, there is potential for the two-priors approach to yield a smaller sample size requirement than the fixed parameter simulation because the empirical error rates obtained are lower than the fixed parameter simulation's. For this reason, we investigate the effect of dimensionality on the sample size requirement using the fixed parameter simulation method and the two-priors simulation method.

In Figure 4.9, we illustrate a three parameter model with one paired order violation at effect size 0.5. To perform the simulation, we use the model diagrammed in Figure 4.4 and the corresponding algorithm from Section 4.4.3. For this problem, we generate data from the 3-dimensional multivariate normal distribution using the design prior values of $\mu_1 = 0.0$, $\mu_2 = 1.0$, $\mu_3 = 0.5$ as the mean vector components from which to generate parameter values. To be consistent with the simulation setup in Van Rossum et al. (2013), we specify a constant variance ($\sigma^2 = 1$) within groups, and no correlation between the groups. To obtain results comparable to Van Rossum et al. (2013), we use the same sample sizes (10, 20, 40), (analysis) priors, burn-ins (1000), posterior samples (10,000), and replications ($N = 1000$). Note that the analysis priors for this problem are depicted in Figure 4.5; however, instead of $j = 1, \ldots, 5$, now $j = 1, 2, 3$.

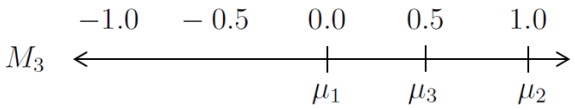

Figure 4.9: 3-Dimensional Model on the Number Line.

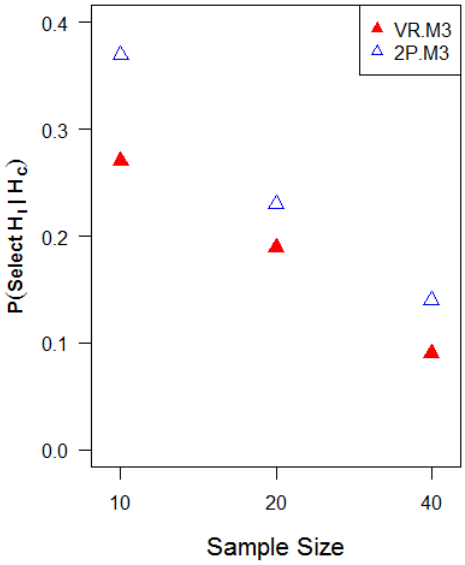

Figure 4.10: Error Rates for 3-Dimensional Hypothesis.

We investigate MCMC convergence for a randomly chosen sample of simulation replications, using parallel chains. Gelman-Rubin plots and other diagnostics reveal no convergence problems for the sampled replications. Figure 4.10 shows that the three parameter model requires a smaller sample size than the five parameter model (Figure 4.8) to obtain acceptably small errors using both simulation methods. Additionally, the Van Rossum method results in smaller error rates at each of the sample sizes examined, which can lead to different sample size requirements. For example, suppose the client wishes to have a maximum error rate of 0.1. Then the Van Rossum fixed parameter approach requires a sample size of 40, but the two-priors approach requires a sample size larger than 40.

Figure 4.11 depicts a seven parameter model with one and two paired order violations. To perform the simulation, we again use the model diagrammed in Figure 4.4 and the corresponding algorithm from Section 4.4.3. For this problem, we

generate data from the 7-dimensional multivariate normal distribution using the design prior values shown in Figure 4.11 as the means for parameter generation. To be consistent with the simulation setup in Van Rossum et al. (2013), we specify a constant variance ($\sigma^2 = 1$) within groups, and no correlation between the groups. To obtain results comparable to Van Rossum et al. (2013), we use the same sample sizes (10, 20, 40), (analysis) priors, burn-ins (1000), posterior samples (10,000), and replications ($N = 1000$). Note that the analysis priors for this problem are captured in Figure 4.5; however, instead of $j = 1, \ldots, 5$, now $j = 1, \ldots, 7$.

We investigate MCMC convergence for a randomly chosen sample of simulation replications, using parallel chains. Gelman-Rubin plots and other diagnostics reveal no convergence problems for the sampled replications. Figure 4.12 shows that the sample size required to obtain an acceptable error rate is higher for these two models than any other dimension examined. Although it is clear that a sample size greater than 40 is required, it is also evident that the two-priors simulation yields smaller error rates than the Van Rossum approach, which could produce a smaller sample size requirement.

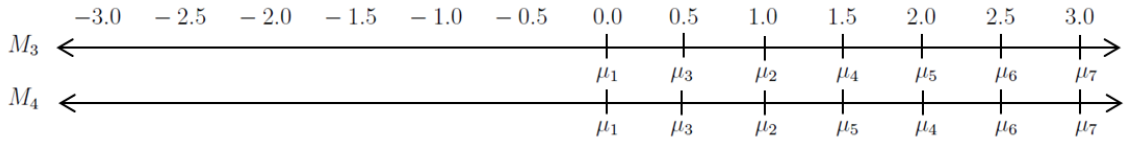

Figure 4.11: 7-Dimensional Model on the Number Line.

4.5.2 Design Prior Variation

Increasing uncertainty in parameter values may correspond to an increase in the requisite sample size for a given error probability specification. This level of uncertainty is reflected by altering the design prior standard deviation, ϕ. Changing the value of ϕ changes the extent of overlap in the design prior distributions, as seen in Figure 4.2. This overlap represents the probability that the design priors yield parameter values violating the order specified by the informative hypothesis. It is not

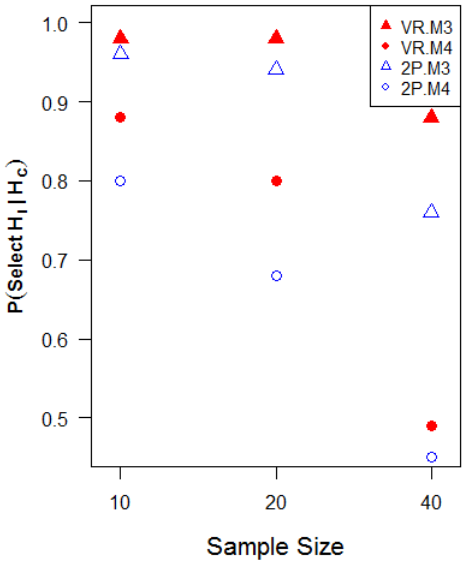

Figure 4.12: Error Rates for 7-Dimensional Hypotheses.

known to what extent this alters the sample size requirement, so we investigate this potential change via simulation.

The sample size simulation in Section 4.4.3 specifies $\phi = 0.1$ for both effect size 0.2 and 0.5. For the latter effect size, this value of ϕ corresponds to an approximate 1% probability of generating parameter values that violate the specified order, for any two of the neighboring design priors. For example, the two neighboring design priors may be $\mathrm{N}(\mu_2 = 0.5, \phi = 0.1)$ and $\mathrm{N}(\mu_3 = 1.0, \phi = 0.1)$. Alternatively, $\phi = 0.1$ corresponds to an approximate 30% probability of generating parameter values that violate the specified order, for any two of the neighboring design priors with the 0.2 effect size. In this simulation, we examine three values of ϕ for both effect sizes while allowing the probability of order violation among generated parameter values to range between roughly 1% and 30%. The values of ϕ designed to achieve the desired probability of order violation are displayed in Table 4.2.

The model diagrammed in Figure 4.4 is employed again, along with the corresponding algorithm from Section 4.4.3. For this problem, we generate data from the 5-dimensional multivariate normal distribution using the design prior values shown in Table 4.1 as the means. To be consistent with the simulation setup in Van Rossum et

al. (2013), we specify no correlation between the groups and $\sigma^2 = 1.0$. We change the value of ϕ in the design priors to achieve the appropriate probability of order violation (Table 4.2). We use the same sample sizes (10, 20, 40), (analysis) priors, burn-ins (1000), posterior samples (10,000), and replications ($N = 1000$) as in Section 4.4.3 and Section 4.5.1.

Table 4.2: Specification of Design Prior Standard Deviation, ϕ.

ϕ	Effect Size	Overlap
0.100	0.5	1%
0.175	0.5	15%
0.250	0.5	30%
0.040	0.2	1%
0.070	0.2	15%
0.100	0.2	30%

We investigate MCMC convergence for a randomly chosen sample of simulation replications, using parallel chains. Gelman-Rubin plots and other diagnostics reveal no convergence problems for the sampled replications. Figure 4.13 shows that increasing the overlap in the design priors leads to higher error probabilities when H_I is true. This indicates that using design priors with large values of ϕ, relative to the effect size, may require the largest sample size to obtain acceptable error rates. This result is expected because as the probability of overlap in design priors increases, so does the probability of an H_I order violation. This plot can be used for sample size determination, as was illustrated with Figure 4.8 in Section 4.4.3. For example, suppose a client believes his data follow Model 2 and wishes to use the smallest sample size that achieves a 0.10 error rate. A sample size of 20 is sufficient to obtain this result for the fixed parameter simulation and the simulations which have design priors with 1% and 15% probability of generating parameter values that violate the order. However, a sample size of 40 is necessary to achieve this for a 30% probability of overlap in these design priors.

Figure 4.13 shows a general decrease in empirical error rates as the extent of overlap in design priors increases when H_C is true. A sample size greater than 40 is

required to produce reasonable error rates for Model 3. When the sample size is 10, Model 4 yields the same order of empirical errors as in Model 3. When the sample size is 40, the pattern does not hold. A sample size of 40 or larger is required to obtain acceptable error rates for Model 4. There are virtually no selection errors for Model 5; these results are consistent with those from Section 4.4.3.

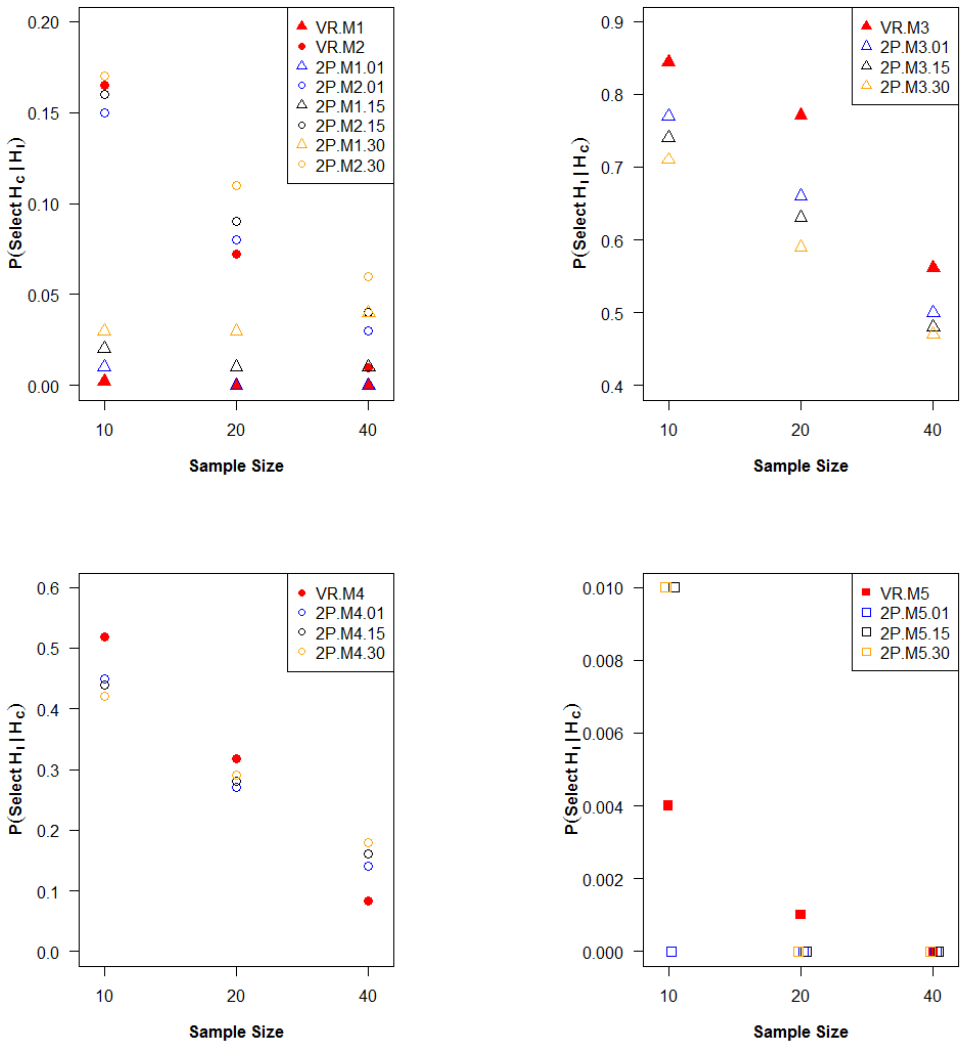

Figure 4.13: Errors for Design Prior Simulation. Top left (Models 1 and 2). Top right (Model 3). Bottom left (Model 4). Bottom right (Model 5).

4.5.3 Hypothesis with Constraint and Unknown Relationships

At times, interest may lie primarily with one group's mean compared to a set of other group means having either an unknown or unimportant relationship. For example, suppose we have four group means and the researcher suspects the order is

given by

$$H_{I2} : \mu_1 > \max\{\mu_2, \mu_3, \mu_4\}. \tag{4.9}$$

We continue to use an ANOVA model encompassing prior for this problem. One possibility is

$$\pi(\boldsymbol{\mu}, \sigma^2 | H_E) = \prod_{j=1}^{J} \mathrm{N}(\mu_j | \mu_0, \tau_0^2) \Gamma^{-1}(\sigma^2 | a, b),$$

where μ_0, τ_0^2, a, and b are all hyperparameters chosen to make $\pi(\boldsymbol{\mu}, \sigma^2 | H_E)$ relatively noninformative. We can create a prior for the informative hypothesis, H_{I2}, by limiting the domain of the encompassing prior to match the order specified by the informative hypothesis via an indicator function. We have

$$\pi(\boldsymbol{\mu}, \sigma^2, H_{I2}) = \frac{\prod_{j=1}^{J} \mathrm{N}(\mu_j | \mu_0, \tau_0^2) \Gamma^{-1}(\sigma^2 | a, b) I_{H_{I2}}}{\int \int \prod_{j=1}^{J} \mathrm{N}(\mu_j | \mu_0, \tau_0^2) \Gamma^{-1}(\sigma^2 | a, b) I_{H_{I2}} d\boldsymbol{\mu} d\sigma^2},$$

where

$$I_{H_{I2}} = \begin{cases} 1 & \text{If the order follows that of } H_{I2}; \\ 0 & \text{Otherwise.} \end{cases}$$

We are now interested in testing the informative hypothesis, H_{I2}, against its complement, $H_{C2} : \overline{H_{I2}}$. To perform this test, we still use BF_{IC}; however, we must redefine the set

$$\mathcal{H}_{\mathbf{p}} = \{\boldsymbol{\mu} \in \mathbb{R}^J : \mu_{p_1} > \max\{\mu_{p_2}, \ldots, \mu_{p_J}\}\},$$

where $\mathcal{P}_J \equiv \mathcal{P}\{1, 2, \ldots, J\}$ denotes the set of permutations of the integers $\{1, 2, \ldots, J\}$ and $\mathbf{p} \equiv (p_1, \ldots, p_J) \in \mathcal{P}\{1, 2, \ldots, J\}$, as before. For any $\mathbf{p} \in \mathcal{P}_J$, $c_I = 1/J$.

The model diagrammed in Figure 4.4 is employed again, along with the corresponding algorithm from Section 4.4.3. Now, the appropriate design prior means are in Table 4.3. To obtain results comparable to Van Rossum et al. (2013), we use the same five models summarized in Table 4.3, sample sizes (10, 20, 40), (analysis) priors (Figure 4.14), burn-ins (1000), posterior samples (10,000), and replications (1000).

The empirical error rates for BF_{IC} at each sample size and model are given in Figure 4.15. Across all fifteen design points, the two priors and Van Rossum methods

Table 4.3: Parameter Specification for H_{I2} Simulation.

Model	μ_1	μ_2	μ_3	μ_4
M_1	0.5	0.0	0.0	0.0
M_2	0.2	0.0	0.0	0.0
M_3	−0.2	0.0	0.0	0.0
M_4	0.5	0.0	1.0	0.0
M_5	0.5	0.0	1.0	1.5

$$Y_i \sim N\left(\sum_{j=1}^{4} \mu_j d_{ji}, \tau\right)$$

$$\tau \sim \Gamma(0.01, 0.01)$$

$$\mu_j \sim N(0.0, 0.001)$$

Figure 4.14: Analysis Priors for H_{I2} Simulation.

likely produce the same sample size requirement, as the empirical error rates are similar.

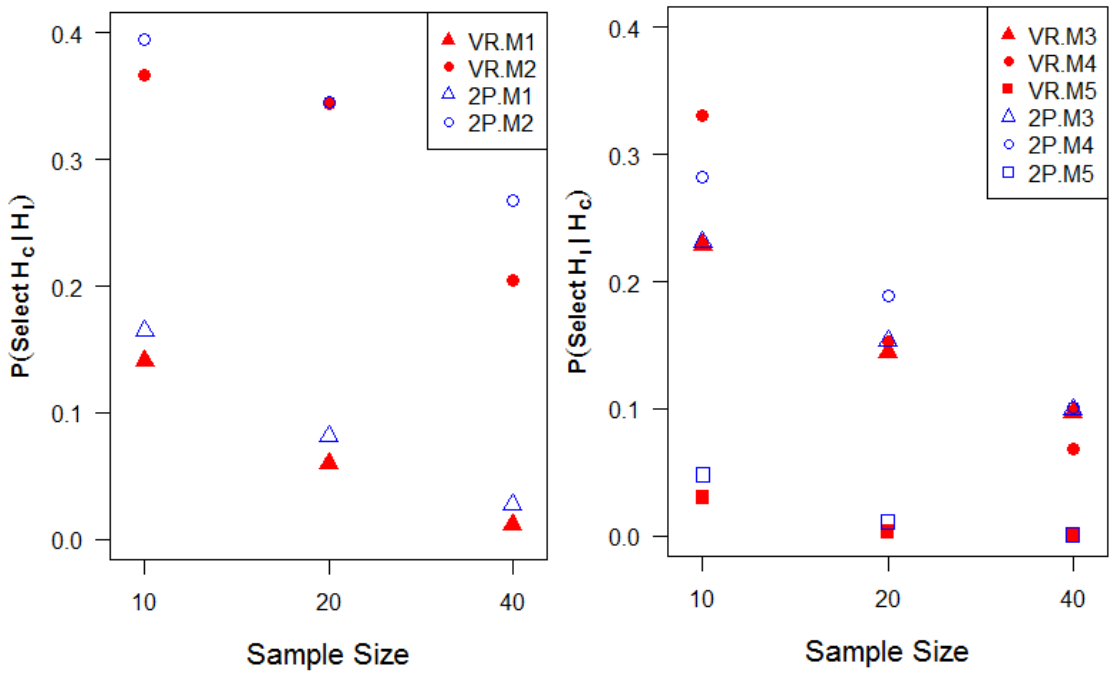

Figure 4.15: Empirical Errors from the H_{I2} Simulation for Models 1 and 2 (left) and 3, 4, and 5 (right).

4.6 Concluding Remarks

In this chapter, we presented a sample size determination technique for informative hypotheses which uses an empirical error rate criterion with the two-priors approach. We performed a simulation study to investigate the effect of substituting design priors for fixed parameters on sample size requirements. We did so at two different effect sizes (0.5 and 0.2) in a five parameter model and then extended the simulation to both three and seven-dimensional models. Upon comparison of our results to those in Van Rossum et al. (2013), we discovered that the sample size requirements produced from these methods may not always coincide. Utilizing the two priors in the simulation provided the flexibility to incorporate a researcher's uncertainty regarding order through the design priors. Additionally, changing the value of the design prior standard deviation, ϕ, altered the requisite sample size as expected. In particular, when H_I is true, increasing the extent of overlap in the design priors resulted in larger error probabilities, which required larger sample sizes to obtain acceptable error rates. When H_C is true, increasing the overlap of design priors resulted in smaller error probabilities, which required smaller sample sizes to obtain acceptable error rates. Further, we performed sample size determination for a model with one inequality constraint and an unknown relationship between three remaining variables. The fixed parameter method and the two-priors method resulted in similar, if not identical, sample size requirements for this simulation. Electing to incorporate uncertainty through design priors can correspond to adjustments in sample size.

CHAPTER FIVE

Conclusion and Future Work

We presented topics in Bayesian models with order constrained parameters. Order among parameters may occur from researchers' expectations, empirical evidence, or any number of other reasons. We examined logistic regression, proportional hazards, and a one-way analysis of variance, all with ordered components.

In Chapter Two, we focused on ordered differential response misclassification. We explored this phenomenon in a logistic regression setting and provided an adjustment using a system of conditional and marginal priors. We compared results from a naive model, a non-differential misclassification adjusted model, a differential misclassification adjusted model, and an ordered differential response misclassification adjusted model using the BFRSS mammography use data as introduced by Njai et al. (2011). In a simulation study, the proposed adjustment achieved "unbiased" estimates with smaller credible interval widths than the differential adjustment's.

Of primary interest for the future is ordered response misclassification in a logistic regression model with covariates having three or more levels. Also of interest is the ordered differential response misclassification generalized linear model regression setting with three or more outcomes. Ordered response misclassification in a logistic regression model with multiple covariates may also be an area of continued work.

In Chapter Three, we explored Bayesian survival models based on a covariate subject to ordered misclassification. We specified a Weibull baseline hazard in order to construct a fully parametric proportional hazards model. We first built a model based on an hypothetical fallible test example using only one misclassified covariate. In this case, the order occurred via a dependence on the misclassification parameters directly. We then built a more complex model based loosely on the example of racial/ethnic misclassification at hospitals (Gomez and Glaser, 2006). This model used one perfectly recorded binary covariate and a misclassified binary covariate with

misclassification rates dependent on the level of the perfectly recorded covariate. In the simulations, the ordered adjustment provided more accurate estimates of the regression parameter and hazard rate than the naive model. In most simulations considered, the ordered adjustment yielded more precise estimates than the unordered adjustment; however, the benefits of incorporating the ordered adjustment were more apparent in Chapter Two.

Applying this ordered misclassification adjustment to another model with a binary covariate could be both interesting and informative. Also of interest is ordered misclassification in covariates having three or more levels.

In Chapter Four, we presented "informative hypotheses" (Hoijtink, 2012) and performed a Bayesian sample size determination technique using the two-priors approach of Brutti et al. (2008). Through simulation, we concluded that applying the two-priors approach in the context of informative hypotheses led to similar empirical error rates than those obtained using a fixed-parameter simulation. However, the two-priors' results may lead to a different sample size requirement, based on a pre-defined maximum-tolerated empirical error rate criterion. In this simulation, we utilized Bayes factors to find preference between an informative hypothesis and its complement. We extended the work to two other dimensions, other design priors, and a different set of hypotheses.

We may examine other sample size determination criteria, such as coverage and credible interval widths. Additionally, we plan to extend the inequality constrained hypotheses to models unexplored in this context. This includes mixture models, models with misclassified data, and correlated binary data models.

APPENDICES

APPENDIX A

Ordered Differential Response Misclassification Chapter

A.1 General Beta

The mean and variance of the general beta are

$$\mu(Y) = \frac{\alpha c + \beta a}{\alpha + \beta}$$

and

$$V(Y) = \frac{\alpha\beta(c-a)^2}{(\alpha+\beta)^2(\alpha+\beta+1)},$$

respectively. Additionally, the mode is

$$M(Y) = \frac{(\alpha-1)c + (\beta-1)a}{\alpha+\beta-2}.$$

The general beta distribution can represent quantities that are not restricted to the support (0,1). For example, suppose we have $X \sim \text{Beta}(50, 50)$ and we have use for this distribution to be on a range of (4,9). Then we apply the transformation $Y = X(9-4)+4$ to achieve $Y \sim \text{Beta}_{[4,9]}(50, 50)$.

A.2 Conditional Means Priors

The conditional means priors (CMP) as described by Bedrick, Christensen, and Johnson (1996) are useful for incorporating prior information about regression parameters. Parameters in a logistic regression model are difficult to interpret and eliciting priors for these parameters is a challenging process. An alternative is to elicit information regarding expected responses from experts. In the logistic regression setting, this shifts the conversation from log-odds ratios to success probabilities at various covariate levels. The process of deriving CMPs involves eliciting information about these observable outcome probabilities from an expert and then inducing priors on the regression parameters.

A.3 Ordered Differential Response Misclassification Code

The following WinBUGS code performs an adjustment for ordered differential response misclassification in a logistic regression model.

```
model{
p[1] <- pi[1]*eta1 + (1 - pi[1])*(1 - theta1)
p[2] <- pi[2]*eta2 + (1 - pi[2])*(1 - theta2)
for(j in 1:k){
ystar[j] ~ dbin(p[j], n[j])
logit(pi[j]) <- beta0 + beta1*x[j]
}
ORCA <-exp(beta1)
# priors:
eta1 ~ dbeta(97,3)
eta2 ~ dbeta(97,3)
theta1 ~ dbeta(49,51)
theta2 <- pre.theta2*(1-theta1)+theta1
pre.theta2 ~ dbeta(24,76)
e_pi[1]~dbeta(19, 13)
e_pi[2]~dbeta(21, 12)
beta0 <- xtili[1,1]*(logit(e_pi[1])) + xtili[1,2]*(logit(e_pi[2]))
beta1 <- xtili[2,1]*(logit(e_pi[1])) + xtili[2,2]*(logit(e_pi[2]))
}
```

APPENDIX B

Ordered Covariate Misclassification Chapter

B.1 Constraint on Weibull Scale Parameter

The following describes work by Blair and Seaman (2014). Suppose we have information regarding the median survival time and wish to determine the value of α, the shape parameter of the Weibull distribution. For $T \sim \text{Weibull}(\alpha, \sigma)$, the median survival time is

$$t_m = \left(\frac{\ln(2)}{\sigma}\right)^{(1/\alpha)},$$

which yields

$$\alpha = \frac{\ln(\ln(2)/\sigma)}{\ln(t_m)}.$$

If $\sigma > \ln(2)$ and $t_m > 1$, α will be negative, which is not acceptable. In most cases, $t_m > 1$; thus, we must restrict $\sigma < \ln(2)$. In our problem, $\sigma = \lambda \exp(\beta)$ and it is not reasonable to restrict $\exp(\beta)$. Therefore, we restrict the nuisance parameter, λ. We require

$$\lambda \exp(\beta) < \ln(2).$$

Additionally, $\sigma = \lambda \exp(\beta)$ must be greater than zero and $\exp(\beta)$ is always positive. Thus, we bound λ by

$$0 < \lambda < \frac{\ln(2)}{k},$$

where k is an unusually large value of the hazard ratio and is chosen based on the level of conservativeness desired. Since we have no additional knowledge of λ, we assume that every value between 0 and $\ln(2)/k$ is equally probable. Thus,

$$\lambda \sim \text{ Unif}\left(0, \frac{\ln(2)}{k}\right).$$

B.2 Generating Survival Data

Bender, Agustin, and Blettner (2003) describe a method to generate survival data. We summarize their work regarding Weibull proportional hazards data gener-

ation. To begin, the survival function for the proportional hazards model is

$$S(t|x) = \exp(-H_0(t)e^{\beta' x}),$$

and the cumulative distribution function (CDF) is

$$F(t|x) = 1 - \exp(-H_0(t)e^{\beta' x}).$$

Let Y be a random variable with CDF $F(Y)$. Using the probability integral transform, $W = F(Y) \sim$ Uniform(0,1). Suppose a random variable, U, is the linear transformation $U = 1 - W$. Then $U \sim$ Uniform(0,1). In terms of the proportional hazards model, $W = F(t|x) \sim$ Uniform(0,1) and

$$U = 1 - W = \exp[-H_0(T)\exp(\beta' x)] \sim \text{ Uniform}(0, 1). \tag{B.1}$$

If $h_0(t) > 0$ for all t, the survival times for the proportional hazards model can be found by inverting Equation B.1 as

$$T = H_0^{-1}\left[-\log(U)\exp(-\beta' x)\right].$$

Additionally, the inverse of the cumulative hazard function for the Weibull distribution is

$$H_0^{-1}(t) = \left(\lambda^{-1}t\right)^{1/\nu}.$$

Using this information allows the survival times of a proportional hazards model with a Weibull baseline to be expressed as

$$T = \lambda^{-1/\nu}\left(-\log(U)\exp(-\beta' x)\right)^{1/\nu} = \left(\frac{-\log(U)}{\lambda\exp(\beta' x)}\right)^{1/\nu}.$$

The hazard function is

$$h(t|x) = \lambda\nu t^{\nu-1}\exp(\beta' x) = \lambda\exp(\beta' x)\nu t^{\nu-1}.$$

Thus, the survival times are $t_i \sim$ Weibull$(\lambda\exp(\beta' x_i), \nu)$.

APPENDIX C

Bayesian Sample Size Determination Chapter

We use the following code from van Rossum et al. (2013) to perform the sample size simulations for the five-dimensional model. The simulation code for other hypotheses and dimensions can be derived from this code.

```
MODEL{
#likelihood
  for(i in 1:full.num){
  y[i]~dnorm(mu[i],invsigma2)
  mu[i] <-mu1*d1[i] + mu2*d2[i] + mu3*d3[i] + mu4*d4[i] + mu5*d5[i]
}
#priors
 mu1~dnorm(0.0,0.001)
 mu2~dnorm(0.0,0.001)
 mu3~dnorm(0.0,0.001)
 mu4~dnorm(0.0,0.001)
 mu5~dnorm(0.0,0.001)
 invsigma2~dgamma(0.01,0.01)
 f1<-step(mu5-mu4)
 f2<-step(mu4-mu3)
 f3<-step(mu3-mu2)
 f4<-step(mu2-mu1)
 fit<-f1*f2*f3*f4
}
```

BIBLIOGRAPHY

Alwin, D. (2014), "Investigating response errors in surveys," *Sociological Methods & Research*, 43, 3–14.

Bang, H., Chiu, Y., Kaufman, J., Patel, M., Heiss, G., and Rose, K. (2013), "Bias correction methods for misclassified covariates in the Cox model: comparison of five correction methods by simulation and data analysis," *Journal of Statisical Theory and Practice*, 7, 381–400.

Bartholomew, D. J. (1961), "Ordered tests in the analysis of variance," *Biometrika*, 48, 325–332.

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

Blair, S. and Seaman, Jr., J. (2014), "Prior elicitation on parametric proportional hazards models through expert information on median survival times," Baylor University.

Brutti, P., De Santis, F., and Gubbiotti, S. (2008), "Robust Bayesian sample size determination in clinical trials," *Statistics in Medicine*, 27, 2290–2306.

Cancer Prevention Institute of California (2015), *Greater Bay Area Cancer Registry*.

Carroll, R., Rupert, D., Stefanski, L., and Crainiceanu, C. (2006), *Measurement error in nonlinear models: a modern perspective*, New York: Chapman & Hall/CRC, 2nd ed.

CDC (2014), *1999-2011 Incidence and Mortality Web-based Report*, United States Cancer Statistics Working Group. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute, Atlanta, available at: www.cdc.gov/uscs.

Dunson, D. (2005), "Bayesian semiparametric isotonic regression for count data," *Journal of the American Statistical Association*, 100, 618–627.

Dunson, D. and Neelon, B. (2003), "Bayesian inference on order-constrained parameters in generalized linear models," *Biometrics*, 59, 286–295.

Dunson, D. and Peddada, S. (2008), "Bayesian nonparametric inference on stochastic ordering," *Biometrika*, 95, 859–874.

Elton, R. A. and Duffy, S. W. (1983), "Correcting for the effect of misclassification bias in a case-control study using data from two different questionnaires," *Biometrics*, 39, 659–663.

FDA (2010), *Guidance for the use of Bayesian statistics in medical device clinical trials*.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005), "Statistical methods for eliciting probability distributions," *Journal of the American Statistical Association*, 100, 680–701.

Gelfand, A., Smith, A., and Lee, T. (1992), "Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling," *Journal of the American Statistical Association*, 87, 523–532.

Gelman, A. (2006), "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, 1, 515–534.

Gelman, A., Hill, J., and Yajima, M. (2012), "Why we (usually) don't have to worry about multiple comparisons," *Journal of Research on Educational Effectiveness*, 5, 189–211.

Gomez, S. and Glaser, S. (2006), "Misclassification of race/ethnicity in a population-based cancer registry (United States)," *Cancer Causes & Control*, 17, 771–781.

Gomez, S., Glaser, S., West, D., Satariano, W., and O'Connor, L. (2003), "Hospital policy and practice regarding the collection of data on race, ethnicity, and birthplace," *American Journal of Public Health*, 93, 1685–1688.

Gustafson, P. (2003), *Measurement error and misclassification in Statistics and Epidemiology: impacts and Bayesian adjustments*, Boca Raton: Chapman & Hall/CRC.

Hausman, J., Abrevaya, J., and Scott-Morton, F. (1998), "Misclassification of the dependent variable in a discrete-response setting," *Journal of Econometrics*, 87, 239–269.

Hoijtink, H. (2012), *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*, Chapman & Hall/CRC.

Host, K., Brugman, D., Travecchio, L., and Beem, A. L. (1998), "Students' perceptions of the moral atmosphere in secondary school and the relationship between moral competence and moral atmosphere," *Journal of Moral Education*, 27, 47–70.

Kinnersley, N. and Day, S. (2013), "Structured approach to the elicitation of expert beliefs for a Bayesian-designed clinical trial: a case study," *Pharmaceutical Statistics*, 12, 104–113.

Kirn, T. (2006), "New alcohol test appears fallible: several medical professionals who say they did not touch a drink are testing positive; losing their jobs," *Clinical Psychiatry News*, 34, 43.

Klugkist, I., Kato, B., and Hoijtink, H. (2005a), "Bayesian model selection using encompassing priors," *Statistica Neerlandica*, 59, 57–69.

Klugkist, I., Laudy, O., and Hoijtink, H. (2005b), "Inequality constrained analysis of variance: a Bayesian approach," *Psychological Methods*, 10, 477–493.

Kopylev, L. (2012), "Constrained parameters in applications: review of issues and approaches," *ISRN Biomathematics*, 2012, 6 pages.

Lyles, R., Tang, L., Superak, H., King, C., Celentano, D., Lo, Y., and Sobel, J. (2011), "Validation data-based adjustments for outcome misclassification in logistic regression: an illustration," *Epidemiology*, 22, 589–597.

Madi, M., Leonard, T., and Tsui, K.-W. (2000), "Bayes inference for treatment effects with uncertain order constraints," *Statistics & Probability Letters*, 49, 277–283.

McGlothlin, A., Stamey, J., and Seaman, Jr., J. (2008), "Binary regression with misclassified response and covariate subject to measurement error: a Bayesian approach," *Biometrical Journal*, 50, 123–134.

Morgan-Cox, M., Stamey, J., and Seaman, Jr., J. (2010), "Count regression models with a misclassified covariate: a Bayesian approach," Ph.D. thesis, Baylor University.

Morita, S., Thall, P., and Muller, P. (2008), "Determining the effective sample size of a parametric prior," *Biometrics*, 64, 595–602.

Mwalili, S. (2006), "Bayesian and frequentist approaches to correct for misclassification error with application to caries research," Ph.D. thesis, Katholieke Universiteit Leuven.

Njai, R., Siefel, P., Miller, J., and Liao, L. (2011), "Misclassification of survey responses and Black-White disparity in mammography use, Behavioral Risk Factor Surveillance System, 1995-2006," *Prev Chronic Dis*, 8, 1–7.

Prescott, G. J. and Garthwaite, P. H. (2002), "A simple Bayesian analysis of misclassified binary data with a validation substudy," *Biometrics*, 58, 454–458.

Rabe-Hesketh, S., Pickles, A., and Skrondal, A. (2003), "Correcting for covariate measurement error in Logistic Regression using nonparametric maximum likelihood estimation," *Statisical Modelling*, 3, 215–232.

Rauscher, G., Johnson, T., Cho, Y., and Walk, J. (2008), "Accuracy of self-reported cancer-screening histories: a meta-analysis," *Cancer Epidemiol Biomarkers Prev*, 17, 748–757.

Ren, D. and Stone, R. (2007), "A Bayesian adjustment for covariate misclassification with correlated binary outcome data," *Journal of Applied Statistics*, 34, 1019–1034.

Robert, C. P. (2007), *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, New York: Springer, 2nd ed.

Robertson, T., Wright, F., and Dykstra, R. (1988), *Order Restricted Statistical Inference*, John Wiley & Sons.

Royston, P. and Parmar, M. (2002), "Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects," *Statistics in Medicine*, 21, 2175–2197.

Selen, J. (1986), "Adjusting for errors in classification and measurement in the analysis of partly and purely categorical data," *Journal of the American Statisticial Association*, 81, 75–81.

Silvapulle, M. and Sen, P. (2004), *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*, London: John Wiley & Sons, Inc.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Chichester: John Wiley & Sons, 1st ed.

Spiegelman, D., Rosner, B., and Logan, R. (2000), "Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs," *Journal of the American Statisticial Association*, 95, 51–61.

Stamey, J., Seaman, Jr., J., and Young, D. (2005), "Bayesian sample-size determination for inference on two Binomial populations with no gold standard Classifier," *Statistics in Medicine*, 24, 2963–2976.

Stoline, M. (1981), "The status of multiple comparisons: simultaneous estimation of all pairwise Comparisons in One-Way ANOVA Designs," *The American Statistician*, 35, 134–141.

Thomas, D., Stram, D., and Dwyer, J. (1993), "Exposure measurement error: influence on exposure-disease relationships and methods of correction," *Annual Reviwes of Public Health*, 14, 69–93.

Thurigen, D. (2000), "Measurement error correction using validation data: a review of methods and their applicability in case-control studies," *Statistical Methods in Medical Research*, 9, 447–474.

Van de Schoot, R., Velden, R., Van der Boom, J., and Brugman, D. (2010), "Can at-risk young adolescents be popular and antisocial? Sociometric status groups, anti-social behaviour, gender and ethnic background," *Journal of Adolescence*, 1, 1–10.

Van Rossum, M., Van de Schoot, R., and Hoijtink, H. (2013), "Is the hypothesis correct or is it not: Bayesian evaluation of one informative hypothesis for ANOVA," *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9, 13–22.

Wang, C. and Song, X. (2013), "Expected estimating equations via EM for proportional hazards regression with covariate misclassifiation," *Biostatistics*, 14, 351–365.

Wang, F. and Gelfand, A. (2002), "A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models," *Statistical Science*, 17, 193–208.

Yi, G., Ma, Y., Spiegleman, D., and Carroll, R. (2014), "Functional and structural methods with mixed measurement error and misclassification in covariates," *Journal of the American Statistical Association*.

Zucker, D. and Spiegelman, D. (2004), "Inference for the proportional hazards model with misclassified discrete-valued covariates," *Biometrics*, 60, 324–34.