ABSTRACT

Topics in Odds Ratio Estimation in the Case-Control Studies
and the Bioequivalence Testing in the Crossover Studies

Denka G. Markova, Ph.D.

Chairperson: Dean M. Young, Ph.D.

The double-sampling paradigm, which has become an important part of the epidemiological designs, includes two stages. First, individuals are classified into groups by disease and exposure levels using a fallible test, and second, some individuals are classified into a subset using a "gold standard" test. The parameter of interest in our study is the odds ratio as an association between disease level and exposure level. Here we compare four confidence intervals for the odds ratio under the assumption of differential or non-differential misclassification. More specifically, we compare the coverage and interval widths of the Wald, score, profile likelihood, and approximate integrated likelihood intervals with different specificity and sensitivity values, as well as different sample sizes and odds ratios for the case-control clinical studies. Our investigations implies the consistent superiority of the approximate integrated confidence interval.

Also, we eliminate the effect of several parameters on a bioequivalence testing procedure that plays an important role in the development of generic drugs. The current FDA criteria is not flexible with respect to highly variable drugs, and this characteristic has caused many good drugs to be rejected. Most often in the literature, we find studies examining the sample size or the within-subject variability as the main factors affecting the outcome of a bioequivalence test. Frequently,

pharmaceutical companies have tried to convince the FDA that their product would meet the bioequivalence criteria just by increasing the sample size. Here we examine the effect of the between-subject variability as well as the effect of the mean ratio difference between the test and reference formulations. We use a Monte Carlo simulation to draw conclusions based on the importance of these two sources of variability and to show that simply increasing the sample size is insufficient to meet the bioequivalence criteria.

Topics in Odds Ratio Estimation in the Case-Control Studies
and the Bioequivalence Testing in the Crossover Studies

by

Denka G. Markova, 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

_____

Dean M. Young, Ph.D., Chairperson

_____

James D. Stamey, Ph.D.

_____

Jack D. Tubbs, Ph.D.

_____

Thomas L. Bratcher, Ph.D.

_____

James Kendrick, Ph.D.

Accepted by the Graduate School
December 2011

_____

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

To the casual observer, a doctoral dissertation may appear to be solitary work. However, to complete a project of this magnitude requires a network of support, and I am indebted to more people than I can list here. Nevertheless, I would like to give credit to the most influential of them.

My Family

First and foremost, I want to thank my parents, Yordanka and Grudi Markovi. Without them I would not be the person I am today and for this I am most grateful. Lately, looking back at my childhood, I have realized that I have been truly blessed with amazing parents. Mom and Dad, thank you for your constant support, belief, and encouragement. Thank you for standing behind me, for giving me the opportunity to come to the US, and for working so hard to realize every dream my sister or I had.

I also want to thank my little sister, Nevena. Sis, you have been such a support for me. Thank you for listening and always wanting to defend me. *Obicham te, milichko!*

Next, I want to thank my paternal grandparents, Denka and Nedelcho, whose deep appreciation of education has long inspired my own. You never let me settle for less than great. Grandma, your stories have cheered and encouraged me to keep going. Also, special thanks to my maternal grandma, Zaprianka, who always gave me unconditional love. Thanks to my little cousin, Bozhidar. Your constant belief that I can do no wrong has encouraged me to try to be the best. I also want to thank my Aunt Rosi, who loves me like her own daughter. I love you and your baby boy.

My Mentors

First, I would like to give special thanks and appreciation to Dr. Dean Young for persevering with me as my advisor throughout this time. Thank you for trusting me and always believing that I am capable. Your kind and encouraging words were greatly appreciated. I do not believe there could have been a better mentor for me. Thank you for all your time spent on correcting my dissertation and for teaching me how to write better, for always joking with me and making me laugh. I also want to thank your lovely wife, Mrs. Young. Your help in the final revisions of this dissertation were greatly appreciated as well.

Next, I want to thank Dr. James Stamey for serving as my second advisor. Thank you for always being available and finding time to talk to me even at 10 pm at night. Thank you for checking up on me, for helping me, and reading my numerous drafts. I am also appreciative of you and your wife for letting me watch your boys from time to time. Often those were my most fun days of the week.

I am very grateful to all the professors in the statistics department. You all have been like a family to me and I appreciate all of your help, advice, and encouragement in the past five years. Thank you for taking the time to get to know us and for showing us how to improve. Dr. Tubbs, thank you your support and readiness to listen and for helping me grow as a statistician. Dr. Bratcher, you always acted tough, but I know it was with the best intentions. Dr. Johnston, thank you for your willingness to explain any question I had, especially when I was consulting. Karen, I also want to thank you for checking up on me and for taking me to lunch and listening to my stories. Dr. Kendrick, thank you for serving on my committee. You and your wife have been so kind to me, and I greatly appreciate your help.

Next, I would like to show my appreciation for Brandi Greer. You are such a great friend, but you are also a mentor. You have influenced not only school but life in general. If it was not for you, I would have been lost in the dissertation process. One can only learn from you how to be a better human being! You and your adorable babies have a very special place in my heart!

I would also like to give special thanks to Dr. Sang Chung and the whole team at the Food and Drug Administration, Division of Clinical Pharmacology II. Thank you all for giving me the opportunity to come and work with you, for treating me as part of the team, and for showing me what a great working environment is.

Last but not least, I want to thank my advisors in my undergraduate career. Dr. Worth, thank you for showing me that research could be really fun, for opening my eyes to the possibility of a graduate degree, and for your encouragement even throughout my graduate years. Dr. Durand, thank you for being the greatest and most caring professor I have ever met. Your words that graduate school is "not about how smart you are, but about how persistent you are" stayed with me and helped me realize that the PhD is really about hard work and never giving up. Dr. Eoff, thank you for introducing Linda and me to Baylor University and the statistics degree. Without all of your help, I would never have made it so far.

My Friends

The first friend I would like to thank is Linda Njoh for making me apply to Baylor, for convincing me that statistics is the right field for me, and for not letting me give up. Thank you for listening to me and my problems, for the endless all-nighters, for spending time with me during much-needed breaks, for the never-ending moral support, for your ability to see the light at the end of the tunnel, for always, always believing in me, and for just being you. I could not have asked for a better friend than you!

Also, I would like to thank my Bulgarian friends for their unconditional love and support throughout all my years away from them and home. Zori, thank you for helping me come to the USA. Without your advise and help I would have never made it here. Elitsa, thank you for checking up on me and never giving up of our friendship. Moni, Eli, Vanya, thank you all for being such good friends and not letting the distance change that.

Next, I want to thank the friends I met at Baylor. Austin Hand, thank you for being such a great friend and a classmate. Thank you for your willingness to listen, for helping me with my career, and for your trust in me as a statistician and person. Gustavo Chavarria, thank you for your moral support and encouragement, for always listening to my problems, and for opening your home for me when I needed it. Nelson, Otsmar, and Ivana, thank you all for your support throughout this process. I am so excited we will be graduating together.

MaryAnn Morgan-Cox, I am so thankful to have you and your family in my life. You have been such a great example of strength and persistence. Thank you for always encouraging me and helping me in every aspect of my life.

Next, I would like to show my greatest appreciation for the Bulgarian family I met in Waco, Roujina and Couman. You have been such a support for me during the last months of this dissertation process. Thank you for opening your home, you lives, and your hearts for me. Thank you for your encouragement and kind words. You are such a great couple, and you have been an amazing example of what a family can achieve with love and hard work.

Last but not least, I would like to thank my wonderful coworkers. You all have been so supportive in this final process. Thank you for all your encouraging words, lucky charms, and prayers. Thank you for making my first job after graduate school such an amazing experience.

DEDICATION

To My Grandma Denka

A woman endued with courage, strength, and the will to fight. Her endless encouragement and generosity have inspired me throughout my whole life. Her unwavering love and friendship will always mean the world to me.

To My Family

I am fortunate to have been loved by all of you.

CHAPTER ONE

Introduction

This dissertation consists of two topics. Therefore, in this chapter we separately introduce the concepts and background information for each topic.

1.1 Introduction to Confidence Intervals for the Odds Ratio

1.1.1 Overview

Essentially, every statistical inference problem begins with an unknown state of nature represented by a parameter of interest value, Basu (1977). We often plan and perform a statistical experiment and generate sample observations to obtain further information about the parameter of interest through a careful analysis of the data. In traditional statistics, we define a probability measure of the events of interest that is dependent on the parameter of interest. However, statistical models where the probability measure depends on only one parameter are rare. Typically, the model is defined by the parameter of interest and an additional set of unknown parameters known as nuisance parameters. Furthermore, to access the information on the parameter of interest, we must account for or eliminate the nuisance parameters.

The analysis of binary data, i.e., data with underlying binomial distribution, is important because such data is collected from a wide range of applications, including medical diagnosis, survey analysis, and political voting, Boese (2005). One area of the medical field where one finds such data is epidemiology, where the disease status and exposure level of a patient treated for a certain condition are collected. The parameter of interest is sometimes the odds ratio that represents a measure of association between disease level and true exposure level. However, when we collect

data, we can make misclassification errors. Such errors may cause incorrect counts that affect our odds ratio inferences. Bross (1954) has shown that estimation of the population proportion parameter based on samples subject to misclassification is biased. To correct for this bias, we must adjust our parameter estimation. A procedure for obtaining such an estimator using binomial data with misclassification, called a double-sampling plan, has been developed by Tenenbein (1970). The procedure is used to derive nearly unbiased estimates and appropriate confidence intervals (CIs) for the odds ratio defined above. Nevertheless, the double-sampling paradigm produces one or more nuisance parameters that must be eliminated or accounted for.

One of the most important problems in statistical inference is, "How can we eliminate the effect of the nuisance parameters?" Basu (1977) has classified the answer to this question in roughly ten overlapping categories. This study considers the use of a maximum-likelihood-estimation approach to derive Wald and Score CIs and compare them to the profile likelihood and approximate integrated likelihood CIs derived using pseudo-likelihood estimation. Hence, we compare four interval estimation methods for the odds ratio for measuring the association between disease and exposure levels under a double-sampling procedure for binomial data with two types of misclassification.

The remainder of this section is organized as follows. In Subsection 1.1.2 we explain the design of the case-control study that is examined in this paper. In Subsection 1.1.3 we detail an overview of the double-sampling scheme defined by Tenenbein (1970). Because we are dealing with misclassified data, we approach the problem assuming differential or non-differential misclassification. The definitions of the two types of misclassification are presented in Subsection 1.1.4. Finally, in Subsection 1.1.5 we explain the confidence interval methods we utilize in this dissertation.

1.1.2 Types of Studies

The methods and analyses presented here are often applied in epidemiology, which is the study of factors that affect the health and illness of populations. Epidemiological research aims to obtain the prevalence of a disease, to examine the causes, and even to decide if a given exposure can cause or prevent the disease. Many different types of study designs can address these questions. The two most common studies are case-control and cohort.

Case-control studies are observational epidemiological studies that we use to identify factors that contribute to a medical condition by comparing a group of patients who have the condition, known as "cases," to a group of patients who do not have the condition, known as "controls." We select the subjects based on their disease status, and we treat cases and controls as independent groups. The studies examine potential exposures that both groups have encountered over time. They are usually cost effective and fast, but very sensitive to bias.

The cohort is another type of epidemiological study that selects subjects based on their exposure status. Therefore, at the beginning of the study the participants are disease free. The cohort group of individuals is observed over time and then tested for the disease. However, in situations where the disease has a very low probability of occurring, the use of a cohort is not appropriate because it will not guarantee that diseased cases will be in the group. Cohort studies tend to be more costly and time consuming and have a greater chance of losing subjects.

In the following chapters we present analyses for the case-control type of study and address some of the difficulties that occur.

1.1.3 Double-Sampling Procedure

The double-sampling procedure has been increasingly incorporated in the epidemiological design studies. Bross (1954) has shown that, under misclassification,

the sample estimate of the binomial proportion is biased and the bias could be substantial. Therefore, we need to obtain sufficient information about the misclassification probabilities to adjust for the bias. Tenenbein (1970) has proposed the use of the double-sampling plan to obtain an unbiased estimator for the population proportion of binary data with misclassification. He has suggested that we compare the results from testing the same group of sampling units by two or more measuring devices. Such a scheme requires the use of a large data set and a smaller data subset.

Suppose we conduct a test or procedure that allows us to obtain a disease status on a large sample of participants. However, such an instrument, although fast, inexpensive, and perhaps noninvasive, can be fallible. Hence, the counts we observe will have errors due to misclassification, thus leading to biased estimators of the parameter of interest, the odds ratio. To calculate the misclassification rate and account for the bias, we obtain a subsample of the original data set and use not only the fallible test, but also a second, inerrant test, referred to as the "gold standard" test. This "gold standard" procedure is often very expensive, invasive, and time consuming. Hence, the sample that we test with both devices is much smaller than the original fallible sample. The fallible sample is called the main or incomplete study, whereas the infallible sample is called a validation or complete study. Dahm et al. (1995) have investigated the value of additional fallible classification for improving estimates of the odds ratio in case-control and cohort studies. Also, Karunaratne (1991) has examined different approaches of estimation in both types of studies under differential and non-differential misclassification. In this dissertation we utilize a double-sampling procedure to assess the misclassification rate and develop several confidence intervals that account for nuisance parameters.

1.1.4 Types of Misclassification

When analyzing misclassified data, a researcher focuses on whether the misclassification rates between the fallible and infallible tests depend on the true disease status of the patient, Tenenbein (1970). Often misclassification is linked to the levels of specificity and sensitivity for the study of interest. Hence, high levels of specificity and sensitivity indicate low misclassification rates.

If the rate is independent of the disease outcome level, then such misclassification is referred to as nondifferential. Therefore, for nondifferential misclassification we can assume that the specificity and sensitivity are independent of the disease status, i.e., they are the same for both diseased (cases) and non-diseased (controls) individuals (see Karunaratne (1991)).

When the misclassification rate is not independent of the disease level, then we define this as differential misclassification. In such instances the sensitivity and specificity between the diseased and non-diseased groups are assumed to be different. This case often results when information for the validation study is obtained from a previous study and not directly from the current study population.

In this dissertation we conduct simulations to compare confidence intervals for the odds ratio using double-sampling for case-control studies under both differential and nondifferential misclassification. We denote the case-control study with differential misclassification as CCDIFF and the nondifferential as CCNDIFF. Thus, we compare four different confidence intervals for each of the previous situations as well as two levels of misclassification: low and high as defined by Dahm et al. (1995).

1.1.5 Confidence Interval Methods

In statistics we infer a population parameter based on a sample of values. Let $\mathbf{x} = (x_1, x_2, ..., x_n)'$ be a random vector of n iid observations sampled from the probability density $f(\mathbf{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)'$ is a parameter vector of dimension

p contained in the parameter space $\boldsymbol{\Theta}$. The likelihood function, $L(\boldsymbol{\theta}|\mathbf{x}) = f(\boldsymbol{\theta}|\mathbf{x})$, is viewed as a function of $\boldsymbol{\theta}$ given the observations $\mathbf{x}$, and by the likelihood principle, contains all the information concerning the experiment of interest, Bjornstad (1996). Thus, inferences about $\boldsymbol{\theta}$ depend on the random variable only through the likelihood function, see Berger et al. (1999). Although $\boldsymbol{\theta}$ is a vector of parameters, we are often interested in only one parameter or a subset of those parameters. Hence, we can partition $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\eta}')'$, where $\boldsymbol{\psi}$ is a parameter of interest and $\boldsymbol{\eta}$ is a subvector of nuisance parameters.

We derive four CIs to estimate the parameter of interest, the odds ratio for the proportion of diseased versus non-diseased under misclassification using double-sampling. We compare the intervals using the criteria of coverage and interval width. Specifically, we consider Wald and score CIs as well as the profile likelihood and the approximate integrated likelihood CIs.

1.2 The Bioequivalence Testing Procedure

1.2.1 Overview

Pharmaceutical manufacturing has become one of the most vital industries in the world. An integral role in new drug development is the bioequivalence (BE) study. The United States Food and Drug Administration (FDA) defines bioequivalence as "the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study" (FDA (2003)). Essentially, we call two drug products bioequivalent when they are expected to perform the same and can be interchanged.

The main reason for the popularity of BE studies is the approval of new generic drugs. When a reference-listed drug on the market might be too expensive for the

general population, companies often attempt to develop a generic version that is more accessible. In such instances, the FDA requests a BE study to determine the rate and extent of absorption of each therapeutic moiety for the generic and the reference products. BE studies are also useful when we have new formulations for old drug products that are proven to work or if we are testing a new dosage or including new inactive ingredients.

Because the BE studies do not depend on the actual outcome of the trial but only on the rate and extent of availability of the tested product, they are generally conducted in a healthy population. Male and female adults are given the drug under standardized conditions and are monitored throughout the length of the trial. In some cases, however, one might be forced to use diseased patient groups for safety reasons, Davit, Nwakama, Buehler, Conner, Haidar, Patel, Yang, Yu, and Woodcock (2009). Also, most BE studies are conducted on the highest strength of a drug.

In this section we define the components considered when we are testing the BE of two pharmaceutical products. In Subsection 1.2.2 we define the parameters of interest investigated and the study design utilized in the simulation study we conduct. In Subsection 1.2.3 we discuss the statistical and mathematical criteria for passing a BE test. We explain how we assess our simulation results and reach our conclusions. In Subsection 1.2.4, we list the sources of variability we control in the simulation study presented in this paper.

1.2.2 Parameters of Interest and Study Design

BE assessment depends on the bioavailability of the administered drugs in the participants' systems. Bioavailability is defined by the FDA (2003) as "the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by

measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action." Hence, when we conduct BE studies, we focus on two main pharmacokinetic parameters: the area under the curve of the drug plasma concentration versus time (AUC) and the maximum concentration (C_{max}). In our study, however, we focus on only the AUC because both criteria have very similar characteristics and the conclusions drawn for one can be expanded to the other.

BE studies are performed using two types of designs according to the drugs to be tested – crossover or parallel designs. The most common approach is the crossover design with two sequences and two periods. This design is referred to as a 2×2 crossover design and is the type of design we utilize in our study. In Chapter 4 we discuss the definitions and reasons behind our choice of design as well as the choice of the parameter to be investigated in detail.

1.2.3 Criteria for Declaring BE

Recall the definition of bioequivalence from Section 1.2.1. The "no significant difference" portion of the definition is assessed by examining a confidence interval (CI) from the geometric mean test/reference ratios for both pharmacokinetic parameters AUC and C_{max}. We want the 90% confidence interval of the geometric mean test versus reference ratio to fall within preset bioequivalence limits of $80\% - 125\%$ (FDA (2003)). The BE limits are based on medical judgment and FDA experience that a difference of 20% or less in drug exposure is not clinically significant for most drugs, Haidar et al. (2008a). An investigation of the properties of the two parameters and the reasons for choosing the particular BE limits above are discussed in Chapter 4.

1.2.4 Sources of Variability

In every clinical trial several sources of variability can affect the outcome of a statistical decision procedure. In this paper we investigate some of the well-known sources and, thus, expand on previous research. We consider within- and between-subject variabilities, sample size, and the value of the mean ratio difference in order to better understand their effects on a BE trial. Several of those parameters have been discussed in the literature; however, in our study we expand on them and consider the between-subject variability as an added layer of complexity in our simulation. Further reasoning and discussion on our choices of parameters can be found in Chapter 4 of this dissertation.

1.3 Dissertation Organization

In this dissertation we present two distinct topics: the first concerns the interval estimation of the odds ratio parameter in the case-control studies, and the second is an investigation of the sources of variability in the bioequivalence study with a 2×2 crossover design.

In the first part of this dissertation, we derive four approximate likelihood-related confidence intervals for the odds ratio when using double-sampling procedures. The underlying distribution of the data is binomial but is subject to misclassification. We consider case-control studies with two kinds of misclassification errors – differential and nondifferential. In Chapter 2 we focus on case-control studies subject to differential misclassification. We define the model and derive the maximum likelihood estimators as well as the restricted maximum likelihood estimators. We then use the MLEs and/or RMLEs to derive and estimate four competing confidence intervals for the odds ratio of interest. The intervals based on the maximum likelihood procedures we consider are Wald and score, and the intervals which are based on the pseudo-likelihood procedures are profile likelihood and approximate

integrated likelihood. In the last part of the chapter, we conduct a simulation study to compare the discussed intervals using two levels of misclassification – low and high.

In Chapter 3 we focus on the nondifferential misclassification in case-control studies. We follow essentially the same outline as in Chapter 2. That is, we introduce the model and explain the difference between the derivations in this chapter compared to the differential procedures. We derive and evaluate the four proposed confidence intervals. Then, we design and perform a simulation study to compare the CI competing under low and high misclassification errors.

In Chapter 4 we investigate a new topic concerning the bioequivalence testing in the pharmaceutical industry. We explain what BE studies are and why they are important in the drug development process. Then we discuss the reasons for creating an extensive simulation investigation and describe the Monte Carlo study design and methods used in this chapter. Last, we present the results of our Monte Carlo simulation study and our conclusions.

CHAPTER TWO

Confidence Intervals for the Odds Ratio in Case-Control Studies with Differential Misclassification

2.1 Introduction

Case-control trials are studies in which we compare one group of people with a medical condition (cases) to a group of people without the medical condition (controls). We often conduct these studies to investigate the association between the occurrence of a disease and a specific binary risk factor (exposure) of primary interest, Forbes and Santner (1995). The odds ratio is one the most popular relative measures of the exposure-disease relation, and we can estimate the odds ratio from retrospective data, Cornfield (1951). We often encounter problems in the form of measurement error because the epidemiological studies are usually observational rather than experimental and because the variables under study are generally subjective and must be ascertained by a subject's self report. If we ignore the misclassification errors and perform a naive analysis on the cases where discrepancies exist between apparent and actual exposure status, we may obtain highly biased estimates, as first shown by Bross (1954).

A large epidemiological literature provides methods for correcting data misclassification. For instance, Walter and Irwig (1988) have discussed the deleterious effects of ignoring misclassification, and Thomas, Stram, and Dwyer (1993) have reviewed correction methods. Gustafson, Le, and Saskin (2001) have studied a Bayesian scenario where reasonable guesses are available for the misclassification probabilities and suggest methods to adjust odds ratio estimators.

One of the first methods proposed for estimating a population proportion of exposure was the double sampling procedure introduced by Tenenbein (1970), who

has derived the maximum likelihood estimators for a binomial parameter when both false-negative and false-positive error rates occur. Other authors have considered the one proportion model with only one type of misclassification. For instance, Lie, Heuch, and Irgens (1994) have corrected for false-negative counts using only multiple fallible classifiers and have also derived the MLE.

In this paper we investigate the estimation of a confidence interval (CI) for the odds ratio parameter when we have binomial data subject to misclassification. This topic is a broad and heavily discussed subject in epidemiology, as well as in statistics. Until recently, many considered the Wald interval that uses the MLE of the binomial parameter to be the standard method for interval estimation. However, many authors now recognize the erratic coverage probability of the Wald intervals. In particular, Brown, Cai, and DasGupta (2001) have investigated in detail the unsatisfactory coverage properties of the Wald interval. Also, Boese, Young, and Stamey (2006) have given five asymptotic confidence intervals in the false-positive misclassification model. The interval estimators examined in Boese, Young, and Stamey (2006) have been based on likelihood and pseudo-likelihood methods. Also, Paul and Thedchanamoorthy (1997) have examined the likelihood-based confidence intervals for the common odds ratio, and Lee and Byun (2008) have derived Bayesian credible sets for the case of binomial data with false-positive misclassification and double sampling.

In this chapter we derive four confidence intervals for the odds ratio parameter in a case control study using double sampling. Two of the intervals, the Wald and score intervals, are likelihood based, and two intervals, the profile likelihood and approximate integrated likelihood, are pseudo-likelihood based. We consider differential misclassification for the two samples of interest (complete and fallible); that is, the specificity and sensitivity of those samples are not assumed to be equal. The double-sampling procedure allows us to estimate nuisance parameters, then employ

the four interval estimation methods mentioned above that eliminate or account for the nuisance parameters, and obtain a CI for the odds ratio. A point estimation method has been presented by Karunaratne (1991), and a real data example using his results has been given by Dahm, Gail, Rosenberg, and Pee (1995).

We have organized the remainder of the chapter as follows. First, in Section 2.2 we present the model and introduce the notation. Also, we discuss the double-sampling procedure in more detail. Then, in Section 2.3 we derive the maximum likelihood estimators of the parameter of interest and the model nuisance parameters. These results are used to derive a Wald CI. To derive the other three confidence intervals, we define the restricted maximum likelihood estimators (RMLEs) of the nuisance parameters and describe this method in detail in Section 2.4. Further, in Section 2.5 we derive the observed information matrix. We then derive the four confidence intervals for the odds ratio: the Wald, score, profile likelihood, and approximate integrated likelihood intervals. In Section 2.6 we present a simulation study for the efficacy of the four intervals by comparing them on the basis of width and coverage and examining the effects of sample sizes and of probabilities for disease outcome and misclassification on the intervals. Last, in Section 2.7 we comment on the simulation results.

2.2 The Model

We assume the underlying distribution for our population to be binomial because subjects either have or do not have a certain medical condition. Let the binary random variable D represent the disease status ($D = 1$ for diseased, $D = 0$ for non-diseased) for each individual in the population. Then, to form a case-control study, we independently chose a number of individuals from the diseased group and the control group. The number of people selected for each group is determined by the researcher. Hence, this type of study is extremely useful when the rate of a disease

is low in the study population and, therefore, using a random sampling procedure might not provide us with an adequate number of observations for each group.

Now, to facilitate the use of the double-sampling design, we assume that we have two testing procedures that can determine the disease status of each participant. In the first stage, we classify all of the individuals in the study using only a fallible procedure. Let Z denote the fallible exposure level, where $Z = 1$ represents an individual that is diagnosed with the disease by the fallible test and $Z = 0$ represents an individual with no disease according to the fallible procedure. Such a procedure is usually fast, inexpensive, and noninvasive and is performed on a relatively large sample. In the second stage of the double-sampling scheme, we use a smaller sub-sample of the fallible test sample and perform a second "gold standard" procedure on this sub-sample P. Let X denote the exposure level measured by the infallible ("gold standard") test, where $X = 1$ indicates a disease and $X = 0$ indicates someone who is free of a disease. This test is often expensive, time consuming, and extremely invasive. Therefore, we perform it on only a small group of individuals. Because we know that the "gold standard" test is absolutely accurate, we can use the parameter estimates from this sub-sample to correct for the estimator biases, Tenenbein (1970).

Now, let the subscripts i, j, and k represent the fallible test outcome, $Z = i$, the "gold standard" outcome, $X = j$, and the true disease status, $D = k$. We denote the complete data cell count where both tests are performed by V_{ijk} and the incomplete data cell count by W_{ik}. Also, we use M_k to indicate the sample size from the disease group $D = k$ for the complete data and N_k for the incomplete data. Notice that $M_k + N_k$ gives us the total sampled from each group of diseased or non-diseased in the first stage of the double-sampling procedure. For a visual representation, please refer to Table 2.1. Also, in the complete sample, let T_k represent the number from each disease group for which the "gold standard" test gave a positive result, $X = 1$. The number of individuals tested with both fallible and "gold standard"

procedures, M_k, is predetermined by the researcher and is a sub-sample of the total number of people participating in the study, $M_k + N_k$ for each $k = 0, 1$.

Table 2.1: Counts for a study with misclassified exposure data

	Validation study (complete)				Main study (incomplete)	
Fallible	Cases (D=1)		Controls (D=0)		Cases (D=1)	Controls (D=0)
(Z)	X=1	X=0	X=1	X=0		
Z=1	V_{111}	V_{101}	V_{110}	V_{100}	W_{11}	W_{10}
Z=0	V_{011}	V_{001}	V_{010}	V_{000}	W_{01}	W_{00}
	T_1	$M_1 - T_1$	T_0	$M_0 - T_0$		
	M_1		M_0		N_1	N_0

Next, we define the probabilities we use in this study. For the complete (validation, infallible) study, we denote the probability of exposure, that is, the probability where $X = 1$ for the k^{th} group ($D = k$) by

$$\pi_k = Pr(X = 1|D = k), \tag{2.1}$$

where $k = 0, 1$ for controls and cases, respectively. Also, after introducing the fallible test we define the sensitivity (true positive rate), S_k, as the probability that an individual tests positive under the fallible test ($Z = 1$), given that an individual has a positive result on the "gold standard" procedure ($X = 1$) for each case or control group. Thus,

$$S_k = Pr(Z = 1|X = 1, D = k), \tag{2.2}$$

for $k = 0$, 1. We denote the probability that an individual does not have the disease according to the fallible test ($Z = 0$) given that the individual tested negative by the "gold standard" test ($X = 0$) for each cases and controls group by

$$C_k = Pr(Z = 0|X = 0, D = k), \tag{2.3}$$

where $k = 0$ or 1. We remark that we have distinct values for the specificity and sensitivity for each of the cases or controls group, which we refer to as differential

misclassification. We can usually reasonably assume this observation when working with case-control studies and will derive the CIs of interest in this chapter based on this assumption.

Based on the equations (2.1) - (2.3), and derivations in Prescott and Garthwaite (2002), we induce the following distributions on the observable counts for the complete study. We assume

$$T_k = V_{01k} + V_{11k} \sim Bin(M_k, \pi_k), \tag{2.4}$$

$$V_{11k}|T_k \sim Bin(T_k, S_k), \tag{2.5}$$

and

$$V_{00k}|T_k \sim Bin(M_k - T_k, C_k), \tag{2.6}$$

where $k = 0$ or 1 indicates actual disease status of the person in the study ("0" is diseased, and "1" is non-diseased). Also, for the incomplete study, we have

$$W_{1k} \sim Bin(N_k, \pi_k S_k + (1 - \pi_k)(1 - C_k)),$$

where $k = 0$, 1. Now, the parameter of interest, the odds ratio ψ, is

$$\psi \equiv \frac{\pi_1(1 - \pi_0)}{\pi_0(1 - \pi_1)}. \tag{2.7}$$

One can find many real-world examples for the case-control design. Greenland (1988) has provided an example from a case-control study on sudden infant death syndrome (SIDS) first used in Drews, Kraus, and Greenland (1990) and also described in Morrissey and Spiegelman (1999). The study examined the relationship between maternal use of antibiotics during pregnancy and incidence of SIDS. Drug use was measured by interview (Z) and validated by medical record (X). Another example has been given by Dahm et al. (1995), where the effect of ambient exposure to radon in the homes of lung cancer-risk patients was investigated. For this study samples from the population of cases (with lung cancer) were drawn independently

from the population of controls (no lung cancer). Then, a "gold standard" test (X) was administered that involved one year of continuous monitoring on radon levels, whereas a fallible test (Z) was assessed by just taking the measurements over one week only, see Karunaratne (1991).

2.3 Maximum Likelihood Estimators

To estimate the CIs for ψ, the parameter of interest in this study, we first derive the MLEs of all the nuisance parameters, $\boldsymbol{\eta}$. Let $\boldsymbol{d} \equiv (W_{00}, W_{01}, W_{11}, W_{10}, V_{111}, V_{001}, V_{110}, V_{000}, V_{011}, V_{101}, V_{010}, V_{100})'$ denote the observed data counts for the full data (both main and validation studies), and let $\ell = \ell(\pi_1, \pi_0, C_1, C_0, S_1, S_0|\boldsymbol{d})$ represent the log likelihood function. Then,

$$\begin{aligned}
\ell \propto\ & W_{11}\ln[\pi_1 S_1 + (1-\pi_1)(1-C_1)] + W_{10}\ln[\pi_0 S_0 + (1-\pi_0)(1-C_0)] \\
& + (N_1 - W_{11})\ln[1 - \pi_1 S_1 - (1-\pi_1)(1-C_1)] + T_1\ln(\pi_1) \\
& + (N_0 - W_{10})\ln[1 - \pi_0 S_0 - (1-\pi_0)(1-C_0)] + T_0\ln(\pi_0) \\
& + (M_1 - T_1)\ln(1-\pi_1) + (M_0 - T_0)\ln(1-\pi_0) \\
& + V_{111}\ln(S_1) + V_{110}\ln(S_0) + V_{001}\ln(C_1) + V_{000}\ln(C_0) \\
& + (T_1 - V_{111})\ln(1-S_1) + (T_0 - V_{110})\ln(1-S_0) \\
& + [(M_1 - T_1) - V_{001}]\ln(1-C_1) \\
& + [(M_0 - T_0) - V_{000}]\ln(1-C_0).
\end{aligned}$$

Because we are interested in estimating the odds ratio, from (2.7) we have

$$\pi_0 = \frac{\pi_1}{\pi_1 - \pi_1\psi + \psi},$$

and then substituting into the log-likelihood function (2.3), we get

$$
\begin{aligned}
\ell_\psi \propto\ & (M_0 - T_0)\ln\left[1 - \frac{\pi_1}{\psi + \pi_1 - \psi\pi_1}\right] + W_{10}\ln\left[\frac{(1-C_0)(1-\pi_1)\psi + \pi_1 S_0}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + T_0 \ln\left[\frac{\pi_1}{\psi + \pi_1 - \psi\pi_1}\right] + (N_0 - W_{10})\ln\left[\frac{C_0\psi + \pi_1 - C_0\psi\pi_1 - \pi_1 S_0}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + (M_0 - T_0 - V_{000})\ln(1 - C_0) + (M_1 - T_1 - V_{001})\ln(1 - C_1) \\
& + V_{110}\ln S_0 + (T_0 - V_{110})\ln(1 - S_0) + V_{000}\ln C_0 \\
& + V_{111}\ln S_1 + (T_1 - V_{111})\ln(1 - S_1) + V_{001}\ln C_1 \\
& + (N_1 - W_{11})\ln(C_1 + \pi_1 - C_1\pi_1 - \pi_1 S_1) \\
& + W_{11}\ln[(1 - C_1)(1 - \pi_1) + \pi_1 S_1] \\
& + T_1 \ln \pi_1 + (M_1 - T_1)\ln(1 - \pi_1),
\end{aligned}
\tag{2.8}
$$

where $\boldsymbol{\eta} = (\pi_1, C_1, C_0, S_1, S_0)'$ is the nuisance parameter vector.

Next, to derive closed-form maximum likelihood estimators of the odds ratio ψ and the nuisance parameters $\boldsymbol{\eta}$, we use Tenenbein (1970)'s re-parametrization of the likelihood function. From Joseph et al. (1995) and Prescott and Garthwaite (2002), the MLEs for ψ and $\boldsymbol{\eta}$ are

$$
\begin{aligned}
\hat{\psi} =\ & \frac{(V_{100} + V_{110})\left[(V_{000} + V_{010})(V_{000} + V_{100}) + V_{000}W_{00}\right] + (V_{000} + V_{010})V_{100}W_{10}}{(V_{100} + V_{110})\left[(V_{000} + V_{010})(V_{010} + V_{110}) + V_{010}W_{00}\right] + (V_{000} + V_{010})V_{110}W_{10}} \\
& \times \frac{(V_{101} + V_{111})[(V_{001} + V_{011})(V_{011} + V_{111}) + V_{011}W_{01}] + (V_{001} + V_{011})V_{111}W_{11}}{(V_{101} + V_{111})\left[(V_{001} + V_{011})(V_{001} + V_{101}) + V_{001}W_{01}\right] + (V_{001} + V_{011})V_{101}W_{11}}, \\
\hat{\pi}_1 =\ & \frac{(V_{101} + V_{111} + W_{11})(V_{011}V_{101} + V_{001}V_{111})}{(V_{001} + V_{011} + V_{101} + V_{111} + W_{01} + W_{11})(V_{101} + V_{111})(V_{001} + V_{011})} \\
& + \frac{(V_{001} + V_{011} + W_{01})(V_{011}V_{101} + V_{011}V_{111})}{(V_{001} + V_{011} + V_{101} + V_{111} + W_{01} + W_{11})(V_{101} + V_{111})(V_{001} + V_{011})}, \\
\hat{S}_k =\ & \frac{(V_{00k} + V_{01k})V_{11k}(V_{10k} + V_{11k} + W_{1k})}{(V_{10k} + V_{11k})((V_{00k} + V_{01k})(V_{01k} + V_{11k}) + V_{01k}W_{0k}) + (V_{00k} + V_{01k})V_{11k}W_{1k}},
\end{aligned}
$$

and

$$
\hat{C}_k = \frac{V_{00k}(V_{10k} + V_{11k})(V_{00k} + V_{01k} + W_{0k})}{(V_{10k} + V_{11k})((V_{00k} + V_{01k})(V_{00k} + V_{10k}) + V_{00k}W_{0k}) + (V_{00k} + V_{01k})V_{10k}W_{1k}},
$$

where $k = 0, 1$ represent the true disease status $D = 0$ or 1. Note that the MLE $\hat{\psi}$ is a direct result from the invariance property of the MLEs, i.e.,

$$
\hat{\psi} = \frac{\hat{\pi}_1(1 - \hat{\pi}_0)}{\hat{\pi}_0(1 - \hat{\pi}_1)}.
$$

For a more detailed overview of this derivation, see Appendix A.1. Also, observe that all the MLEs are in terms of only the observed cell counts, $\boldsymbol{d}$. However, the likelihood function is in terms of both the cell counts and the category totals (M_k, N_k, T_k, for $k = 0, 1$). Therefore, when expressing the MLEs, we substitute the row totals from Table 2.1 with the corresponding sum of observed cell counts. For instance, in (2.8) we have $(N_0 - W_{10})$, but this quantity is equivalent to W_{00}. Therefore, we use W_{00} in further derivations. For all other substitutions and simplifications, refer to Table 2.1.

2.4 Restricted Maximum Likelihood Estimation

Some of the CI methods we consider require more than one evaluation at the MLEs. Recall that to calculate the score, profile likelihood, and approximate integrated likelihood CIs, we must evaluate the likelihood function at the conditional (restricted) estimates to eliminate the nuisance parameters. Therefore, in this section we derive the restricted maximum likelihood estimators of the nuisance parameters.

First, we consider the cell counts for the fallible study (refer to Table 2.1).

Table 2.2: Counts for the main study with unobserved, misclassified data

	Main study (incomplete)			
Fallible	Cases (D=1)		Controls (D=0)	
(Z)	X=1	X=0	X=1	X=0
Z=1	U_{111}	$W_{11} - U_{111}$	U_{110}	$W_{10} - U_{110}$
Z=0	U_{011}	$W_{01} - U_{011}$	U_{010}	$W_{00} - U_{010}$
	$U_{111} + U_{011}$	$N_1 - (U_{111} + U_{011})$	$U_{110} + U_{010}$	$N_0 - (U_{110} + U_{010})$
	N_1		N_0	

These counts are the subject to misclassification. We define a set of new unknown latent variables, U_{ijk}, for the unobserved misclassified data counts under the hypothesis that an infallible test was performed. Again, i, j, and k have values 0

for diseased or 1 for non-diseased and correspond to the outcomes of the fallible test (Z), "gold standard" test (X), and true disease condition (D), respectively, for each patient in the study. In Table 2.2 we show how the latent variables are distributed in the main study. Hence, U_{ijk} are the unobserved portions of the observed W_{ik} counts.

Further, based on previous derivations and the assumption that the complete study is a sub-sample of the infallible sample we can write the unobserved count distributions as

$$\begin{aligned}(U_{11k} + U_{01k}) &\sim Bin(N_k, \pi_k),\\ U_{11k}|(U_{11k} + U_{01k}) &\sim Bin(U_{11k} + U_{01k}, S_k),\\ (W_{0k} - U_{01k})|(U_{11k} + U_{01k}) &\sim Bin(N_k - (U_{11k} + U_{01k}), C_k),\end{aligned}$$

where $k = 0$ indicates the control group and $k = 1$ indicates the cases group distributions. See Joseph et al. (1995) for more details. Also, let $\boldsymbol{d}^{full} = (W_{00}, W_{01}, W_{11}, W_{10}, V_{111}, V_{001}, V_{110}, V_{000}, V_{011}, V_{101}, V_{010}, V_{100}, U_{111}, U_{011}, U_{110}, U_{010})'$ represent the full data vector including the latent variables, and let $\ell_U(\psi, \boldsymbol{\eta}|\boldsymbol{d}^{full})$ denote the log-likelihood function for the full data, from the main study and validation study. Then,

$$\begin{aligned}\ell_U \propto\ & (U_{111} + U_{011} + T_1 + U_{110} + U_{010} + T_0)\ln\pi_1 + (W_{11} + W_{01} + M_1 - U_{111}\\ & -\ U_{011} - T_1 + W_{10} + W_{00} + M_0 - U_{110} - U_{010} - T_0)\ln(1-\pi_1)\\ & - (W_{10} + W_{00} + M_0)\ln(\pi_1 - \pi_1\psi + \psi)\\ & + (W_{00} + V_{000} - U_{010})\ln C_0 + (W_{10} + M_0 - T_0 - V_{000} - U_{110})\ln(1 - C_0)\\ & + (W_{01} + V_{001} - U_{011})\ln C_1 + (W_{11} + M_1 - T_1 - V_{001} - U_{111})\ln(1 - C_1)\\ & + (V_{110} + U_{110})\ln S_0 + (T_0 - V_{110} + U_{010})\ln(1 - S_0)\\ & + (V_{111} + U_{111})\ln S_1 + (T_1 - V_{111} + U_{011})\ln(1 - S_1)\\ & + (W_{10} + W_{00} + M_0 - U_{110} - U_{010} - T_0)\ln\psi.\end{aligned} \tag{2.9}$$

Because we cannot derive a closed-form solution for the RMLEs, we construct an EM algorithm to determine the RMLEs for a fixed value of ψ. This procedure consists of the following two steps:

E-step: Let $\mathbf{\Phi}^{(r)} = (\psi, \pi_1^{(r)}, S_0^{(r)}, S_1^{(r)}, C_0^{(r)}, C_1^{(r)})'$ be the current parameter vector at the r^{th} iteration. First, we derive the conditional expectations for the latent variables $\mathbf{U} = (U_{111}, U_{011}, U_{110}, U_{010})'$, given the observable counts and current parameter estimates. These expressions are

$$U_{111}|\boldsymbol{d}, \mathbf{\Phi}^{(r)} \sim Bin\left(W_{11}, \frac{{\pi_1}^{(r)} {S_1}^{(r)}}{{\pi_1}^{(r)} S_1^{(r)} + (1 - {\pi_1}^{(r)})(1 - C_1^{(r)})}\right),$$

$$U_{011}|\boldsymbol{d}, \mathbf{\Phi}^{(r)} \sim Bin\left(W_{01}, \frac{{\pi_1}^{(r)}(1 - S_1^{(r)})}{{\pi_1}^{(r)}(1 - S_1^{(r)}) + (1 - {\pi_1}^{(r)}) C_1^{(r)}}\right),$$

$$U_{110}|\boldsymbol{d}, \mathbf{\Phi}^{(r)} \sim Bin\left(W_{10}, \frac{{\pi_1}^{(r)} S_0^{(r)}}{{\pi_1}^{(r)} S_0^{(r)} + \psi(1 - {\pi_1}^{(r)})(1 - C_0^{(r)})}\right),$$

and

$$U_{010}|\boldsymbol{d}, \mathbf{\Phi}^{(r)} \sim Bin\left(W_{00}, \frac{{\pi_1}^{(r)}(1 - S_0^{(r)})}{{\pi_1}^{(r)}(1 - S_0^{(r)}) + \psi(1 - {\pi_1}^{(r)}) C_0^{(r)}}\right).$$

Thus, the conditional expectations of the unobserved counts are

$$U_{111}^* \equiv E[U_{111}|\boldsymbol{d}, \mathbf{\Phi}^{(r)}] = \frac{W_{11}{\pi_1}^{(r)} S_1^{(r)}}{{\pi_1}^{(r)} S_1^{(r)} + (1 - {\pi_1}^{(r)})(1 - C_1^{(r)})},$$

$$U_{011}^* \equiv E[U_{011}|\boldsymbol{d}, \mathbf{\Phi}^{(r)}] = \frac{W_{01}{\pi_1}^{(r)}(1 - S_1^{(r)})}{{\pi_1}^{(r)}(1 - S_1^{(r)}) + (1 - {\pi_1}^{(r)}) C_1^{(r)}},$$

$$U_{110}^* \equiv E[U_{110}|\boldsymbol{d}, \mathbf{\Phi}^{(r)}] = \frac{W_{10}\pi_1 S_0}{{\pi_1}^{(r)} S_0^{(r)} + \psi(1 - {\pi_1}^{(r)})(1 - C_0^{(r)})},$$

and

$$U_{010}^* \equiv E[U_{010}|\boldsymbol{d}, \mathbf{\Phi}^{(r)}] = \frac{W_{00}{\pi_1}^{(r)}(1 - S_0^{(r)})}{{\pi_1}^{(r)}(1 - S_0^{(r)}) + \psi(1 - {\pi_1}^{(r)}) C_0^{(r)}}.$$

M-step: We then update the parameter estimates using the solutions to the full-data estimating equations

$$\frac{\partial \ell_U}{\partial \pi_1} = -\frac{W_{11} + W_{01} + M_1 - U_{111}^* - U_{011}^* - T_1 + W_{10} + W_{00} + M_0 - U_{110}^* - U_{010}^* - T_0}{1 - \pi_1}$$
$$+ \frac{U_{111}^* + U_{011}^* + T_1 + U_{110}^* + U_{010}^* + T_0}{\pi_1} - \frac{(W_{10} + W_{00} + M_0)(1 - \psi)}{\pi_1 - \pi_1\psi + \psi} = 0,$$

$$\frac{\partial \ell_U}{\partial S_0} = \frac{V_{110} + U^*_{110}}{S_0} - \frac{T_0 - V_{110} + U^*_{010}}{1 - S_0} = 0,$$
$$\frac{\partial \ell_U}{\partial S_1} = \frac{V_{111} + U^*_{111}}{S_1} - \frac{T_1 - V_{111} + U^*_{011}}{1 - S_1} = 0,$$
$$\frac{\partial \ell_U}{\partial C_0} = \frac{W_{00} + V_{000} - U^*_{010}}{C_0} - \frac{W_{10} + M_0 - T_0 - V_{000} - U^*_{110}}{1 - C_0} = 0,$$

and

$$\frac{\partial \ell_U}{\partial C_1} = \frac{W_{01} + V_{001} - U^*_{011}}{C_1} - \frac{W_{11} + M_1 - T_1 - V_{001} - U^*_{111}}{1 - C_1} = 0.$$

Now, solving these estimating equations for the respective nuisance parameters $\boldsymbol{\eta}$, we get the full-data MLEs in terms of ψ. We then use these estimates to update the parameter estimates $\pi_1^{(r)}$, $S_k^{(r)}$, and $C_k^{(r)}$ in the r^{th} iteration for $k = 0, 1$, corresponding to $D = 0$ for non-diseased and $D = 1$ for diseased. Hence,

$$\pi_1^{(r+1)} = \frac{B - \sqrt{B^2 - 4AC}}{2A},$$

where

$$A = (\psi - 1)(M_1 + W_{01} + W_{11}),$$

$$B = M_0 + W_{00} + W_{10} + \psi(M_1 + W_{01} + W_{11})$$
$$+ (\psi - 1)(T_0 + T_1 + U^*_{010} + U^*_{011} + U^*_{110} + U^*_{111}),$$

and

$$C = \psi(T_0 + T_1 + U^*_{010} + U^*_{011} + U^*_{110} + U^*_{111}),$$

and

$$C_k^{(r+1)} = \frac{W_{0k} + V_{00k} - U^*_{01k}}{W_{0k} + W_{1k} + M_k - T_k - U^*_{01k} - U^*_{11k}},$$

and

$$S_k^{(r+1)} = \frac{V_{11k} + U^*_{11k}}{T_0 + U^*_{01k} + U^*_{11k}}.$$

2.5 Four Confidence Intervals for ψ

We next derive four particular confidence intervals for the odds ratio ψ.

2.5.1 The Observed Information Matrix

Recall that when constructing a likelihood-based interval, we need to calculate or at least approximate the variance of the MLE of interest. In the multivariate case, we often use the observed information matrix. Let $\boldsymbol{\theta} = (\psi, \boldsymbol{\eta}')'$ represent our vector of parameters, where ψ is the odds ratio, which is the parameter of interest, and $\boldsymbol{\eta}$ is a nuisance parameter vector. Recall that we have five nuisance parameters because we assume differential misclassification so that the specificity and sensitivity for cases and controls are different. Hence, $\boldsymbol{\eta} = (\pi_1, S_0, S_1, C_0, C_1)'$ is the nuisance parameter vector. Then, the observed information matrix is

$$\mathbf{J}(\psi, \boldsymbol{\eta}) = -\begin{bmatrix} \frac{\partial^2 \ell}{\partial \psi^2} & \frac{\partial^2 \ell}{\partial \psi \partial \pi_1} & \frac{\partial^2 \ell}{\partial \psi \partial S_0} & \frac{\partial^2 \ell}{\partial \psi \partial S_1} & \frac{\partial^2 \ell}{\partial \psi \partial C_0} & \frac{\partial^2 \ell}{\partial \psi \partial C_1} \\ \cdot & \frac{\partial^2 \ell}{\partial \pi_1^2} & \frac{\partial^2 \ell}{\partial \pi_1 \partial S_0} & \frac{\partial^2 \ell}{\partial \pi_1 \partial S_1} & \frac{\partial^2 \ell}{\partial \pi_1 \partial C_0} & \frac{\partial^2 \ell}{\partial \pi_1 \partial C_1} \\ \cdot & \cdot & \frac{\partial^2 \ell}{\partial S_0^2} & \frac{\partial^2 \ell}{\partial S_0 \partial S_1} & \frac{\partial^2 \ell}{\partial S_0 \partial C_0} & \frac{\partial^2 \ell}{\partial S_0 \partial C_1} \\ \cdot & \cdot & \cdot & \frac{\partial^2 \ell}{\partial S_1^2} & \frac{\partial^2 \ell}{\partial S_1 \partial C_0} & \frac{\partial^2 \ell}{\partial S_1 \partial C_1} \\ \cdot & \cdot & \cdot & \cdot & \frac{\partial^2 \ell}{\partial C_0^2} & \frac{\partial^2 \ell}{\partial C_0 \partial C_1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \frac{\partial^2 \ell}{\partial C_1^2} \end{bmatrix}, \tag{2.10}$$

where ℓ is the log-likelihood function as derived in (2.8). In Appendix A.2 we give expressions for each of the terms in (2.10). Also, let the observed information matrix be partitioned as

$$\boldsymbol{J}(\psi, \boldsymbol{\eta}) = \begin{bmatrix} J_{11} & \boldsymbol{J}_{12} \\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22} \end{bmatrix}, \tag{2.11}$$

where $J_{11} \equiv J_\psi$ is a scalar.

2.5.2 A Wald CI

The first CI we consider is a Wald CI, which is likelihood related. The Wald statistic for ψ when nuisance parameters $\boldsymbol{\eta}$ are involved is

$$W = (\hat{\psi} - \psi)^2 [J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]^{-1},$$

where $J^{11} = (J_{11} - \boldsymbol{J}_{12}\boldsymbol{J}_{22}^{-1}\boldsymbol{J}_{21})^{-1}$ (see Pawitan (2001)). An approximate $100(1-\alpha)\%$ Wald confidence interval for the odds ratio consists of the values of ψ that satisfy

$$(\hat{\psi} - \psi)^2 [J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]^{-1} < \chi_1^2(1-\alpha), \tag{2.12}$$

where $\chi_1^2(1-\alpha)$ denotes the $1-\alpha$ quantile of a central chi-squared distribution with one degree of freedom. Then, solving (2.12) directly for ψ, we get

$$\hat{\psi} - \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]} < \psi < \hat{\psi} + \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]} \tag{2.13}$$

as an approximate $(1-\alpha)\%$ Wald confidence interval for ψ under the double-sampling procedure with differential misclassification. The information matrix used in the interval is evaluated at the MLEs derived in Section 2.3 for both the odds ratio and the nuisance parameters. We utilize the bisectional method to determine the bounds of the CI.

2.5.3 A Score CI

Next, we derive the score CI, which is also a maximum likelihood-based interval. It is an alternative method to the Wald and likelihood ratio tests; however, we cannot confirm its superiority uniformly. Hence, we compare the score CI with the Wald CI for the odds ratio of interest. The score interval is based on the score statistic and is given by

$$Sc = \left[S(\psi, \hat{\boldsymbol{\eta}}_\psi)\right]^2 \left[J^{11}(\psi, \hat{\boldsymbol{\eta}}_\psi)\right] \dot{\sim} \chi_p^2,$$

where ψ is the parameter of interest and $\boldsymbol{\eta}$ is the vector of nuisance parameters, and $\boldsymbol{\theta} = (\psi, \boldsymbol{\eta}')'$. An approximate $100(1-\alpha)\%$ score CI consists of the values of ψ that satisfy

$$\left[S(\psi, \hat{\boldsymbol{\eta}}_\psi)\right]^2 \left[J^{11}(\psi, \hat{\boldsymbol{\eta}}_\psi)\right] < \chi^2_1(1-\alpha), \tag{2.14}$$

where $\hat{\boldsymbol{\eta}}_\psi$ is the restricted MLE (RMLE) of the nuisance parameter, $\boldsymbol{\eta}$, evaluated at a fixed value of ψ and $\chi^2_1(1-\alpha)$ denotes the $1-\alpha$ quantile of a central chi-squared distribution with one degree of freedom. Here we evaluate the information matrix at the RMLEs of the nuisance parameters for a fixed value of ψ. We cannot directly solve for ψ and, therefore, we use an EM procedure described in Section 2.4 and a bisectional method to determine the above CI for ψ.

2.5.4 A Profile Likelihood CI

The next confidence interval we consider is the profile likelihood CI, which is pseudo-likelihood based. For this CI we replace the nuisance parameters in the likelihood function with their RMLEs to eliminate them. Recall that the profile likelihood CI is based on the profile likelihood function

$$L_P(\psi) \equiv \max_{\boldsymbol{\eta}} L(\psi, \boldsymbol{\eta}) = L(\psi, \hat{\boldsymbol{\eta}}_\psi),$$

where $\hat{\boldsymbol{\eta}}_\psi$ is the vector of RMLEs in terms of the parameter of interest ψ. Again we employ the use of the RMLEs computed using the EM algorithm described in Section 2.4. Then, an approximate $100(1-\alpha)\%$ profile likelihood CI consists of the values of ψ that satisfy

$$-2\left[\ell(\psi, \hat{\boldsymbol{\eta}}_\psi) - \ell(\hat{\psi}, \hat{\boldsymbol{\eta}})\right] < \chi^2_1(1-\alpha), \tag{2.15}$$

where $\chi^2_1(1-\alpha)$ denotes the $1-\alpha$ quantile of a central chi-squared distribution with one degree of freedom, where $\ell(\hat{\psi}, \hat{\boldsymbol{\eta}}) \equiv \ln\left[L(\hat{\psi}, \hat{\boldsymbol{\eta}})\right]$ is evaluated at the MLEs $\hat{\psi}$ and $\hat{\boldsymbol{\eta}}$, given in Section 2.3. We use a bisectional method on the log-likelihood $\ell(\psi, \hat{\boldsymbol{\eta}}_\psi)$ evaluated at the RMLEs from Section 2.4 to compute the CI of interest. Because

we cannot derive the exact RMLEs, no closed form derivation exists for the profile likelihood interval. However, as mentioned by Riggs (2006), the profile likelihood interval is often used for its simplicity and ease of understanding, although it does not account for the uncertainty in the nuisance parameters estimation and can yield optimistically short intervals for ψ.

2.5.5 An Approximate Integrated Likelihood CI

The last interval we consider utilizes the integration procedure as a method of eliminating the nuisance parameters from the likelihood function. Therefore, this interval is called the approximate integrated likelihood CI because it approximates the integrated likelihood function. Often, the integration of the nuisance parameters can be an impossible task, so we use the Laplace approximation method to derive the approximate integrated likelihood function

$$L_{AI}(\psi) = \int L(\psi, \boldsymbol{\eta}) d\boldsymbol{\eta} \approx \frac{cL_P(\psi)}{\left|\hat{\boldsymbol{J}}_{\boldsymbol{\eta}}(\psi, \hat{\boldsymbol{\eta}})\right|^{1/2}}, \tag{2.16}$$

where $c = (2\pi)^{\nu/2}$, ν is the dimension of the nuisance parameter, and $L_P(\psi)$ is the profile likelihood defined as

$$L_P(\psi) \equiv \max_{\boldsymbol{\eta}} L(\psi, \boldsymbol{\eta}) = L(\psi, \hat{\boldsymbol{\eta}}_\psi). \tag{2.17}$$

Notice that $\boldsymbol{J}_{\boldsymbol{\eta}}$ is the nuisance parameter sub-matrix of the observed information. Thus, from (2.11) we have $\hat{\boldsymbol{J}}_{\boldsymbol{\eta}}$ is $\boldsymbol{J}_{22}$ evaluated at the MLEs of the nuisance parameters. We use the expressions for the information matrix included in Appendix A.2 and the expressions for the MLEs in Section 2.3. Then, we define an approximate $100(1-\alpha)\%$ approximate integrated likelihood CI as the set of values of ψ that satisfy

$$-2\left[\ell_{AI}(\psi) - \ell_{AI}(\hat{\psi}_{AI})\right] < \chi^2_1(1-\alpha). \tag{2.18}$$

Here, $\ell_{AI}(\psi)$ is the approximate integrated log likelihood evaluated at a fixed value of ψ where $\ell_{AI}(\hat{\psi}_{AI})$ is the log likelihood.

2.6 A Simulation Study

Here, we examine the behavior of four different confidence intervals for the estimation of the odds ratio, ψ, in a case-control study with differential misclassification when using double sampling. We wish to compare the performance of those intervals based on coverage and width. We performed a Monte Carlo simulation to examine the effect of the sample sizes of both cases and controls, M_k and N_k, respectively, for $k = 0$ and 1, the effect of the disease probability π_j, for $j = 0, 1$, for the "gold standard" test outcome $X = 0, 1$, and the magnitude of the odds ratio ψ, on the coverage and width of the CIs defined in intervals (2.13) to (2.18).

2.6.1 Parameter and Sample Size Configurations

For the simulation, we first chose the assumed values for the odds ratio ψ. We considered two values for ψ: $\psi = 2$ and $\psi = 4$. Also, we let $\pi_0 \equiv Pr(X = 1|D = 0)$ and $\pi_1 \equiv Pr(X = 1|D = 1)$ where $X = 0$, which indicates that the "gold standard" test showed non-diseased and $D = 0$ indicated the true condition of the participant as non-diseased.

Table 2.3: Odds ratio and probabilities for the simulation in CCDIFF

ψ	π_0	π_1
2	0.25	0.40
4	0.38	0.71

Next, we defined the probabilities of differential misclassification, i.e., the sensitivity and specificity for cases or controls. Dahm et al. (1995) have considered the misclassification to be in one of three categories – low, medium and high. We considered only low and high misclassification with the corresponding probability values in Table 2.4.

Table 2.4: Specificity and Sensitivity for CCDIFF

Misclassification	Specificity		Sensitivity	
	C_0	C_1	S_0	S_1
Low	0.99	0.97	0.98	0.96
High	0.90	0.85	0.80	0.75

Again, the specificity was $C_k = Pr(Z = 0|X = 0, D = k)$ and sensitivity was $S_k = Pr(Z = 1|X = 1, D = k)$ for $k = 0, 1$, for non-diseased and diseased, respectively. We examined the CIs for eight different sample sizes in each of the described situations – low or high misclassification with $\psi = 2$ or 4. The sample sizes for each of the complete or incomplete studies are shown in Table 2.5.

Table 2.5: Sample sizes used in the simulation for CCDIFF

	A1	A2	A3	A4	A5	A6	A7	A8
M_0	50	50	75	100	125	150	175	200
M_1	30	30	37	50	62	75	87	100
N_0	250	500	750	1000	1250	1500	1750	2000
N_1	120	250	370	500	620	750	870	1000

Notice that the sample sizes for the complete studies are substantially smaller than the corresponding sample sizes for the incomplete studies. Thus, the summarized parameter configurations for this simulation are $\psi \in \{2, 4\}$, $\pi_1 \in \{0.40, 0.71\}$, $S_0 \in \{0.98, 0.80\}$, $S_1 \in \{0.96, 0.75\}$, $C_0 \in \{0.99, 0.90\}$, and $C_1 \in \{0.97, 0.85\}$. We generated 10,000 data sets and calculated the four 95% confidence intervals for the binomial odds ratio, ψ, under double sampling with differential misclassification in case-control studies.

2.6.2 Results

We first consider the scenario where we have low differential misclassification. Refer to Table 2.4 to recall the appropriate probabilities. The assumed odds ratio parameter is $\psi = 2$. Figure 2.1 shows the coverage of all four intervals and Figure 2.2 shows the boxplots of the four competing CI widths.

Consider the Wald interval first. From figure 2.1 we see that the Wald interval is very conservative and overestimated the desired 95% confidence level for all of the sample size scenarios, $A1$ - $A8$ (Table 2.5). This statement is also supported by the boxplot Figure 2.2 showing that the Wald interval has a consistently higher median and mean width compared to the profile likelihood and the approximate-integrated likelihood. This property makes containing the true parameter value easier and more frequent, thus causing the over- coverage of the Wald CI. When the sample size increased, we observed significant improvement in the interval widths. However, the coverage did not seem to be affected.

From the score interval, we see an interesting result. Although the interval had the highest median and average width for the first five sample size scenarios, it still had the lowest coverage for all of the sample size scenarios considered. This result is perhaps explained by the interval variability. We see that even after we increased the sample sizes, the range of the score interval was quite large compared to that of the competing intervals. This fact is somewhat unusual because the score interval is generally believed to be an improvement of the Wald interval.

Next, consider the profile likelihood CI. From Figure 2.1 we see that even though the interval underestimates the coverage throughout all the sample size examples, a definite improvement occurred with the sizes increasing from $A1$ to $A8$ (Table 2.5). The profile likelihood showed better estimation than the score and Wald; however, its coverage was still less than the approximate-integrated CI. This behavior is seen better on Figure 2.1. Notice that the profile likelihood CI widths'

median and mean values were consistently smaller than those of the approximate-integrated likelihood CI, thus explaining the profile likelihood CI's undercoverage. Again, notice the definite improvement of interval variability with the increasing of the sample sizes.

The last interval we consider in this paper is the approximate-integrated interval. As per our expectations, this method seemed to give the best results in the estimation of the odds ratio. Figure 2.1 shows that the AI interval also underestimated the coverage. However, with the increased sample sizes, he interval was it gets much closer to the desired 95% confidence level than the three competing intervals, a statement also supported by Figure 2.2. The approximate-integrated interval median and average width were higher than the profile likelihood, making for more conservative coverage and constantly improving with the sample size increase from $A1$ to $A8$.

Next, we considered the case with low differential misclassification (see Table 2.4) and the parameter of interest of $\psi = 4$. The results of the simulation analysis are shown in Figures 2.3 and 2.4. The conclusions we obtained in this case were almost identical to the conclusion from the previous case with low differential misclassification and $\psi = 2$ discussed earlier in this section. We noticed that the Wald interval overestimated the desired 95% confidence level for all the sample size examples – $A1$ to $A8$. Again, we see from the interval width boxplot in Figure 2.4 that this interval had a higher median and average width compared to those of the other pseudo-likelihood CIs and that the interval tightened when the sample size was increased. The results for the score interval were similar to the previous score interval preformance discussed before. However, in this case we noticed that the coverage actually declined with an increase in the sample size. Looking at Figure 2.4, we see that this result was supported by the decrease of the CI widths. Thus, the tighter CIs had smaller chances of containing the true parameter, ψ. The profile

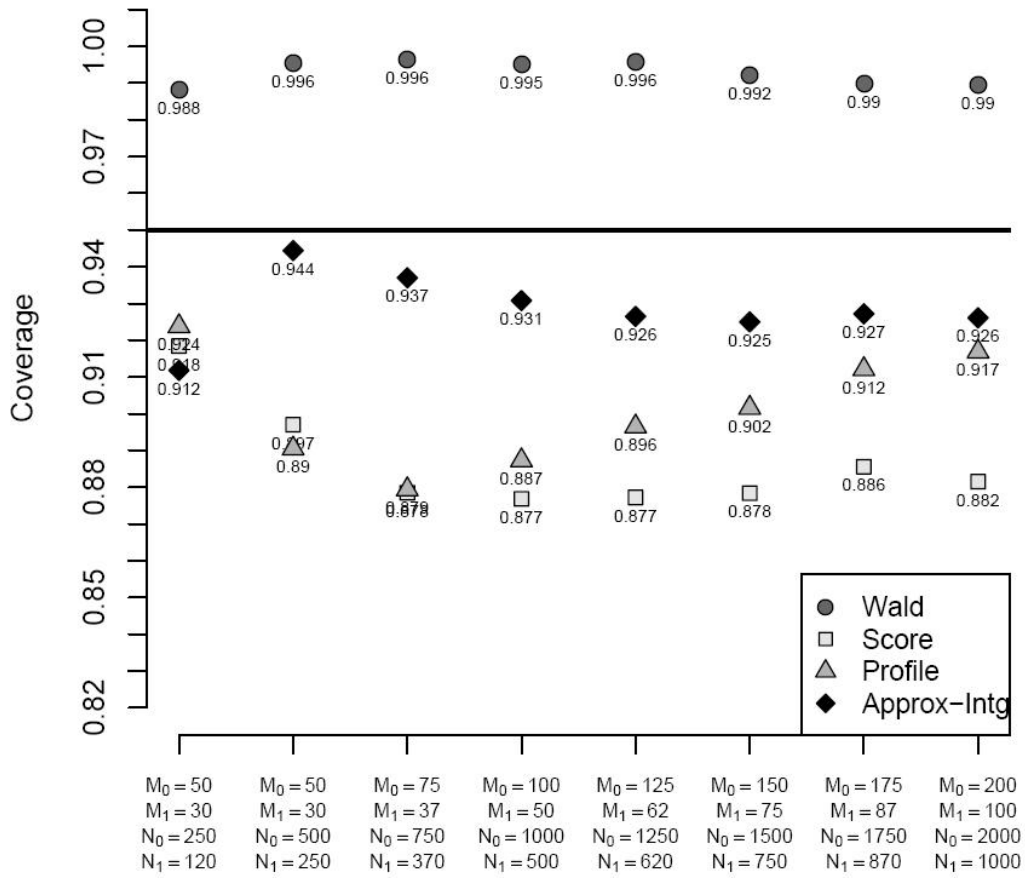

Figure 2.1: Coverage probabilities of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with low differential misclassification and an odds ratio of $\psi = 2$

likelihood interval followed the exact same pattern as the profile likelihood from the case with $\psi = 2$. It also underestimated the 95% confidence limit, and its average and median widths were less than the approximate-integrated likelihood interval. Again, the approximate-integrated interval underestimated the coverage, but the coverage improved significantly with the sample size increase. Also, we conclude that the best interval in terms of coverage is the AI interval.

Now, we consider the second scenario, where we have high differential misclassification (refer to Table 2.4) and an odds ratio parameter of $\psi = 2$. Figures 2.5 and 2.6 show the results of the analysis of the simulated datasets under those conditions.

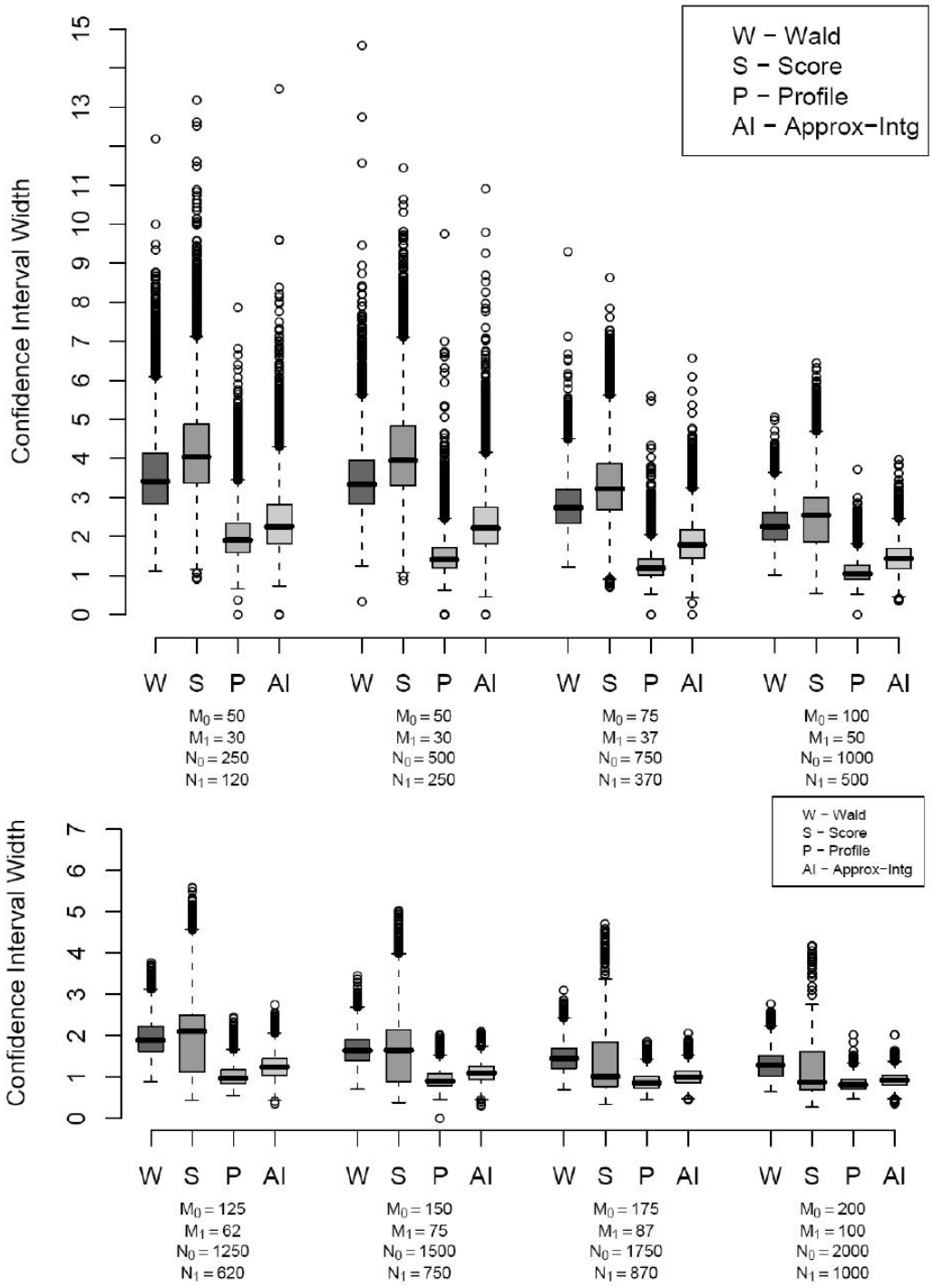

Figure 2.2: Interval widths of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with low differential misclassification and an odds ratio of $\psi = 2$

Notice again the coverage plot patterns for the odds ratio for all four intervals are similar. With high misclassification we found more consistent results. We see in Figures 2.5 that we can order the confidence intervals by worst to best coverage in the ordering score, Wald, profile likelihood, and approximate integrated.

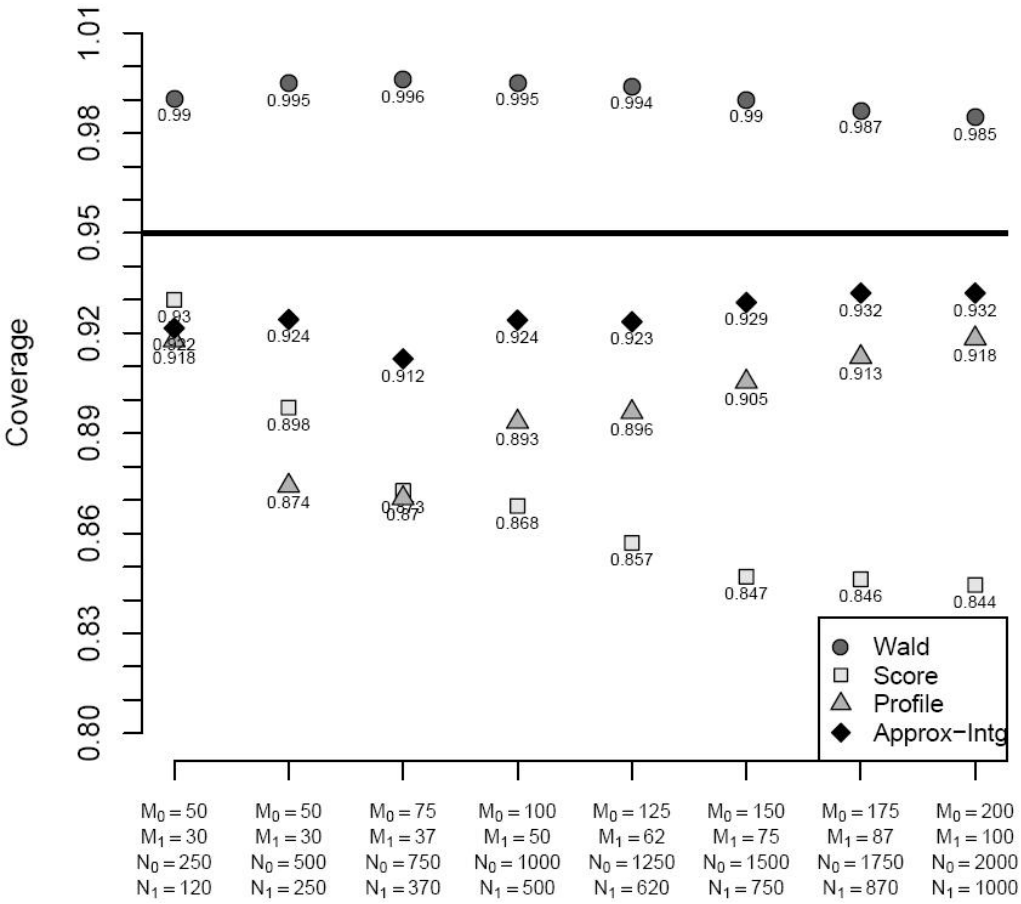

Figure 2.3: Coverage probabilities of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with low differential misclassification and odds ratio of $\psi = 4$

The Wald CI underestimates the desired confidence level; nevertheless, it shows significant improvement with the sample sizes increasing from $A1$ through $A8$. The boxplot indicates that similar results should be expected since the mean and range of the widths of the Wald CI are consistently smaller than the PL and the AI CIs.

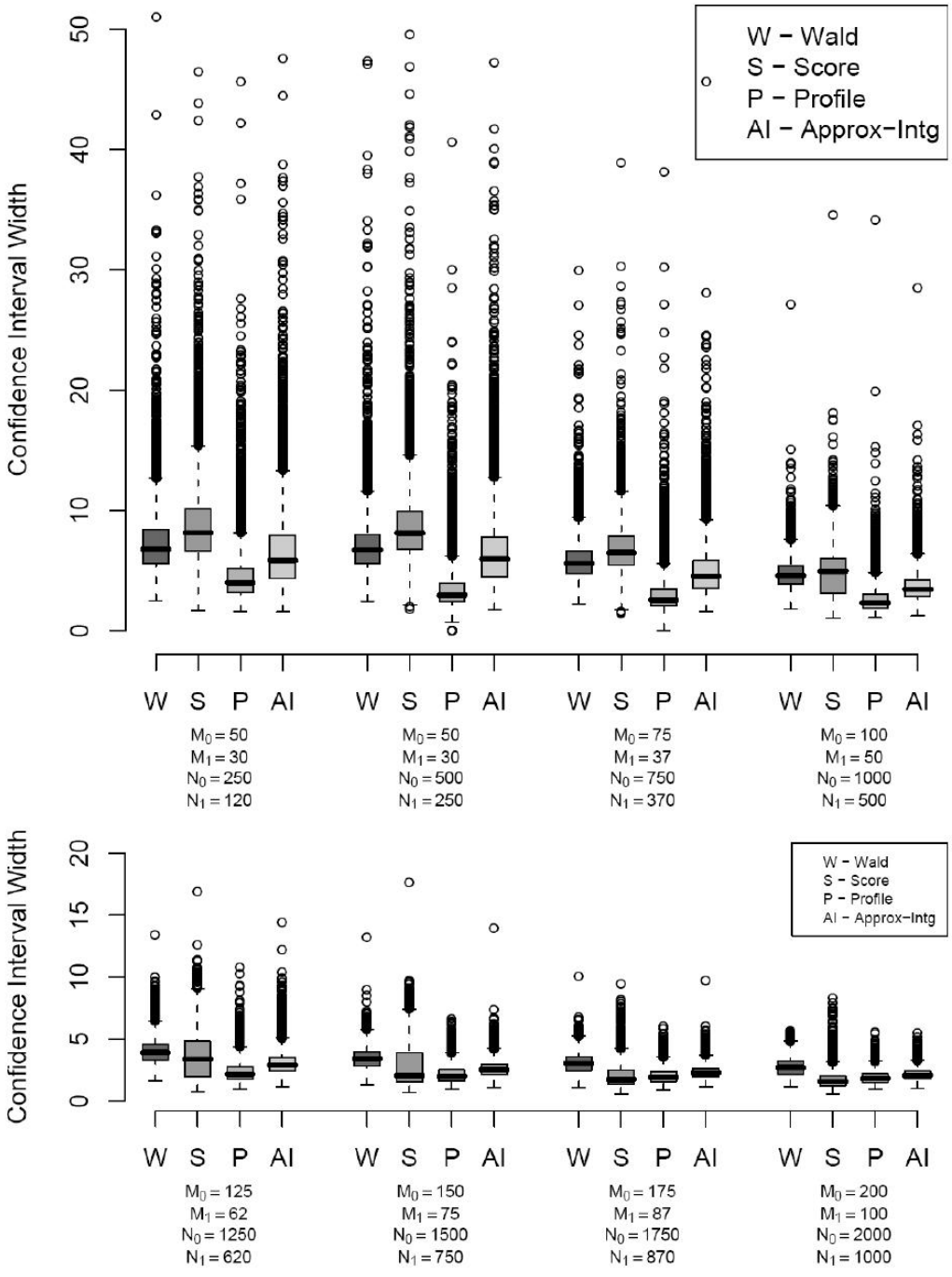

Figure 2.4: Interval widths of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with low differential misclassification and an odds ratio of $\psi = 4$

The next interval of interest is the score CI. As our previous discussion from the case of low misclassification shows we expect the score to perform worse than

the rest. The figures support this claim. We see on Figure 2.5 that the score interval vastly underestimated the confidence level but still showed improvement for the larger sample sizes. This observation is supported by Figure 2.6, as well, where we see that the score interval widths were smaller than all three other intervals throughout the eight sample size scenarios. The ranges and means of the widths of the interval are smaller, as well as the median, which made this interval much less conservative than the rest.

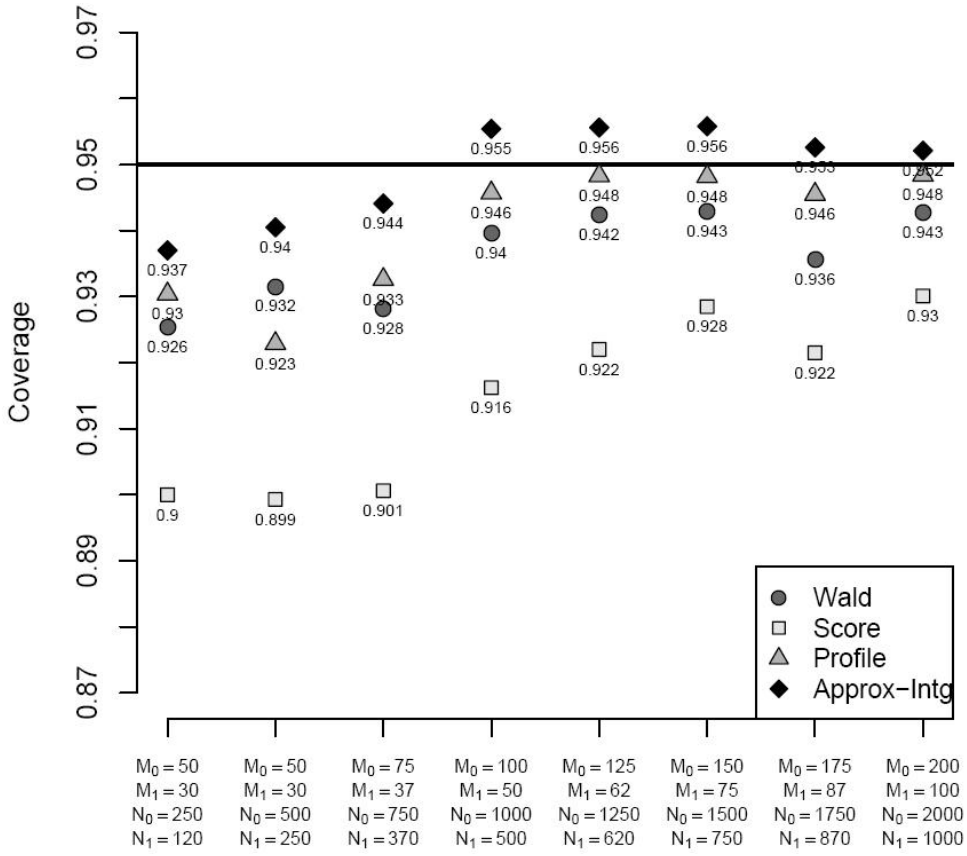

Figure 2.5: Coverage probabilities of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with high differential misclassification and odds ratio of $\psi = 2$

Next we look at the profile likelihood interval. Again, it underestimated the coverage over all the sample sizes; nevertheless, the estimation was much closer to

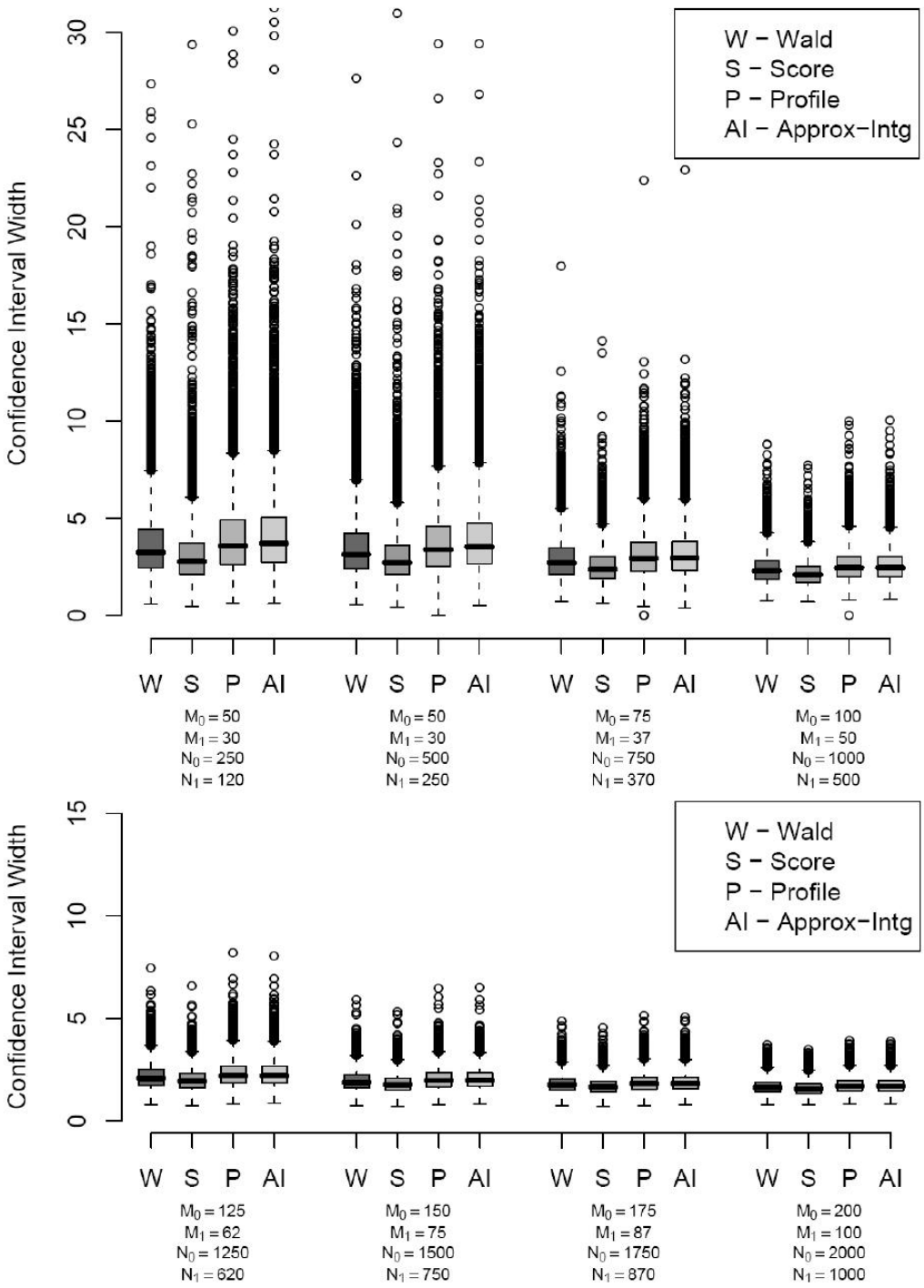

Figure 2.6: Interval widths of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with high differential misclassification and an odds ratio of $\psi = 2$

the desired 95% confidence level. The boxplot Figure 2.6 is also consistent with this finding. The profile likelihood interval had higher medians and ranges of the widths than the two full likelihood based intervals making it more likely to contain the true parameter value.

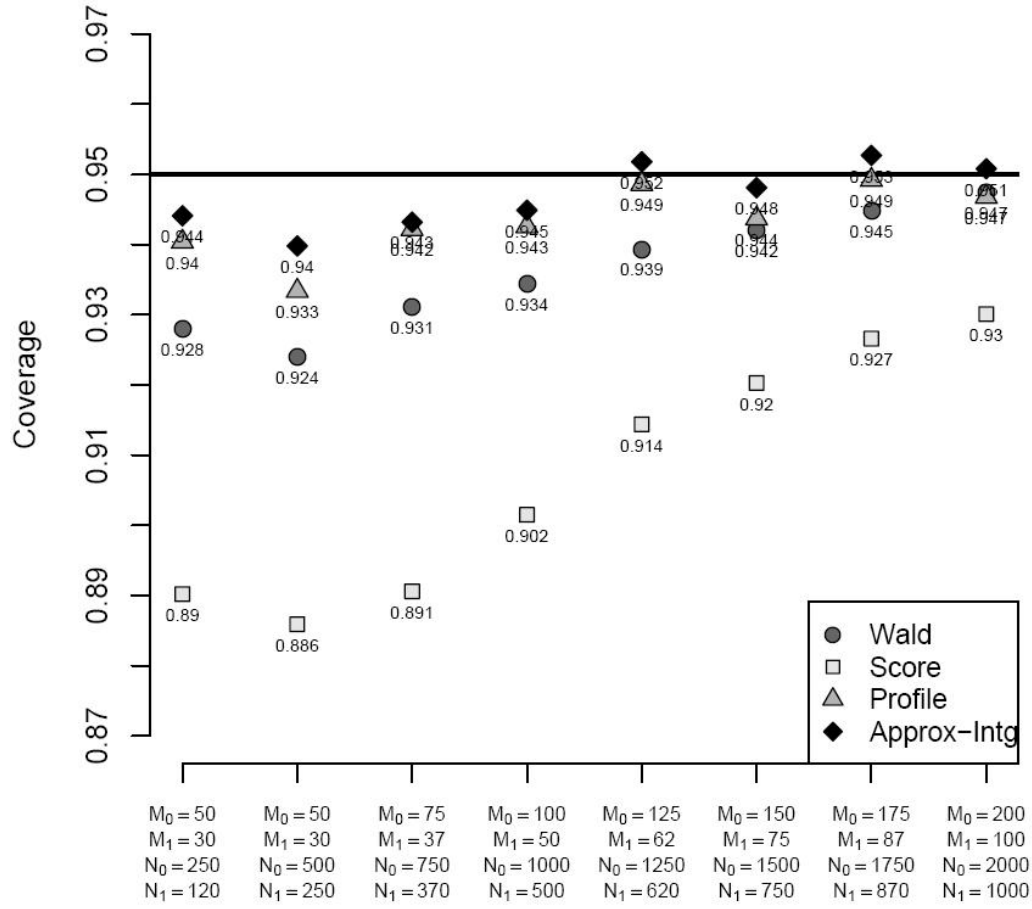

Figure 2.7: Coverage probabilities of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with high differential misclassification and odds ratio of $\psi = 4$

The last interval considered in this paper is the approximate-integrated interval. For this set-up the interval performed better than the rest of the intervals. For a couple of the sample size scenarios, it overestimated the confidence level desired; however, with the increase of sample size values, the interval had much better

results. Again, we see that the boxplot yielded similar results. The interval was definitely wider than the Wald and score intervals, and it was very close to the profile likelihood one. The overall variability of the approximate-integrated interval was better than the profile likelihood since its interquartile range was consistently smaller. Overall, all four intervals underestimated for the smaller sample sizes. However, the coverage properties became considerably better for the larger sample sizes. The smaller variability in the observed results might be due to the fact that we have more information for the fallible samples when we have a higher misclassification.

The last case we consider is the case of high misclassification and an odds ratio parameter of $\psi = 4$. Figures 2.7 and 2.8 summarize our findings. Again we see that the overall patterns and results were similar to those of the high misclassification case discussed above. The Wald confidence interval performed much better than the score but somewhat worse than the profile likelihood and the approximate-integrated intervals. The boxplot also supports this conclusion. The Wald interval was comparatively tighter than the two pseudo-likelihood intervals but wider than the score interval. The median value was consistently smaller than the profile likelihood and the approximate-integrated intervals, which explains its lower coverage.

Next, we look at the score interval, and again as expected, it performed more poorly than the rest. From the boxplot Figure 2.8, we see that the score interval had the smallest median value as well as the smallest interquartile range. This result explains its lower coverage percentages. Nevertheless, both Wald and score intervals coverage properties improved greatly when the sample sizes were increased.

Let us consider the profile likelihood interval next. The profile likelihood interval showed consistent improvement when the sample sizes changed from $A1$ to $A8$. From Figure 2.8 we see that the profile likelihood and the approximate-integrated intervals showed little difference in interval width. Also, consistent with all of the previous cases discussed, the approximate-integrated interval improved with the in-

crease of the sample size. The similarity in interval width between the two pseudo-likelihood-based intervals is illustrated in Figure 2.8 as well.

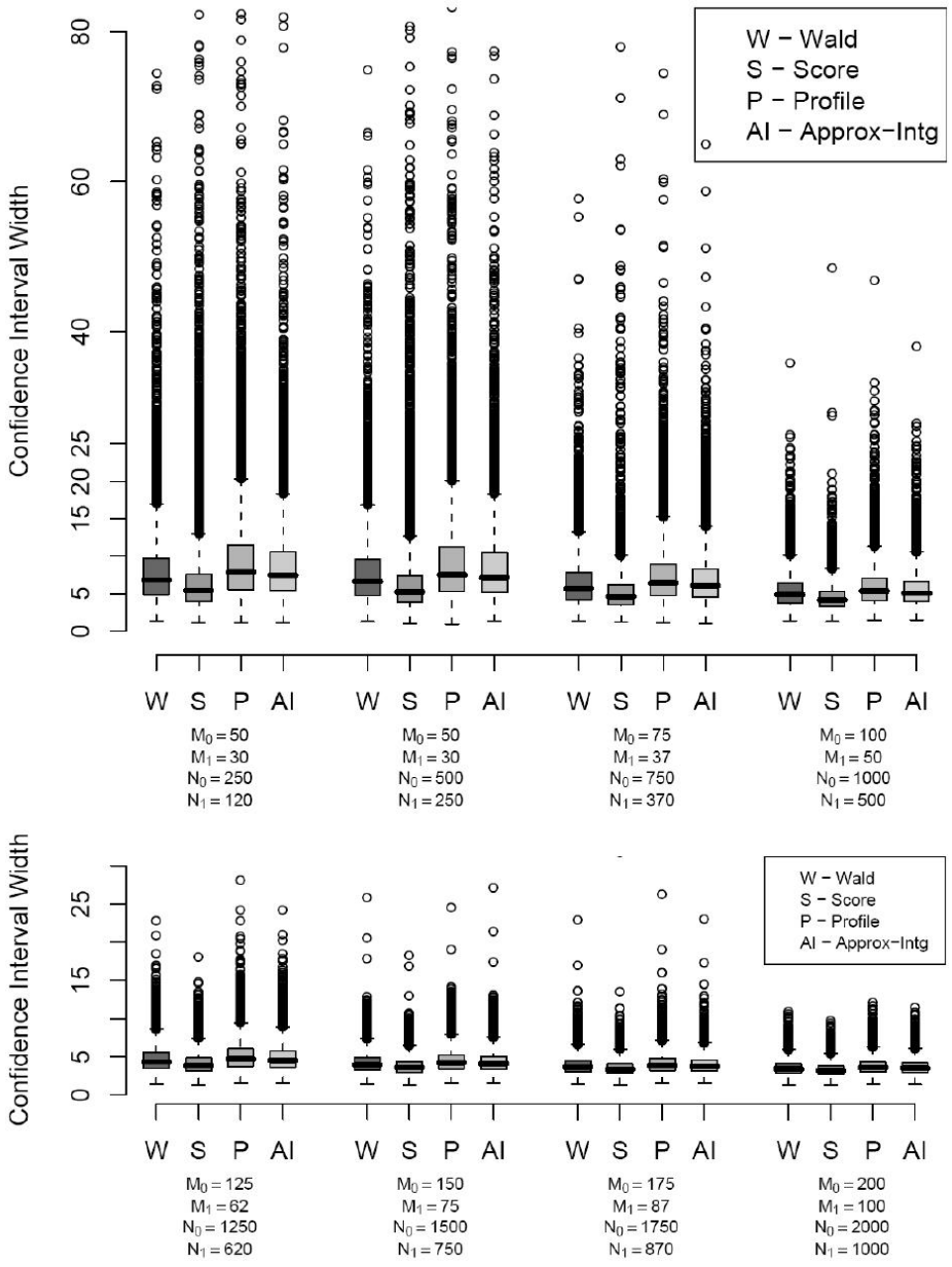

Figure 2.8: Interval widths of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with high differential misclassification and an odds ratio of $\psi = 4$

2.7 Overall Conclusions and Comments

Overall, we found that the approximate integrated interval performed best under the scenarios of low or high misclassification and different values of the odds ratio parameter considered here. This finding is consistent with the results mentioned by other authors that examined this interval under different conditions (Greer (2008) and Boese (2005)). Significantly, the interval is not too computationally or theoretically complicated. The score interval performed worst among all CIs in our simulations in terms of coverage properties.

Also, although the simulations were performed with the same seed assignment, odds ratio value, and sample sizes for each combination of misclassification, we still had an issue with the small probabilities for each of the cells in Table 2.1 and Table 2.2. Therefore, when the simulated results produced a zero count, we replaced the missing empty cell values between 0.0001 and 0.05. Thus, for a small number of cases this replacement affects the coverage and the variability of the interval widths.

CHAPTER THREE

Confidence Intervals for the Odds Ratio in Case-Control Studies with Non-differential Misclassification

3.1 Introduction

Epidemiological research is currently one of the most popular research areas. This science relies heavily on statistical analysis for establishing and quantifying the relationship between risk factors and disease, and for analyzing the presence of a particular disease in a given geographic area. Hence, statisticians have been designing many different types of studies to produce this information.

Here, we consider one particular type of statistical study; the case-control study, that is based on events that have already occurred (retrospective, observational study) rather than on events that will occur in the future (prospective study). Case control studies date back as early as the eighteen hundreds and are highly utilized in the domain of cancer research, Breslow (1996). The main idea behind this study design is to compare a cohort of people that exhibit a disease of interest (cases) to a cohort of individuals without the disease (controls) to obtain conclusions about the rates, risk factors (exposure), and consequences of that disease. The most popular relative measure of such exposure-disease relation is the odds ratio. Its utility is not because of its intelligibility as an effect measure for epidemiological research results, but rather for its convenient mathematical properties, Nurminen (1995).

Although, case-control design is an extremely useful and efficient way of conducting epidemiological studies, it is also known for its numerous shortcomings and limitations, such as selection bias, recall bias, and confounding (see Breslow (1996) and Langholz (2010)). Epidemiological studies are usually observational and ret-

rospective. They often rely on self reporting and recollection of events and factors in the past, for which the researcher has no control. This fact causes measurement error. We recognize two major types of misclassification – differential and non-differential. For differential misclassification we assume that the cases group and controls group exhibit different specificity and sensitivity measures, whereas for the non-differential case those probabilities are assumed to be equal. Bross (1954) first addressed the crucial impact of misclassification errors when performing statistical analysis.

Many statisticians and scientists have proposed various methods for correcting the bias and estimating the misclassification error in case-control studies. Walter and Irwig (1988) have discussed the deleterious effect of ignoring misclassification and Thomas, Stram, and Dwyer (1993) have offered a review of a few methods of correction. Espeland and Hui (1987) have proposed several ideas for adjusting for misclassification using information on error rates assessed in different ways and Gustafson, Le, and Saskin (2001) have examined a scenario where partial knowledge on the misclassification was available and used to correct the odds ratio estimates. Maximum-likelihood and closed-form estimators of some epidemiologic measures have been considered by Greenland (2008). Also, Morrissey and Spiegelman (1999) have presented matrix methods for estimating the odds ratio subject to exposure misclassification. However, one of the most utilized procedures for estimating the population proportion of exposure is the double sampling procedure first introduced by Tenenbein (1970). He suggested that by comparing the results obtained by two or more measuring devices, one can obtain information on the extend of misclassification, and thus, make corrections to the biased estimators. Tenenbein's double sampling idea has also been implemented by Dahm, Gail, Rosenberg, and Pee (1995), who investigated the value of additional fallible data in order to improve the point estimation of the odds ratio.

Several authors have discussed different methods of calculating confidence intervals under these circumstances. Likelihood-based confidence intervals for the proportion parameters with binary data have been considered by Paul and Thedchanamoorthy (1997) and Boese (2005). A Bayesian approach to CI estimation for the population proportion under the double sampling and false-positive misclassification has been offered by Lee and Byun (2008). For the odds ratio of two binomial samples, Troendle and Frank (2001) have derived unbiased CIs and Reiczigel et al. (2008) have derived an exact unconditional CI. In regards to the case of non-differential misclassification, Correa-Villaseor et al. (1995) examined in detail the effects of bias in case-control studies with three levels of exposure. In particular, our attention lies in determining the CIs for the odds ratio under non-differential misclassification assumption. We utilize the double sampling paradigm as the method of choice for correcting for misclassification errors.

In this chapter we derive four confidence intervals for the odds ratio parameter of interest in a case-control study with non-differential misclassification. We consider two likelihood-based intervals - the Wald and score CIs, and two pseudo-likelihood-based intervals - the profile likelihood and approximate integrated likelihood CIs. Here, the double-sampling process of sampling usually produces not only the parameter of interest, but also a set of nuisance parameters. To eliminate the effect of the nuisance parameters we employ the four methods of calculating CIs mentioned above. Karunaratne (1991) has discussed and developed point estimates of the odds ratio parameter for the case-control study under double sampling and non-differential misclassification, which is the basis for this paper.

The reminder of the chapter is organized as follows. First, in Section 3.2 we present the model, basic terms, and notation used for this study. We also discuss the double sampling procedure in more detail, as well as the assumption of non-differential misclassification. Then, in Section 3.3 we derive maximum likelihood

estimating equations for both the parameter of interest and the nuisance parameters. We utilize the Newton-Raphson method for estimating those parameters when we cannot reach closed form solutions. The results of this section are used to develop a Wald confidence interval. The other three intervals require the derivation of restricted maximum likelihood estimators (RMLEs). We present an EM algorithm to yiels the RMLEs in Section 3.4. Further, in Section 3.5, we give an overview of the four CIs after implementing the results from the previous two sections. Also, we derive the observed information matrix, which is an essential part of the derivation of the Wald and score CIs. In Section 3.6 we describe a simulation study comparing the coverage and interval widths of each of the four intervals for two levels of misclassification - low and high. Finally, in Section 3.7 we give several comments, concerns, and concluding remarks on the four CIs compared here.

3.2 The Model

The case-control study of interest here involves the double-sampling method. Let D be the true indicator of a disease status. Then, suppose we have two different procedures that one can use to test for a particular illness. One procedure is a very accurate, well-known and proven method for detection of the disease, whereas the other gives less reliable and faulty results. We refer to the first method as the "gold standard" and the second as the fallible test. Let X and Z be the binary outcomes measuring the exposure level of each of the tests respectively. Thus, $X = 1$ represents an individual that has been positively diagnosed by the infallible test and $Z = 1$ represents an individual positively diagnosed by the fallible test. For our study we first choose a larger group of people that are tested using only the erroneous device. Then, we choose a smaller sub-sample that is tested with the absolutely accurate device. This smaller sample is referred to as the validation or complete sample. Tenenbein (1970) has referred to this method a double-sampling method.

Let, the subscripts i, j, $k = 0, 1$ represent the outcomes of the fallible test, $Z = i$, the "gold standard" test, $X = j$, and the true disease status, $D = k$, where "1" indicates diseased and "0" indicates healthy. Now, define V_{ijk} to be the cell counts for the complete data (where both tests are performed) and W_{ik} for the incomplete data (where only erroneous test is used). Refer to Table 3.1 for a visual explanation of the notation. Also, let M_k denote the sample size from the disease group $D = k$ from the complete study. For the incomplete study, N_k represent the number of patients in the patient group where $D = k$. Notice that $M_k + N_k$ gives the total amount of participants in the study for each group of cases or controls. The first step sample sizes from the double-sampling procedure are N_k and $M_k + N_k$. We define the counts T_k as the number of people from each disease group ($D = 0$ or 1) that tested positive from the infallible test, $X = 1$.

Table 3.1: Counts for a study with misclassified exposure data

	Validation study (complete)				Main study (incomplete)	
Fallible	Cases (D=1)		Controls (D=0)		Cases (D=1)	Controls (D=0)
(Z)	X=1	X=0	X=1	X=0		
Z=1	V_{111}	V_{101}	V_{110}	V_{100}	W_{11}	W_{10}
Z=0	V_{011}	V_{001}	V_{010}	V_{000}	W_{01}	W_{00}
	T_1	$M_1 - T_1$	T_0	$M_0 - T_0$		
	M_1		M_0		N_1	N_0

For the validation study we denote the probability of exposure ($X = 1$) for each $D = k$ group by π_k. Thus,

$$\pi_k = Pr(X = 1|D = k), \tag{3.1}$$

where $k = 0$ is the control group and $k = 1$ is the cases group. In the complete study we also utilize a fallible test measure for which we define specificity and sensitivity probabilities. The sensitivity, S, is also known as the true positive rate, indicates the probability that an individual tests positive by the fallible test ($Z = 1$) given that

he/she tested positive by the "gold standard" test as well ($X = 1$). The specificity, C, represents the probability that a subject is not positive as obtained by the fallible test ($Z = 0$), given that he/she was also found to be not positive by the "gold standard" ($X = 0$). In Section 3.1, we briefly discussed the different types of misclassification error. Here, we are interested in estimating only the odds ratio under the assumption of non-differential measurement error. Non-differential misclassification of exposure between diseased and non-diseased subjects occurs when the exposure, Z and X, are equally misclassified among cases the ($D = 1$) and controls ($D = 0$). Then, the exposure status is independent of the disease outcome level. We define the specificity and sensitivity probabilities, respectfully, as

$$\begin{aligned} S &= Pr(Z = 1|X = 1, D = 0) \\ &= Pr(Z = 1|X = 1, D = 1) \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} C &= Pr(Z = 0|X = 0, D = 0) \\ &= Pr(Z = 0|X = 0, D = 1). \end{aligned} \tag{3.3}$$

Based on (3.1) - (3.3), and the derivations in Prescott and Garthwaite (2002), we induce the following distributions for the complete study and the observed cell counts:

$$T_k = V_{01k} + V_{11k} \sim Bin(M_k, \pi_k), \tag{3.4}$$

$$V_{11k}|T_k \sim Bin(T_k, S_k), \tag{3.5}$$

and

$$V_{00k}|T_k \sim Bin(M_k - T_k, C_k), \tag{3.6}$$

where $k = 0$ or 1 indicates the true disease status of the participant. We define the observable counts' distributions and probabilities in the incomplete data set as

$$W_{1k} \sim Bin(N_k, \pi_k S + (1 - \pi_k)(1 - C)),$$

where k is indexed as described above. Finally, our parameter of interest, the odds ratio, is

$$\psi \equiv \frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}. \tag{3.7}$$

The estimation of the odds ratio and other statistics of interest in the case-control study design have been examined by many scientists. One of the most examined studies is one concerning sudden infant death syndrome (SIDS) that was first examined by Drews et al. (1990). The study examined the maternal use of antibiotics during pregnancy and the incidences of SIDS. The drug use was measured by unreliable personal interview (Z) and was validated be the mother's medical records (X). Later, Greenland (1988) and Morrissey and Spiegelman (1999) have considered this data for in both the differential and non-differential misclassification cases. Karunaratne (1991) has explained that the non-differential misclassification is often unreasonable for case-control studies because disease outcome is known prior to exposure classification.

3.3 Maximum Likelihood Estimation

In this section we derive the maximum likelihood estimator (MLE) of the parameter of interest, ψ, as well as the nuisance parameters defined in Section 2.2. Let the vector of all parameters be $\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\psi, \boldsymbol{\eta}')'$ and $\boldsymbol{\eta}$ is the vector of nuisance parameters. To derive the CIs for the odds ratio we first derive the MLEs for the nuisance parameters. Let $\boldsymbol{d} = (W_{00}, W_{01}, W_{11}, W_{10}, V_{111}, V_{001}, V_{110}, V_{000}, V_{011}, V_{101}, V_{010}, V_{100})'$ denote the observed data counts for the full data (both main and validation studies), let $L = L(\pi_1, \pi_0, C, S|\boldsymbol{d})$ denote the full-likelihood function, and let ℓ represent the log-likelihood function. Then,

$$
\begin{aligned}
\ell \propto\ & W_{11}\ln[\pi_1 S + (1-\pi_1)(1-C)] + W_{10}\ln[\pi_0 S + (1-\pi_0)(1-C)] \\
& + (N_1 - W_{11})\ln[1-\pi_1 S - (1-\pi_1)(1-C)] + T_1\ln(\pi_1) \\
& + (N_0 - W_{10})\ln[1-\pi_0 S - (1-\pi_0)(1-C)] + T_0\ln(\pi_0) \\
& + (M_1 - T_1)\ln(1-\pi_1) + (M_0 - T_0)\ln(1-\pi_0) \\
& + (V_{011} + V_{010})\ln(1-S) + (V_{101} + V_{100})\ln(1-C) \\
& + (V_{111} + V_{110})\ln S + (V_{001} + V_{000})\ln C.
\end{aligned}
$$

We use the substitution

$$
\pi_0 = \frac{\pi_1}{\pi_1 - \pi_1\psi + \psi}
$$

to get

$$
\begin{aligned}
\ell_\psi \propto\ & (M_0 - T_0)\ln\left[1 - \frac{\pi_1}{\psi + \pi_1 - \psi\pi_1}\right] + W_{10}\ln\left[\frac{(1-C)(1-\pi_1)\psi + \pi_1 S}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + T_0\ln\left[\frac{\pi_1}{\psi + \pi_1 - \psi\pi_1}\right] + (N_0 - W_{10})\ln\left[\frac{C\psi + \pi_1 - C\psi\pi_1 - \pi_1 S}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + (V_{001} + V_{000})\ln C + (V_{101} + V_{100})\ln(1-C) \\
& + (N_1 - W_{11})\ln(C + \pi_1 - C\pi_1 - \pi_1 S) \\
& + (V_{111} + V_{110})\ln S + (V_{011} + V_{010})\ln(1-S) \\
& + W_{11}\ln\left[(1-C)(1-\pi_1) + \pi_1 S\right] \\
& + T_1\ln\pi_1 + (M_1 - T_1)\ln(1-\pi_1),
\end{aligned}
\tag{3.8}
$$

where ℓ_ψ is the log-likelihood in terms of ψ and the nuisance parameters $\boldsymbol{\eta} = (\pi_1, C, S)'$, given the observed data counts. For this particular model we cannot obtain closed-form MLEs, Karunaratne (1991), therefore, we use the a numerical method for finding extremes of functions.

3.4 Restricted Maximum Likelihood Estimation

In Section 3.3 we have derived the maximum likelihood estimators of the odds ratio and the nuisance parameters. However, a couple of the intervals we consider not only require evaluation at the MLEs but also at the restricted MLEs (RMLEs)

of the nuisance parameters.In this section we derive the RMLEs of the nuisance parameters evaluated at a fixed value of ψ.

First, we consider the incomplete (main) study and its cell counts as described in Table 3.1 that used only the fallible test as a classification procedure. Therefore, we assume we have misclassified data in the observed cell counts, W_{ik}. Also, suppose we performed the "gold standard" test on this data set. Then, we define a new set of variables, U_{ijk}, that are unobserved, because we have not actually performed the test. Again, i, j, and k have values 0 for non-diseased and 1 for diseased and correspond to the outcomes of the fallible test (Z), "gold standard" (X), and the actual disease condition (D), respectively, for each patient in the study. For details consider Table 3.2, where we see how the latent variables are distributed across the cells Notice that U_{ijk} are the unobserved portions of the observed W_{ik} counts.

Table 3.2: Counts for the main study with unobserved, misclassified data

	Main study (incomplete)			
Fallible	Cases (D=1)		Controls (D=0)	
(Z)	X=1	X=0	X=1	X=0
Z=1	U_{111}	$W_{11} - U_{111}$	U_{110}	$W_{10} - U_{110}$
Z=0	U_{011}	$W_{01} - U_{011}$	U_{010}	$W_{00} - U_{010}$
	$U_{111} + U_{011}$	$N_1 - (U_{111} + U_{011})$	$U_{110} + U_{010}$	$N_0 - (U_{110} + U_{010})$
	N_1		N_0	

Further, recall that the main study is a sub-sample of the complete study and, hence, we assume the same probability of misclassification. We can then write the distributions of the new latent variables as

$$(U_{11k} + U_{01k}) \sim Bin(N_k, \pi_k),$$

$$U_{11k}|(U_{11k} + U_{01k}) \sim Bin(U_{11k} + U_{01k}, S),$$

$$(W_{0k} - U_{01k})|(U_{11k} + U_{01k}) \sim Bin(N_k - (U_{11k} + U_{01k}), C),$$

where $k = 0$ represents the "controls" group and $k = 1$ represents the "cases" group, see Joseph et al. (1995). Also, recall that the S and C are the same for "cases" and "controls", which is the assumption of non-differential misclassification. Let $\boldsymbol{d}^{full} = (W_{00}, W_{01}, W_{11}, W_{10}, V_{111}, V_{001}, V_{110}, V_{000}, V_{011}, V_{101}, V_{010}, V_{100}, U_{111}, U_{011}, U_{110}, U_{010})'$ represent the full data vector including the latent variables. Now, we can write the likelihood function in terms of the latent variables as

$$
\begin{aligned}
L_U \propto & \begin{pmatrix} N_1 \\ U_{111}+U_{011} \end{pmatrix} \begin{pmatrix} U_{111}+U_{011} \\ U_{111} \end{pmatrix} \begin{pmatrix} N_1-(U_{111}+U_{011}) \\ W_{01}-U_{011} \end{pmatrix} \\
& \times \begin{pmatrix} N_0 \\ U_{110}+U_{010} \end{pmatrix} \begin{pmatrix} U_{110}+U_{010} \\ U_{110} \end{pmatrix} \begin{pmatrix} N_0-(U_{110}+U_{010}) \\ W_{00}-U_{010} \end{pmatrix} \\
& \times (\pi_1 S)^{U_{111}} [\pi_1(1-S)]^{U_{011}} (\pi_0 S)^{U_{110}} [\pi_0(1-S)]^{U_{010}} \\
& \times [(1-\pi_0)(1-C)]^{W_{10}-U_{110}} [(1-\pi_0)C]^{W_{00}-U_{010}} \\
& \times [(1-\pi_1)(1-C)]^{W_{11}-U_{111}} [(1-\pi_1)C]^{W_{01}-U_{011}} \\
& \times C^{V_{001}} (1-C)^{(M_1-T_1)-V_{001}} C^{V_{000}} (1-C)^{(M_0-T_0)-V_{000}} \\
& \times S^{V_{111}} (1-S)^{T_1-V_{111}} S^{V_{110}} (1-S)^{T_0-V_{110}} \\
& \times \pi_1^{T_1} (1-\pi_1)^{M_1-T_1} \pi_0^{T_0} (1-\pi_0)^{M_0-T_0},
\end{aligned}
\tag{3.9}
$$

where for ease of visualization we have omitted the constant terms. Refer to Appendix B.1 for a detailed derivation and simplification of (3.9).

Because we cannot derive closed-form solutions for the RMLEs, we use the EM (expectation-maximization) algorithm to determine the RMLEs for a fixed value of ψ. We consider the two steps in the algorithm individually.

E-step: First, outline at the expectation step. Here we are interested in deriving the conditional expected values for the latent variables $\mathbf{U} = (U_{111}, U_{011}, U_{110}, U_{010})'$, given the observed data and the current parameter values $\boldsymbol{\Phi}^{(r)} = (\psi, \pi_1^{(r)}, S^{(r)}, C^{(r)})'$, where r stands for the current iteration number. From the likelihood function (3.9)

and the derivations in B.1, the conditional distributions of the latent variables are

$$U_{111}|\boldsymbol{d}, \boldsymbol{\Phi}^{(r)} \sim Bin\left(W_{11}, \frac{\pi_1^{(r)} S}{\pi_1^{(r)} S^{(r)} + (1-\pi_1^{(r)})(1-C^{(r)})}\right),$$

$$U_{011}|\boldsymbol{d}, \boldsymbol{\Phi}^{(r)} \sim Bin\left(W_{01}, \frac{\pi_1^{(r)}(1-S^{(r)})}{\pi_1^{(r)}(1-S^{(r)}) + (1-\pi_1^{(r)})C^{(r)}}\right),$$

$$U_{110}|\boldsymbol{d}, \boldsymbol{\Phi}^{(r)} \sim Bin\left(W_{10}, \frac{\pi_1^{(r)} S^{(r)}}{\pi_1^{(r)} S^{(r)} + \psi(1-\pi_1^{(r)})(1-C^{(r)})}\right),$$

and

$$U_{010}|\boldsymbol{d}, \boldsymbol{\Phi}^{(r)} \sim Bin\left(W_{00}, \frac{\pi_1^{(r)}(1-S^{(r)})}{\pi_1^{(r)}(1-S^{(r)}) + \psi(1-\pi_1^{(r)})C^{(r)}}\right).$$

Thus, the conditional expectations of the unobserved variables are

$$U_{111}^* \equiv E[U_{111}|\boldsymbol{d}, \boldsymbol{\Phi}^{(r)}] = \frac{W_{11}\pi_1^{(r)} S^{(r)}}{\pi_1^{(r)} S^{(r)} + (1-\pi_1^{(r)})(1-C^{(r)})},$$

$$U_{011}^* \equiv E[U_{011}|\boldsymbol{d}, \boldsymbol{\Phi}^{(r)}] = \frac{W_{01}\pi_1^{(r)}(1-S^{(r)})}{\pi_1^{(r)}(1-S^{(r)}) + (1-\pi_1^{(r)})C^{(r)}},$$

$$U_{110}^* \equiv E[U_{110}|\boldsymbol{d}, \boldsymbol{\Phi}^{(r)}] = \frac{W_{10}\pi_1^{(r)} S^{(r)}}{\pi_1^{(r)} S^{(r)} + \psi(1-\pi_1^{(r)})(1-C^{(r)})},$$

and

$$U_{010}^* \equiv E[U_{010}|\boldsymbol{d}, \boldsymbol{\Phi}^{(r)}] = \frac{W_{00}\pi_1^{(r)}(1-S^{(r)})}{\pi_1^{(r)}(1-S^{(r)}) + \psi(1-\pi_1^{(r)})C^{(r)}}.$$

M-step: In the maximization step we update the parameter of interest in each iteration by using the solutions to the full data log-likelihood equations. First, define $\boldsymbol{d}^{full} = (W_{00}, W_{01}, W_{11}, W_{10}, V_{111}, V_{001}, V_{110}, V_{000}, V_{011}, V_{101}, V_{010}, V_{100}, U_{111}, U_{011}, U_{110}, U_{010})'$ as the full data vector including the latent variables. Also, let $\ell_U(\psi, \boldsymbol{\eta}|\boldsymbol{d}^{full})$ denote the log-likelihood of both the main and validation studies (full data). Then, in (3.9) we let $\pi_0 = \frac{\pi_1}{\pi_1 - \pi_1\psi + \psi}$ and group with respect to the nuisance parameters $\boldsymbol{\eta} = (\pi_1, S, C)'$. Hence, the conditional log-likelihood function is

$$
\begin{aligned}
\ell_U \propto\ & (N_0 + M_0)\ln\psi - (N_0 + M_0)\ln(\pi_1 - \pi_1\psi + \psi) \\
& + (N_0 + M_0 - U_{110} - U_{010} - T_0 + N_1 + M_1 - U_{111} - U_{011} - T_1)\ln(1 - \pi_1) \\
& + (U_{110} + U_{010} + T_0 + U_{111} + U_{011} + T_1)\ln\pi_1 \\
& + (V_{110} + U_{110} + V_{111} + U_{111})\ln S + (V_{010} + U_{010} + V_{011} + U_{011})\ln(1 - S) \\
& + (W_{10} + V_{100} - U_{110} + W_{11} + V_{101} - U_{111})\ln(1 - C) \\
& + (W_{00} + V_{000} - U_{010} + W_{01} + V_{001} - U_{011})\ln C.
\end{aligned}
$$

Therefore, the full-data estimating equations are

$$
\begin{aligned}
\frac{\partial \ell_U}{\partial \pi_1} \equiv & -\frac{W_{11} + W_{01} + M_1 - U^*_{111} - U^*_{011} - T_1 + W_{10} + W_{00} + M_0 - U^*_{110} - U^*_{010} - T_0}{1 - \pi_1} \\
& + \frac{U^*_{111} + U^*_{011} + T_1 + U^*_{110} + U^*_{010} + T_0}{\pi_1} - \frac{(W_{10} + W_{00} + M_0)(1 - \psi)}{\pi_1 - \pi_1\psi + \psi} = 0,
\end{aligned} \tag{3.10}
$$

$$
\begin{aligned}
\frac{\partial \ell_U}{\partial C} \equiv & \frac{W_{00} + V_{000} - U^*_{010} + W_{01} + V_{001} - U^*_{011}}{C} \\
& - \frac{W_{10} + M_0 - T_0 - V_{000} - U^*_{110} + W_{11} + M_1 - T_1 - V_{001} - U^*_{111}}{1 - C} = 0,
\end{aligned} \tag{3.11}
$$

and

$$
\frac{\partial \ell_U}{\partial S} \equiv \frac{V_{110} + U^*_{110} + V_{111} + U^*_{111}}{S} - \frac{T_0 - V_{110} + U^*_{010} + T_1 - V_{111} + U^*_{011}}{1 - S} = 0. \tag{3.12}
$$

Solving (3.10) - (3.12) for the r^{th} iteration, we have

$$
\pi_1^{(r+1)} = \frac{B - \sqrt{B^2 - 4AC}}{2A},
$$

where

$$
\begin{aligned}
A &= (\psi - 1)(M_1 + W_{01} + W_{11}), \\
B &= M_0 + W_{00} + W_{10} + \psi(M_1 + W_{01} + W_{11}) \\
&\quad + (\psi - 1)(T_0 + T_1 + U^*_{010} + U^*_{011} + U^*_{110} + U^*_{111}),
\end{aligned}
$$

and

$$
C = \psi(T_0 + T_1 + U^*_{010} + U^*_{011} + U^*_{110} + U^*_{111}),
$$

so that

$$C^{(r+1)} = \frac{W_{00} + V_{000} - U^*_{010} + W_{01} + V_{001} - U^*_{011}}{W_{00} + W_{10} + M_0 - T_0 - U^*_{010} - U^*_{110} + W_{01} + W_{11} + M_1 - T_1 - U^*_{011} - U^*_{111}},$$

and

$$S^{(r+1)} = \frac{V_{110} + U^*_{110} + V_{111} + U^*_{111}}{T_0 + U^*_{010} + U^*_{110} + T_0 + U^*_{011} + U^*_{111}}.$$

3.5 The Confidence Intervals

Here we consider four confidence intervals for the odds ratio parameter: the Wald CI, the score CI, the profile likelihood CI, and the approximate integrated (AI) likelihood CI. Below we derive and describe each one of CIs based on the results from the previous sections.

3.5.1 The Observed Information Matrix

The Wald and the score confidence intervals are likelihood-based intervals. For their construction we have to calculate or at least approximate the variance of the MLE of interest. In the multivariate parameter case, in order to estimate the variance we often use the observed information matrix. Let us denote the vector of parameters as $\boldsymbol{\theta} = (\psi, \boldsymbol{\eta}')'$, where ψ is the odds ratio and the parameter of interest and $\boldsymbol{\eta}$ is the nuisance parameter vector. Recall that we have three nuisance parameters because we are considering the non-differential case of misclassification with equal sensitivities and specificities. Then, we have $\boldsymbol{\eta} = (\pi_1, S, C)'$, where S is the common sensitivity and C is the common specificity. Let

$$\boldsymbol{J}(\psi, \boldsymbol{\eta}) = -\begin{bmatrix} \frac{\partial^2 \ell}{\partial \psi^2} & \frac{\partial^2 \ell}{\partial \psi \partial \pi_1} & \frac{\partial^2 \ell}{\partial \psi \partial S} & \frac{\partial^2 \ell}{\partial \psi \partial C} \\ \cdot & \frac{\partial^2 \ell}{\partial \pi_1^2} & \frac{\partial^2 \ell}{\partial \pi_1 \partial S} & \frac{\partial^2 \ell}{\partial \pi_1 \partial C} \\ \cdot & \cdot & \frac{\partial^2 \ell}{\partial S^2} & \frac{\partial^2 \ell}{\partial S \partial C} \\ \cdot & \cdot & \cdot & \frac{\partial^2 \ell}{\partial C^2} \end{bmatrix}, \tag{3.13}$$

denote the observed information matrix where ℓ is the log likelihood function given in (3.8). In Appendix B.2 we include the formulas for each of the terms in the matrix (3.13).

3.5.2 A Wald CI

First, we consider the Wald CI. Let W denote the statistic. Then,

$$W = (\hat{\psi} - \psi)^2 [J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]^{-1},$$

where

$$\boldsymbol{J}(\psi, \boldsymbol{\eta}) = \begin{bmatrix} J_{11} & \boldsymbol{J}_{12} \\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22} \end{bmatrix}, \tag{3.14}$$

is the partitioning of the observed information matrix and $\hat{\psi}$ and $\hat{\boldsymbol{\eta}}$ are the MLEs of ψ and $\boldsymbol{\eta}$ respectively and $J^{11} = (J_{11} - \boldsymbol{J}_{12}\boldsymbol{J}_{22}^{-1}\boldsymbol{J}_{21})^{-1}$. Hence, an approximate $100(1-\alpha)\%$ confidence interval for the odds ratio are the values of ψ that satisfy

$$(\hat{\psi} - \psi)^2 [J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]^{-1} < \chi_1^2(1-\alpha), \tag{3.15}$$

where $\chi_1^2(1-\alpha)$ denotes the $1-\alpha$ quantile of a central chi-squared distribution with one degree of freedom. Notice that in (3.15) $\boldsymbol{J}^{11}$ is evaluated at the MLEs (as derived in Section 3.3) for both the odds ratio and the nuisance parameters. Thus, the Wald CI is

$$\hat{\psi} - \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]} < \psi < \hat{\psi} + \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]}. \tag{3.16}$$

3.5.3 A Score CI

An alternative to the Wald CI and the score CI which is also a maximum likelihood-based interval. The score interval is based on the score statistic which is

$$Sc = \left[S(\psi, \hat{\boldsymbol{\eta}}_\psi)\right]^2 \left[J^{11}(\psi, \hat{\boldsymbol{\eta}}_\psi)\right] \dot{\sim} \chi_p^2,$$

where

$$S(\boldsymbol{\theta}) \equiv \frac{\partial}{\partial \boldsymbol{\theta}} \ln(L(\boldsymbol{\theta}))$$

is the score function and $\boldsymbol{\theta} = (\psi, \boldsymbol{\eta}')'$ is the vector of nuisance parameters and parameter of interest. An approximate $100(1-\alpha)\%$ score interval is the values of ψ that satisfy

$$\left[S(\psi, \hat{\boldsymbol{\eta}}_\psi)\right]^2 \left[J^{11}(\psi, \hat{\boldsymbol{\eta}}_\psi)\right] < \chi_1^2(1-\alpha), \tag{3.17}$$

where $\hat{\boldsymbol{\eta}}_\psi$ is the restricted MLE (RMLE) of the nuisance parameter and $\chi^2_1(1-\alpha)$ denotes the $1-\alpha$ quantile of a central chi-squared distribution with one degree of freedom (see Riggs (2006)). For this interval the observed information matrix is evaluated at the RMLEs of the nuisance parameters for a fixed value of ψ. To determine the limits of the score CI for the odds ratio we use the bisectional method. However, because we cannot directly solve for ψ, we use the EM algorithm in Section 3.4 to help determine those limits.

3.5.4 A Profile Likelihood CI

The third confidence interval we consider in this paper is the profile likelihood interval. It is pseudo-likelihood based and it is believed to give better results than the usual basic Wald and score intervals. To eliminate the nuisance parameters we replace them by their RMLEs in the likelihood function. The profile likelihood function is, therefore,

$$L_P(\psi) \equiv \max_{\boldsymbol{\eta}} L(\psi, \boldsymbol{\eta}) = L(\psi, \hat{\boldsymbol{\eta}}_\psi), \tag{3.18}$$

where $\hat{\boldsymbol{\eta}}_\psi$ is the vector of RMLEs in terms of the parameter of interest ψ. Again we have no closed form for the RMLEs therefore, we use an EM algorithm to compute the RMLEs (see Section 3.4). Thus, an approximate $100(1-\alpha)\%$ profile likelihood CI for the odds ratio is the set of values of ψ that satisfy

$$-2\left[\ell(\psi, \hat{\boldsymbol{\eta}}_\psi) - \ell(\hat{\psi}, \hat{\boldsymbol{\eta}})\right] < \chi^2_1(1-\alpha), \tag{3.19}$$

where $\chi^2_1(1-\alpha)$ denotes the $1-\alpha$ quantile of a central chi-squared distribution with one degree of freedom, $\ell(\hat{\psi}, \hat{\boldsymbol{\eta}})$ is the log-likelihood function evaluated at the MLEs (see Section 3.3), and $\ell(\psi, \hat{\boldsymbol{\eta}}_\psi)$ is the log-likelihood function evaluated at the RMLEs. The profile likelihood CI procedure has a major flaw in that it does not account for the uncertainty in the nuisance parameter estimation. This flaw can lead to optimistically precise estimates for the odds ratio parameter, see Riggs (2006).

3.5.5 An Approximate Integrated Likelihood CI

The last interval we consider is the approximate integrated interval, which is also a pseudo-likelihood based CI. The procedure of calculating this interval involves integration to eliminate the nuisance parameters. It is usually an improvement over the profile likelihood method and yields closer estimates of the odds ratio without ignoring the uncertainty in the nuisance parameters. However, often the integration is not a feasible task and we have to resort to different methods such as the Laplace approximation method. An approximate integrated likelihood function is

$$L_{AI}(\psi) = \int L(\psi, \boldsymbol{\eta}) d\boldsymbol{\eta} \approx \frac{cL_P(\psi)}{\left|\hat{\boldsymbol{J}}_{\boldsymbol{\eta}}(\psi, \hat{\boldsymbol{\eta}})\right|^{1/2}}, \tag{3.20}$$

where $c = (2\pi)^{\nu/2}$, ν is the dimension of the nuisance parameter, and $L_P(\psi)$ is the profile likelihood as defined in (3.18). The notation $\boldsymbol{J}_{\boldsymbol{\eta}}$ represent the nuisance parameter sub-matrix of the observed information matrix. Hence, from (3.14) we have $\hat{\boldsymbol{J}}_{\boldsymbol{\eta}}$ is $\boldsymbol{J}_{22}$ evaluated at the MLEs of the nuisance parameters. This sub-matrix is calculated using the formulas for the MLEs from Section 3.3 and the information matrix terms included in Appendix B.2. A $100(1-\alpha)\%$ approximate integrated likelihood CI for the odds ratio is the set of values of ψ that satisfy

$$-2\left[\ell_{AI}(\psi) - \ell_{AI}(\hat{\psi}_{AI})\right] < \chi^2_1(1-\alpha), \tag{3.21}$$

where $\ell_{AI}(\psi)$ represent the approximate integrated log-likelihood as defined in (3.20) evaluated at a fixed value of ψ and $\ell_{AI}(\hat{\psi}_{AI})$ is the log-likelihood of (3.20).

3.6 A Monte Carlo Simulation Study

In this paper we aim to estimate the odds ratio in a case-control study by implementing a double sampling procedure and assuming non-differential misclassification between the cases and controls. We considered four confidence intervals that differed from each other by the method used to eliminate the nuisance parameters. In this section we describe a simulation study to compare the performance

of these intervals in terms of coverage and interval width. We use a Monte Carlo simulation to examine the effect of the sample sizes of the cases and controls ($k = 0$ or 1 respectively) for the complete, M_k, and the incomplete studies, N_k. Also, we are interested in the effect of the disease probability π_j, for $j = 0, 1$ for the "gold standard" test outcome $X = 0, 1$, and the magnitude of the odds ratio , ψ, on the coverage and interval width of the CIs defined in (3.16) - (3.21).

3.6.1 Parameter and Sample Size Configurations

We considered two values for the odds ratio parameter $\psi = 2$ and $\psi = 4$ with the corresponding probabilities shown in Table 3.3. The probability π_0 is when the

Table 3.3: Odds ratio and probabilities for simulation in CCNDIFF

ψ	π_0	π_1
2	0.25	0.40
4	0.38	0.71

"gold standard" test indicates a diseased individual ($X = 1$) but the true condition is non-diseased ($D = 0$), i.e. $\pi_0 \equiv Pr(X = 1|D = 0)$. Then the probability of being truly diseased and the non-fallible test indicating that one is indeed diseased is represented by $\pi_1 \equiv Pr(X = 1|D = 1)$. Recall that we are only interested in the case of non-differential misclassification, i.e the specificities and sensitivities are the same for "cases" and "controls". Often the misclassification is defined in three categories – low, medium, and high. In our simulation we only consider the low and high with the values given in Table 3.4.

Again, specificity is defined as $C = Pr(Z = 0|X = 0)$ and sensitivity is $S = Pr(Z = 1|X = 1)$ regardless of the given true status of the participant. To obtain reasonable results for comparisons we derive CIs for eight different sample sizes for low and high misclassification and for $\psi = 2$ or 4. By definition we know

Table 3.4: Specificity and sensitivity for CCNDIFF

Misclassification	Specificity (C)	Sensitivity (S)
Low	0.99	0.98
High	0.85	0.75

that the complete sample sizes are much smaller than the fallible sample sizes due to factors discussed previously. We give the sample size values we simulate in Table 3.5.

Table 3.5: Sample sizes used in the simulation for CCNDIFF

	A1	A2	A3	A4	A5	A6	A7	A8
M_0	50	50	75	100	125	150	175	200
M_1	30	30	37	50	62	75	87	100
N_0	250	500	750	1000	1250	1500	1750	2000
N_1	120	250	370	500	620	750	870	1000

and summarize the parameter configuration for this simulation as $\psi \in \{2, 4\}$, $\pi_1 \in \{0.40, 0.71\}$, $S \in \{0.98, 0.75\}$, and $C \in \{0.99, 0.85\}$. We generated 10,000 data sets under each combination of conditions and calculate the 95% confidence intervals for the binomial odds ratio under double sampling and non-differential misclassification in case-control studies for each of the CI methods described in this chapter.

3.6.2 Results

We first consider the scenario with low non-differential misclassification and an assumed odds ratio parameter is $\psi = 2$. Refer to Table 3.4 to recall the appropriate misclassification probabilities. In this case we encounter a problem when using the score interval for estimating the parameter. Hence, in Figure 3.1 and Figure 3.2, we see results for only the three other CIs. Figure 3.1 gives the coverage of each of the intervals and Figure 3.2 shows the boxplots of the interval widths.

Consider the Wald interval first. From Figure 3.1 we see that the Wald interval overestimated the desired 95% confidence level for most of the sample size scenarios, $A1$ to $A8$ (Table 3.5). Nevertheless, the interval showed significant improvement

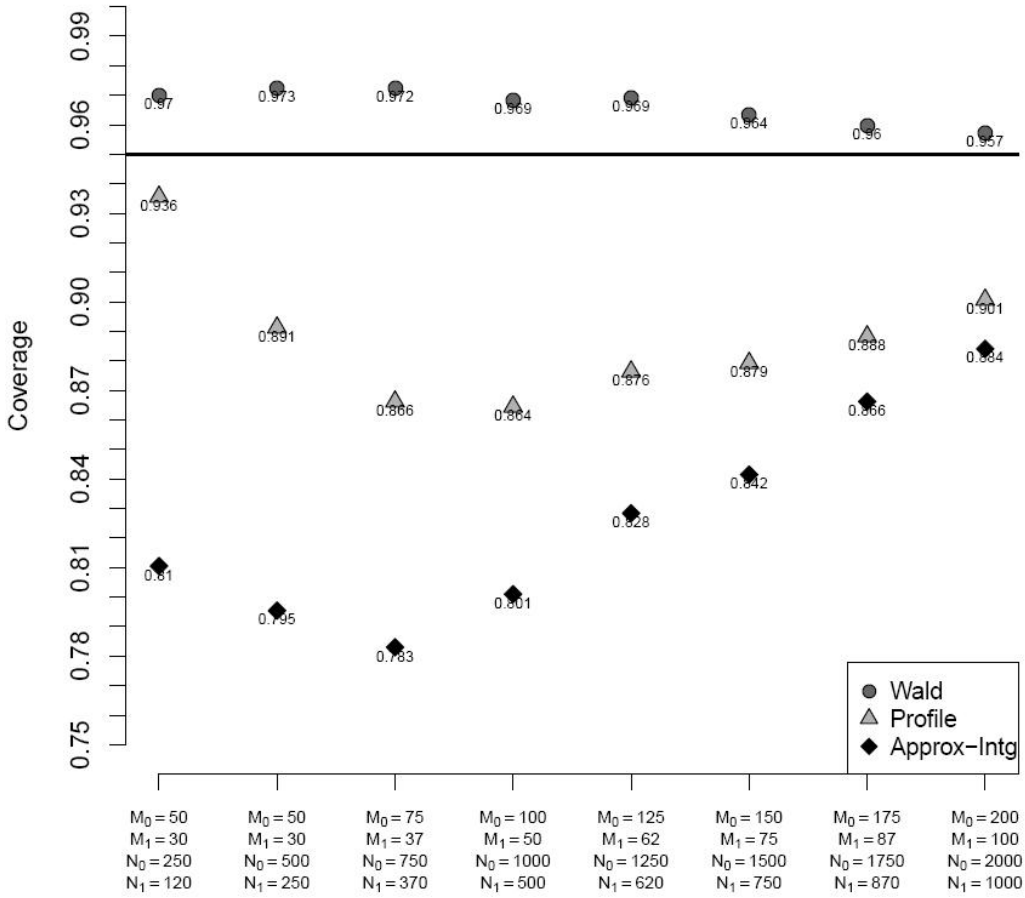

Figure 3.1: Coverage probabilities of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with low nondifferential misclassification and an odds ratio of $\psi = 2$

towards the desired confidence level when the sample size increased. This statement is consistent with the results presented in Figure 3.2. The Wald interval had a higher median and mean interval width compared to the profile likelihood and the approximate-integrated likelihood intervals. This property caused over-coverage of the Wald CI. However, notice that for the larger sample size values the average

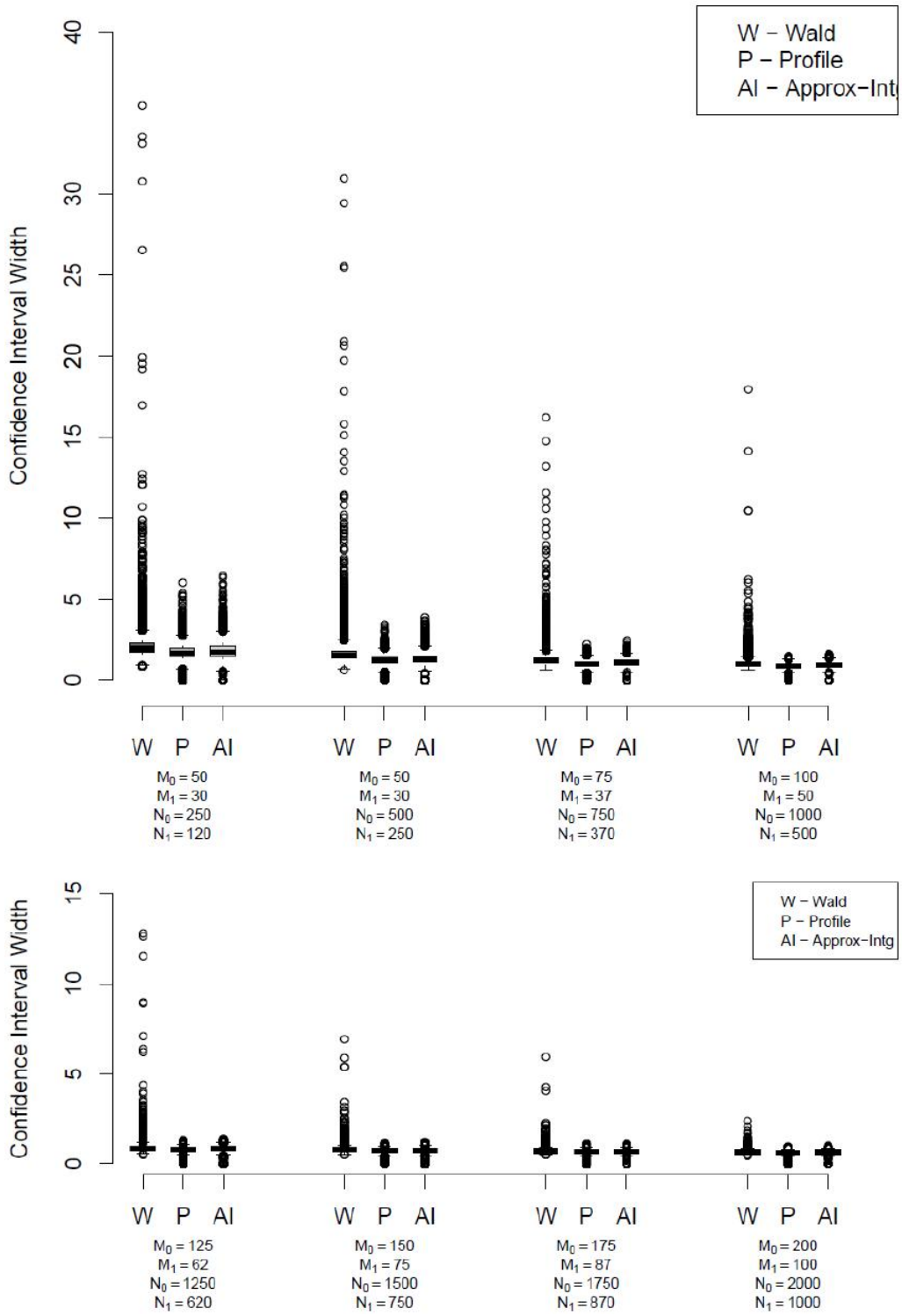

Figure 3.2: Interval widths of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with low nondifferential misclassification and an odds ratio of $\psi = 2$

widths of the Wald CIs were comparable to the interval width characteristics of the other two intervals.

Next, we discuss the profile likelihood and the approximate-integrated likelihood CIs. They both follow a very similar pattern and the resulting coverage and interval width characteristics are essentially the same. In Figure 3.1 we noticed that the profile likelihood interval started with good coverage properties. However, with the increase of the sample size they worsened and then improved once we reached the larger sample sizes. The same behavior is seen with the approximate-integrated interval. Nevertheless, neither the profile nor the AI CIs achieved the desired 95% confidence level because they both greatly underestimate the desired coverage. This shortcoming can be explained by the boxplots shown in Figure 3.2. The intervals widths are extremely small thus yielding very tight CIs and, hence, lower probability of capturing the true parameter value.

Next, we review the scenario with low nondifferential misclassification (Table 3.4) and odds ratio parameter $\psi = 4$. In this case we were able to calculate the score interval.

First, we consider the Wald interval coverage shown in Figure 3.3. The Wald CI was extremely conservative and overestimated the desired 95% confidence level throughout all the given sample sizes. Nevertheless, with an increased sample size, we noticed a great improvement of the coverage of this interval. The boxplots presented in Figure 3.4 showed that the average interval width is similar to the one of the pseudo-likelihood intervals. No obvious evidence existed to show a reason for the overestimation of the Wald interval in this case.

Next, we examined the score interval. Consider the coverage plot in Figure 3.3. The behavior of the score interval was extremely different than the rest of the intervals. Notice that the increased sample size actually significantly decreased the probability of the score interval capturing the true parameter. The interval started

highly conservative and ended highly under-covering. This behavior is supported by the boxplots in Figure 3.4. Notice in this figure that the score interval had a much higher average width than the other three intervals for the smaller sample size cases.

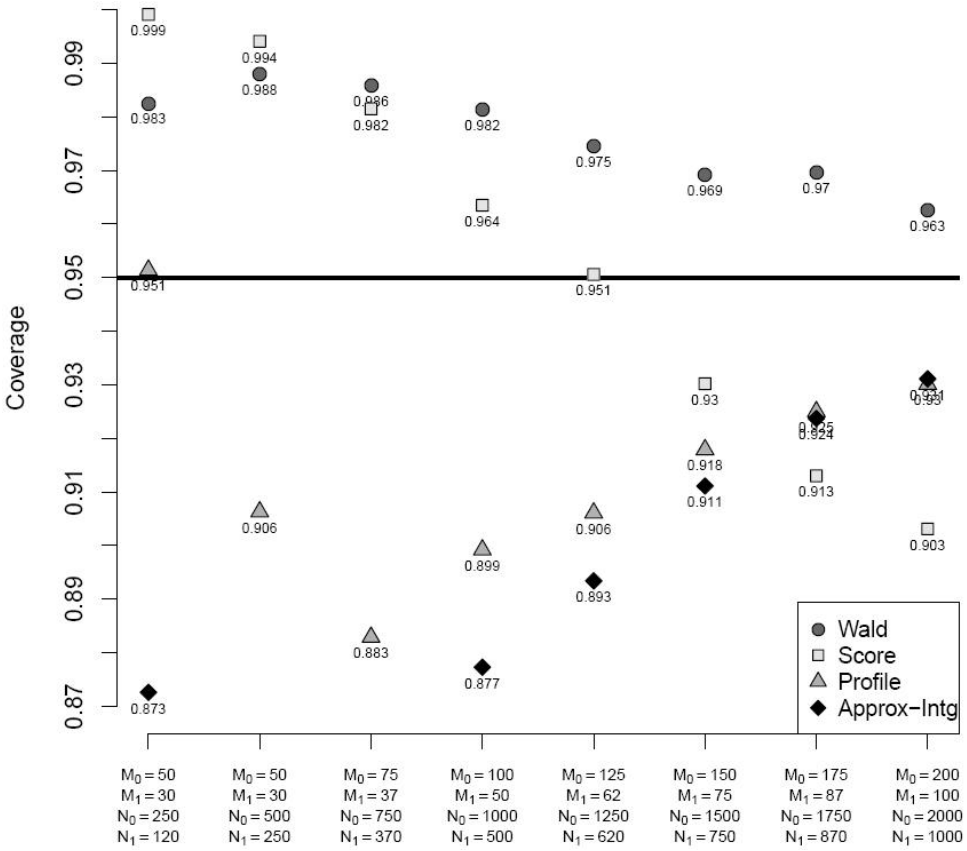

Figure 3.3: Coverage probabilities of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with low nondifferential misclassification and odds ratio of $\psi = 4$

This fact explains why the score interval overestimated at the beginning. However, we see that with the change of sample size values from $A1$ to $A8$, the mean width and interquartile range of the score interval decreased quickly, thus supporting the behavior we observed in the coverage plot in Figure 3.3.

The next interval we considered is the profile likelihood interval. The shape of the coverage curve for the different sample sizes was similar to the coverage plot

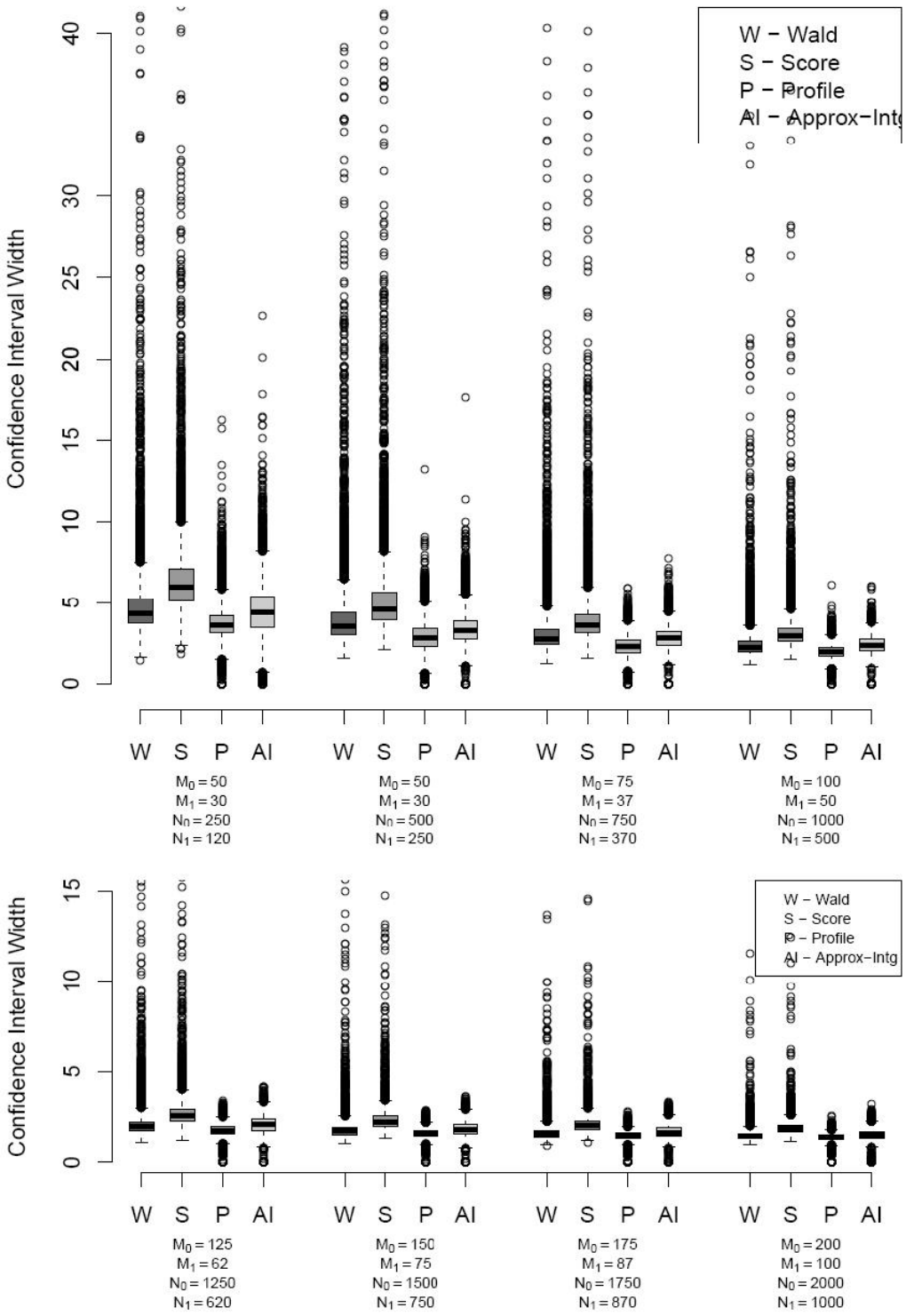

Figure 3.4: Interval widths of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with low nondifferential misclassification and an odds ratio of $\psi = 4$

for the $\psi = 2$ case. The intervals started with almost perfect coverage for the smallest sample size and then decreased coverage for the sample size scenario $A3$, and improved coverage going from $A4$ to $A8$ (Table 3.5). In Figure 3.4 we see that the profile likelihood interval had a much smaller interquartile range of the interval widths throughout all sample sizes, hence causing undercoverage.

The last interval examined here is the approximate-integrated interval. The coverage probabilities for the AI CI followed a similar pattern to the profile-likelihood interval coverage properties. Notice that although both intervals underestimated the desired confidence level of 95%, they both showed great improvement when the sample size was increased.

The next simulation scenario presented in this chapter is the case of high nondifferential misclassification and odds ratio parameter value of $\psi = 2$. The coverage and boxplots interval width of the four competing confidence intervals are shown in Figures 3.3 and Figure 3.4, respectively.

Here we combine the discussion for the Wald, profile likelihood, and approximate-integrated likelihood intervals since they have almost identical coverage and follow the same coverage pattern. The Wald interval started slightly underestimating; however, with an increase in sample size, it approached the desired 95% confidence level. The profile-likelihood and the approximate-integrated likelihood intervals stayed consistently on the 95% confidence level, making them great CI estimators for the odds ratio parameter. In Figure 3.3 we see that these three intervals had very similar confidence interval widths. The widths became gradually smaller with the increase of sample size. The interval width boxplots were consistent with the conclusion that these three intervals are excellent interval estimators of ψ.

Next, we consider the score interval. In Figure 3.3 we see that the coverage of this interval became increasingly worse with the increase of the sample size. This behavior might be explained by the width boxplots shown in Figure 3.4. We notice

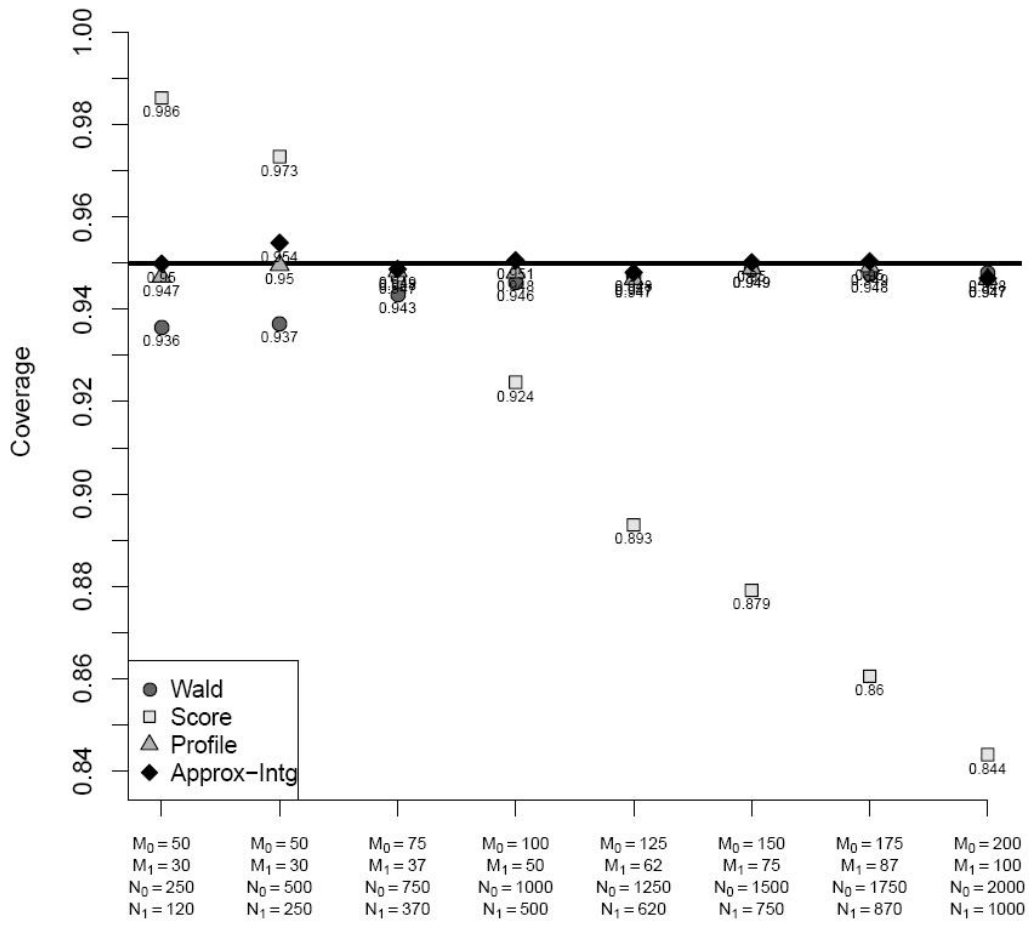

Figure 3.5: Coverage probabilities of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with high nondifferential misclassification and odds ratio of $\psi = 2$

that the average width of the score interval was greater than the competing intervals at the smaller sample size examples, thus explaining the fact that the interval overestimates the coverage for greater sample sizes. However, the score coverage was very different from the coverage of the competing intervals. This fact led us to conclude that this particular interval should be further investigated.

The last scenario we examined was the case of high nondifferential misclassification and odds ratio parameter $\psi = 4$. The results are shown in Figure 3.7 for the coverage of the four intervals of interest and in Figure 3.8 for the confidence interval widths.

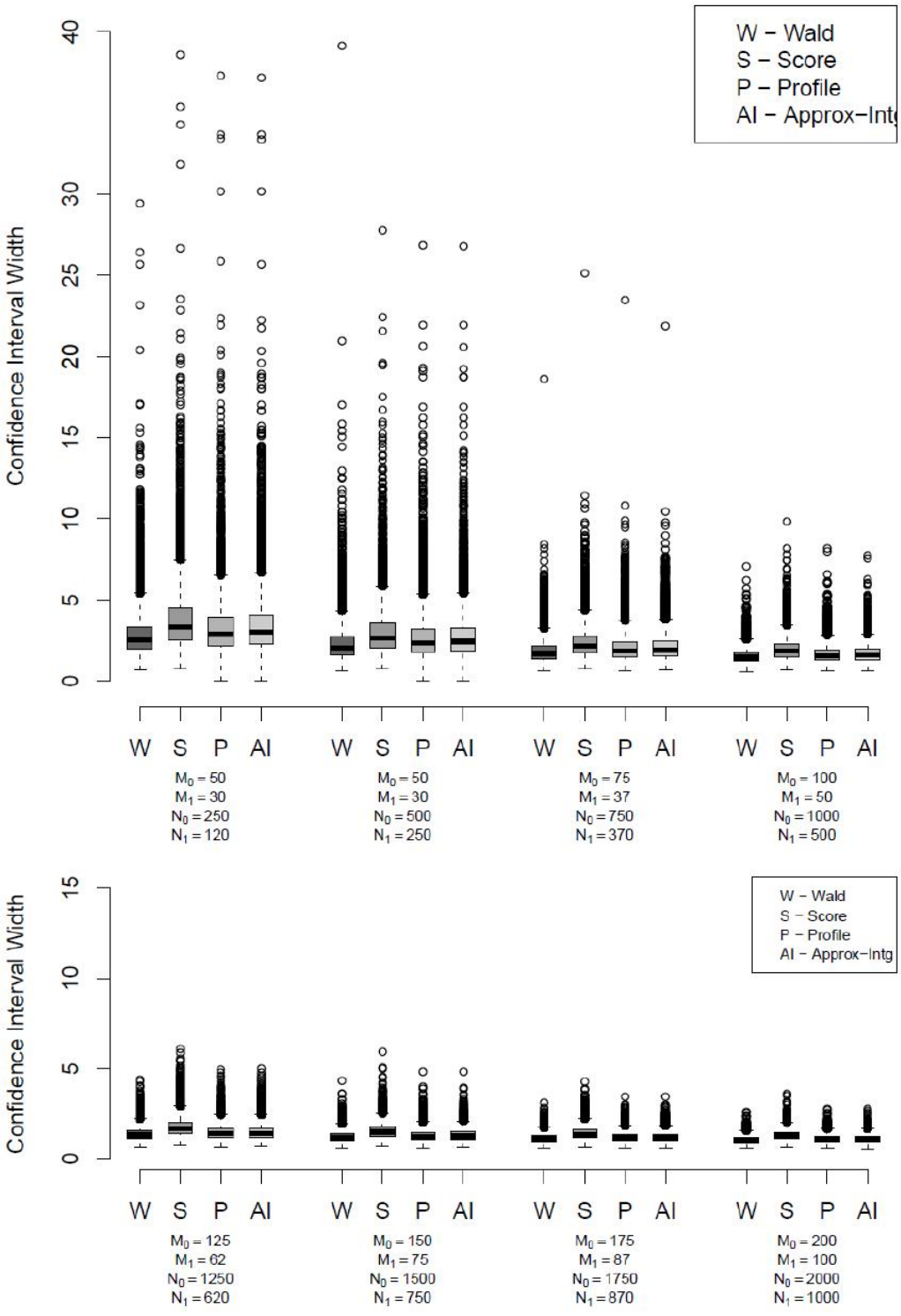

Figure 3.6: Interval widths of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with high nondifferential misclassification and an odds ratio of $\psi = 2$

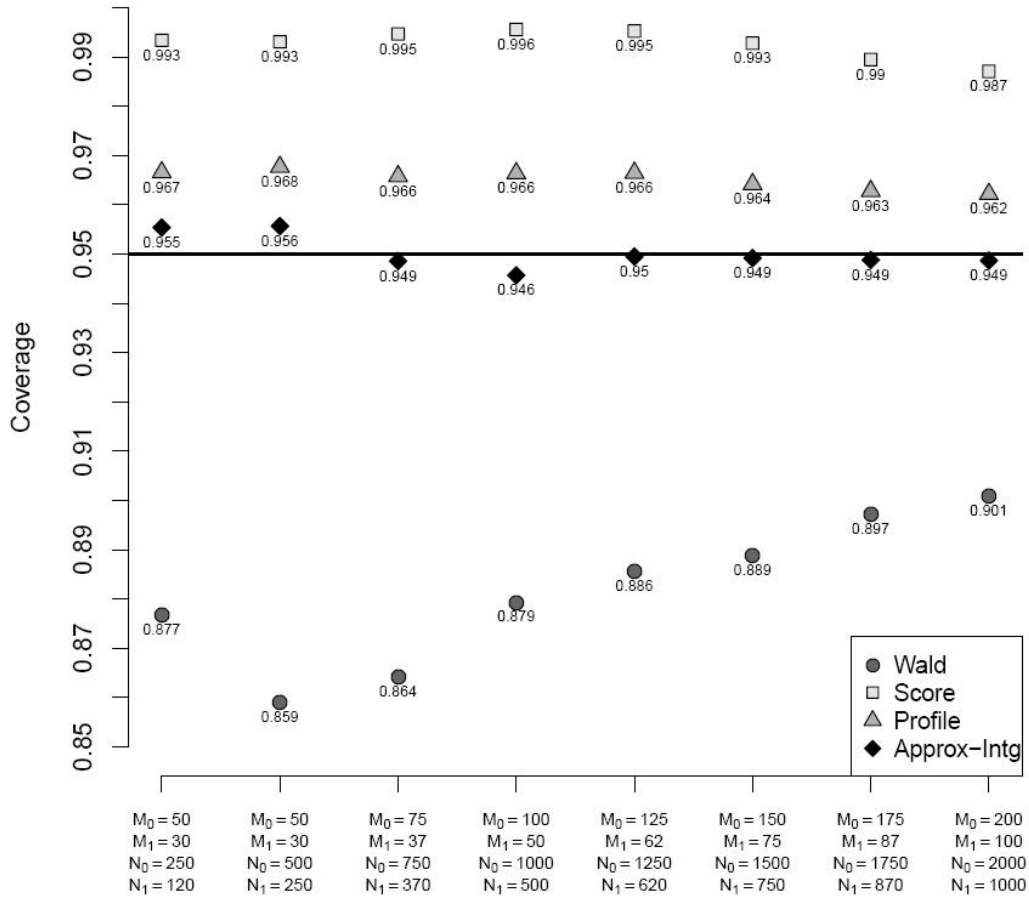

Figure 3.7: Coverage probabilities of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with high nondifferential misclassification and odds ratio of $\psi = 4$

The first interval we observed was the Wald CI. Notice this was the only case where the CI underestimated the coverage throughout all the possible sample sizes. However, the Wald CI improved with an increase of sample size values. In Figure 3.8 we see that this observation was supported by the boxplots of the CI widths.

The Wald interval had consistently smaller average and median interval widths in all of the sample size scenarios. Certainly, the variability decreased with the increase in N; nevertheless, the Wald interval was much smaller than the competing interval widths, making it difficult to capture the true parameter value.

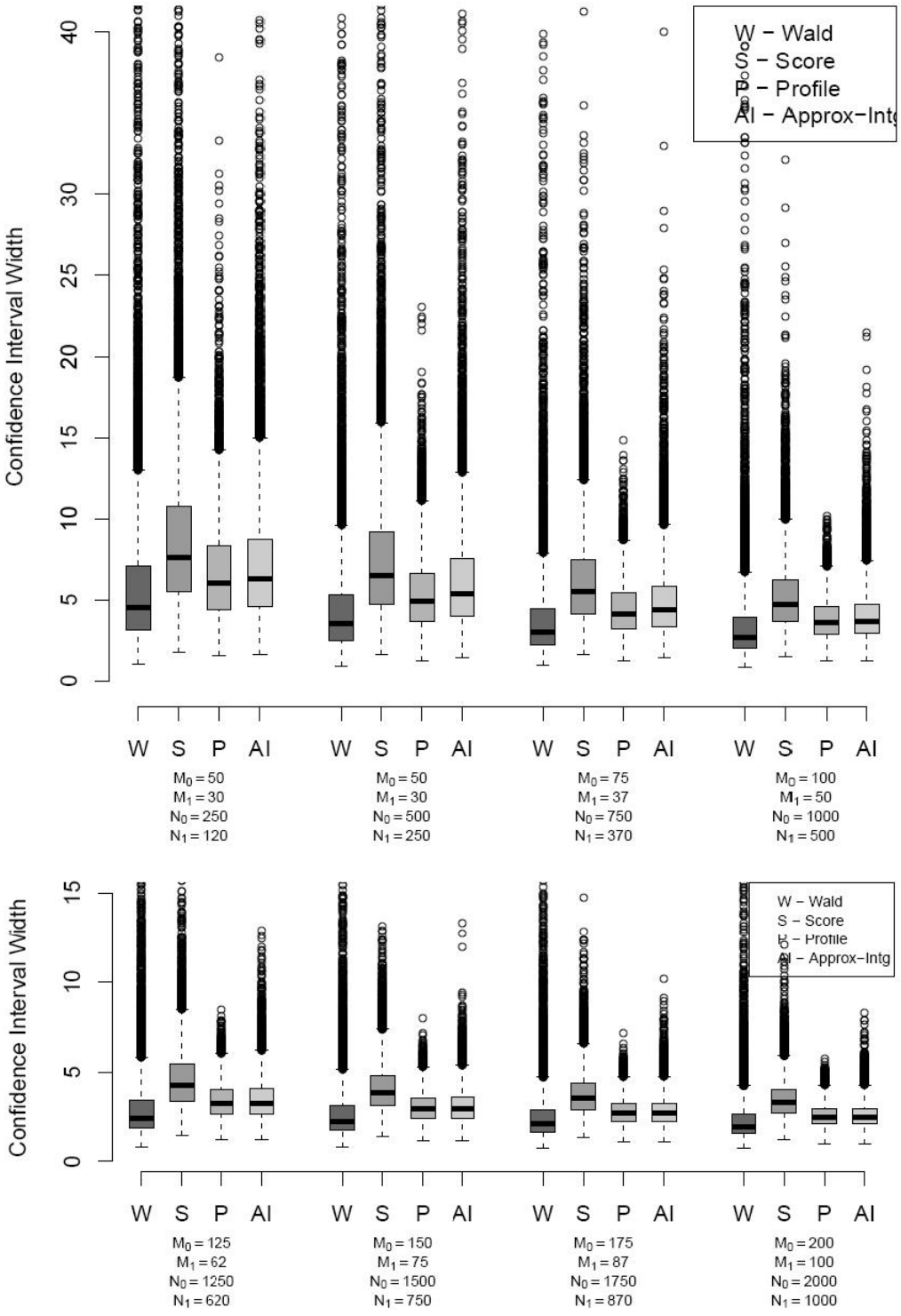

Figure 3.8: Interval widths of the Wald, score, profile, and approximate integrated likelihood CIs for the case-control study with high nondifferential misclassification and an odds ratio of $\psi = 4$

Next, we discuss the score interval. We see in Figure 3.7 that the score interval was very conservative, and an increase of sample size did not affect the coverage probability. Again, this fact was supported by the results shown in Figure 3.8. The score interval had a much greater median width value than the rest of the intervals. Thus, the score intervals were wider and allowed for easier capturing of the true odds ratio parameter. An increase of the sample size decreased the variability. However, an increase in sample size had almost no effect on the coverage of the interval.

The profile likelihood interval was an improvement compared to both the Wald and the score intervals in terms of coverage. Nevertheless, it was also consistently conservative, and an increase in sample size did not affect the coverage properties.

Last, we discuss the approximate-integrated CI. Notice in Figure 3.7 that this interval had almost perfect coverage throughout all of the sample size scenarios. With an increase of the sample size, we saw a decrease in variability and the coverage levels very close to 95%, which was the exact confidence level we desired. Also, in Figure 3.8 we see that the approximate-integrated intervals mean and median widths shrunk steadily with an increase of the sample size. Hence, the AI CI is an excellent estimator for the odds ratio under the parameter scenario.

3.7 Comments

In conclusions, the different scenarios and combinations of low and high misclassification, $\psi = 2, 4$, and the sample sizes from $A1$ to $A8$ showed very inconsistent results. First, we considered the case of low nondifferential misclassification (recall Table 3.4). The Wald interval was consistently overestimating and the profile likelihood and approximate-integrated likelihood intervals were underestimating. None of the intervals showed great potential as an omnibus interval for estimating ψ.

Then, we considered the case of high nondifferential misclassification. Recall that high probability of misclassification gives more information about the errors

and, hence, leads to better estimates. The results presented in this section showed that the approximate-integrated interval is an extremely good estimator and gives almost exact coverage. The competing intervals performed differently for the different odds ratio values, and they cannot be ordered in terms of coverage properties.

The consistent conclusion, however, was that the score interval behaved extremely differently from the competing intervals but also behaved differently for each of the combination scenarios. Hence, a further investigation on this interval is warranted.

CHAPTER FOUR

Simulation Study on Bioequivalence Tests Under Several Variability Conditions

4.1 Introduction

Pharmaceutical manufacturing has become one of the most important industries in the world. Bioequivalence (BE) studies play an integral role in the new drug development. The United States Food and Drug Administration (FDA) defines bioequivalence as "the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study", FDA (2003). Also, Birkett (2003) has stated that "two pharmaceutical products are bioequivalent if they are pharmaceutically equivalent and their bioavailabilities (rate and extent of availability) after administration in the same molar dose are similar to such a degree that their effects, with respect to both efficacy and safety, can be expected to be essentially the same. Pharmaceutical equivalence implies the same amount of the same active substance(s), in the same dosage form, for the same route of administration and meeting the same or comparable standards." Essentially, both definitions state the same basic idea. We call products bioequivalent when they are expected to perform the same and can be interchanged.

BE studies have been used in the pharmaceutical industry for several years. The main reason is for the approval of new generic drugs. When a reference-listed drug on the market is too expensive for the general population, companies often attempt to develop a generic version that is more accessible. In such instances, the FDA requests a BE study to determine the rate and extent of absorption of each

therapeutic moiety for the test drug and reference drug. In addition, BE studies are employed to test new formulations for old drug products or new dosages or inactive ingredients.

BE studies are not dependent on the actual outcome of a clinical trial, but only on the rate and extent of availability of the tested product, and are generally conducted in healthy populations. Male and female adults are given the drug under a standardized condition and are monitored throughout the trial. Most BE studies are conducted on the highest strength of a product line. In some cases, however, the BE study must be conducted on diseased patients for safety reasons, Davit, Nwakama, Buehler, Conner, Haidar, Patel, Yang, Yu, and Woodcock (2009).

An interesting question that arises about BE testing is, "What happens with a new product if it does not meet the BE criteria?" BE studies are performed in Phase III of the drug development process, and usually by that time, the pharmaceutical company has invested considerable time and resources into the new product. Hence, the pharmaceutical company is interested in marketing the drug. Because BE studies are generally conducted on a very small group of people, a common suggestion for a solution to the failed BE case is to increase the number of subjects participating in the study and to thereby narrow the CI used for testing BE, see Tothfalusi, Endrenyi, Midha, Rawson, and Hubbard (2001).

In this paper we investigate whether or not the increase of sample size is enough of an adjustment to alter the results of a failed BE test. Our hypothesis is that several different types of variability affect the outcome of the BE test and, thus, the increase in sample size is not always a solution. Here we extensively investigate the variables that are known to impact the BE test. Such variables are within-subject variability, assumed mean ratio difference between treatment and reference products, and sample size. By design, the overall variability of a crossover design is not affected by the between-subject variability. Nevertheless, we consider the

between-subject variability in our paper because the BE studies are now conducted on a wide range of people, Midha, Rawson, and Hubbard (1997).

This chapter is organized as follows. First, in Section 4.2 we present some background information on the BE testing procedures and variables of interest. We also present a derivation of the geometric mean ratio confidence interval as well as a step-by-step approach for reaching a conclusion in a BE trial. In Section 4.3 we discuss the study design and the types of variability parameters we consider in our investigation. We next explain the simulation set-up and the parameter configurations we consider. We also discuss the statistical methods for deciding on the BE test outcome and presenting the results. In Section 4.4 we give the results and our conclusions on the effect of each variable of interest on the BE test outcome. In Section 4.5 we give overall comments.

4.2 Background

Bioequivalence studies are designed to assess "the absence of a significant difference" (FDA (2003)) between a well-established product and a new generic or investigational test drug. We denote the reference product by R and the test treatment or generic formulation by T.

4.2.1 BE Study Design

The type of BE design usually is related to the type of products tested. The two main types of BE designs are parallel-group and crossover. In this paper we examine the crossover design with two sequences and two periods because the FDA generally asks applicants to conduct BE studies with pharmacokinetic endpoints using a single-dose crossover design, see Haidar et al. (2008a). A graphic representation of the crossover design is shown in Figure 4.1. Subjects are usually healthy and receive a single dose of T or R products on separate occasions with random assignment to the two possible sequences of product administration. Such design is preferred by

the FDA because it is proven to be more sensitive to detecting potential differences between products, FDA (2003). Recall that in a crossover design, subjects are randomly assigned to receive a sequence of treatments that contain all the treatments in the study, Chow and Wang (2001). Hence, in our case, subjects are assigned to receive one of two sequences of treatments – TR or RT – where TR indicates that the test drug is administered first in the first period and the reference drug is given second in the second period and the reverse for the RT sequence. Because each subject receives both treatments, we need to ensure that the drugs do not interact. That goal is achieved by allowing an adequate duration washout period between drug administrations such that the drug of interest cannot be detected in the subjects' plasma. Based on the length of the washout period, BE studies implementing this design can be fairly quick. Another advantage of this design is the use of fewer subjects because each subject is given both drugs. Also, we can use each subject as its own control, which allows us to compare the within-subject variability (WSV) between the treatments.

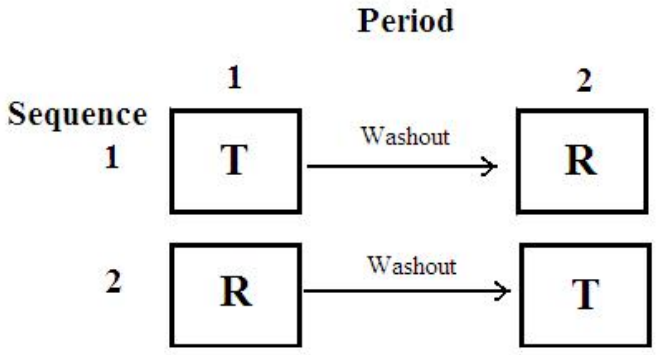

Figure 4.1: Graphical representation of the 2×2 crossover design.

4.2.2 Parameters of Interest

BE assessment depends on the bioavailability of the administered drugs in the participants' systems. Bioavailability is defined as "the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes

available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action," FDA (2003). Hence, when we conduct BE studies, we focus on two main pharmacokinetic parameters: the area under the curve (AUC) and the maximum concentration (drug peak plasma concentration) (C_{max}). Several other parameters are also monitored, such as time to achieve maximum concentration (t_{max}), half-time of maximum concentration (t-half), and elimination rate constant (λ). However, the area under the curve, AUC, of the drug plasma concentration versus time is the most important measure, see Davit et al. (2009).

Another point of interest when discussing the AUC and C_{max} criteria is that both variables have skewed distributions. Therefore, we assume that both have a log-normal distribution. Thus, before statistical analysis, the FDA requires the logarithmic transformation of the data for several reasons that are not only statistical, but also clinical and pharmacokinetic in nature. Rani and Pargal (2004) have reviewed the explanation of this assumption. From a statistical standpoint, because we are analyzing skewed biological data, the log-transformation is a natural transformation to obtain approximately normal data. When the data is skewed, the parameter variances tend to increase with the means. Thus, the log-transformation makes the data nearly symmetric resulting in variances that are more independent of the means. The second reason for the log-transformation is pharmacokinetic. The assumption in the crossover design when we use ANOVA to obtain BE statistics is that the data is a function of additive effects due to subject, period, treatment, and error (Rani and Pargal (2004), Davit et al. (2009)). However, the pharmacokinetic equation for AUC has a multiplicative character because $AUC = (F * D)/(V * Ke)$, where F is the fraction of drug absorbed, D is the given dose, V is the volume of distribution, and Ke is the elimination rate constant. For this reason the FDA has

concluded that if data is analyzed in its original scale, the additive assumption is violated. Thus, because $\ln AUC = \ln F + \ln D - \ln V - \ln Ke$, the logarithmic transformation allows one to use ANOVA to analyze the BE experiment. Notice that a similar argument can be made for the C_{max} parameter. Both AUC and C_{max} must pass the BE criteria for a test drug to be considered bioequivalent to its reference product.

4.2.3 BE Criteria

The AUC and C_{max} criterias are statistically analyzed using two one-sided test procedures to determine whether the average values of these measures, estimated after administration of the test and reference products, are comparable, see Schuirmann (1987). The average values of the data are considered as their geometric mean ratio (GMR) between the test and reference products. However, because both AUC and C_{max} are assumed to be log-normally distributed, the GMR can be computed as the difference of the log-transformed parameter values. According to the FDA definition, bioequivalence is reached if a 90% confidence interval for the differences of log means (or the interval of the GMR of test versus reference) lies within preset BE limits, Davit et al. (2009). The traditional approach to BE testing suggests that those limits are $80\% - 125\%$. These limits are based on medical judgment and FDA experience that a difference of 20% or less in drug exposure is not clinically significant for most drugs, see Haidar et al. (2008a). The 80% limit can be interpreted as occurring when the test product is no less than 80% of the reference, whereas a 125% limit indicates that the reference product is no less than 80% of the test product. This CI method for testing BE was first proposed by Westlake (1972).

4.2.4 Steps in Analyzing Bioequivalence Data

Based on the requirements for these parameters of interest and the study design, we next outline a fundamental approach to perform a BE test. First, we

transform the AUC and C_{max} data using the natural log-transformation. Second, we calculate the differences of the transformed values for each parameter for each subject. Usually, we subtract the reference values from the test values. Third, we calculate the average value for the differences of each metric, i.e., we determine $(\ln AUC_T - \ln AUC_R)/number\ of\ subjects$. Fourth, we calculate the standard deviations of the differences between the transformed data. Fifth, we determine the appropriate value of the test statistic. Sixth, we calculate the upper and lower limits of the mean difference for each AUC and C_{max}. More explicitly, we let $\ln \mu$ and $\ln \sigma$ denote the mean difference and standard deviation of the log-transformed data for either of the parameters. Then, the upper and lower bounds of the CI are $UB \equiv e^{\ln \mu + t_{\alpha,n-1} \ln \sigma/\sqrt{n}}$ and $LB \equiv e^{\ln \mu - t_{\alpha,n-1} \ln \sigma/\sqrt{n}}$, where n is the number of subjects. The interval is appropriate for the AUC or C_{max}. After calculating the CI limits, we take the anti-log in order to return to the original scale and, thus, use the preset bioequivalence limits (BEL) of .80 to 1.25. The last step is to decide whether our test product is bioequivalent to the reference product. As mentioned before, the result of the BE test depends on the interval UB and LB calculated in the previous step. The requirement is that $LB \geq .8$ and $UB \leq 1.25$.

4.2.5 Sources of Variability

The outcome of a BE trial is highly affected by the variables used in calculating the CI of the GMR. Recall that the upper and lower limits of the mean ratio interval must lie within the BE limits. The CI bounds are $UB \equiv e^{\ln \mu + t_{\alpha,n-1} \ln \sigma/\sqrt{n}}$ and $LB \equiv e^{\ln \mu - t_{\alpha,n-1} \ln \sigma/\sqrt{n}}$, where n is the number of subjects, $\ln \mu$ is the mean of the log-transformed reference product, and $\ln \sigma$ is the standard deviation of the reference product. Hence, we identify the quantities of interest in a BE study.

First, we consider the variability in each sample. The two main types of variability are within-subject (WSV) and between-subject variability (BSV). The WSV

is assumed to have the greatest effect on the study outcome. WSV is a measure of each individual's response to the two drug treatments because each person serves as its own control. A person reacts differently even when the same drug is administered under the same conditions (Kytariolos, Karalis, Macheras, and Symillides (2006)). The next source of variability we control for in this study is the BSV. This is a measure of group homogeneity. In this paper we study the crossover design. Its main advantage is that because the treatments are compared on the same subject, the BSV does not contribute to the error variability and the calculation of the appropriate CIs. Nevertheless, because we use a mixed group of participants in the BE trials, the BSV will be affected. The effect of this variability source on a BE crossover design has received little or no attention in the literature.

Another important variable in the BE test outcome is the number of people, N, participating in the trial. BE studies are usually performed on a small group of healthy subjects because pharmaceutical companies try to minimize cost, time, and effort by using the smallest possible sample size. Employing a crossover design helps minimize the required sample size because only about half of the parallel-group study size are needed to achieve the desired power. It is a well-known fact that when N is small, we have wider confidence intervals, which could possibly push the CI bounds beyond the BE limits.

The last variable we investigate is the assumed tolerable geometric mean ratio between the test and reference products. Here we attempt to answer the question "How much of an initial difference between the two drugs should be allowed?" Currently, the FDA allows a 20% difference in the concentration of active ingredients in the blood of one drug to the other, Davit et al. (2009). However, in this study we examine the effect of changing this value on the BE test results.

4.3 Study Design and Simulation Set-Up

The purpose of this study is to investigate the impact of the sample size, BSV, WSV, and the mean ratio difference on the outcome of a Phase III BE study. We performed an extensive Monte Carlo simulation study that compared each of the parameters of interest and their effect on the BE trial outcome. The two criteria of concern are the area under the curve (AUC) and the maximum concentration of the drug (C_{max}). However, because both AUC and C_{max} are assumed to have a log-normal distribution with only slightly different variability, we performed a simulation for AUC using only SAS 9.2.

We simulated two-treatment, two-period, crossover bioequivalence studies assuming sample sizes of $N = 10, 20, ..., 90, 100$. Although, $N = 100$ is not a common sample size for such a study, we examined it to assess the effect of large sample size on the BE test. The underlying distribution was log-normal for the reference drug product. By design of the crossover studies, the between-subject variability was not calculated in the total variability, Midha et al. (1997). However, BE studies were first performed on only male adults, Davit et al. (2009). Now BE studies have expanded to a wider range of participants, and the difference between subjects should no longer be disregarded. Although investigators put a great effort on obtaining a homogenous sample of participants, homogeneity is nearly impossible, and therefore, the BSV should be considered as a source of variability that could affect the BE test outcome. We controlled for the between-subject variability and used it to calculate the log-normal sample standard deviation as $SD = BSV * MEAN$, where the $MEAN$ is the average parameter value for the reference formulation and was set to 100 arbitrary units. The possible BSV values we considered were 10%, 20%, ... , 60%. Then, we assumed that the within-subject variability was known and gave it different values, and we generated a second sample for the test formulation. The possible percentages for WSV were 10%, 20%, ... , 60%. To achieve this result

for the 10% WVS, we set the values of WVS in the simulation to range between 5% and 15%. Each subject was assumed to have its own WSV based on a uniform distribution for the given range. We remark that the FDA defines drugs that have higher than 30% WSV as highly variable drugs. Therefore, in our simulation we addressed the possible problems arising when highly variable drugs are investigated. Thus, the last parameter for which we controlled for was the allowed mean ratio (MR) difference between the reference and test products. By the FDA regulations, a mean ratio difference of 0.2 is acceptable, and the BE criteria is satisfied if the drugs are equivalent, Davit et al. (2009). Thus, the possible MR values we considered were 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, and 20%, where each of the values was assumed to be the same for all subjects in that particular scenario. Refer to Table 4.1 for visual representation of the possible values of each of the parameters of interest. Notice that we investigated each possible combination of values. Hence, we had $8 \times 10 \times 6 \times 6 = 2880$ scenarios of parameter combinations. We generated ten thousand simulated BE trials under each condition.

Last, we evaluated the simulated trials and declared bioequivalence if a 90% confidence interval around the ratio of the estimated geometric means for the two drug products fell between the 80% to 125% BE limits. The percentages of accepted studies were recorded, and power curves were then plotted.

4.4 Results and Discussions

In this section we discuss the results of our Monte Carlo simulation. Here, we divide the results into two groups based on the WSV. As mentioned earlier, pharmaceutical products with higher than 30% WSV are considered to be highly variable. Since different methods of analyzing such drugs exist, we present the results in two groups – normal (WSV less than 30%) and highly variable (WSV greater than 30%).

Table 4.1: Sample sizes used in the simulation for BE trials

	N	BSV %	WSV %	MR % Diff
Sample 1	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50, 60	5−15, 15−25, 25−35 35−45, 45−55, 55−65	2.5
Sample 2	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50, 60	5−15, 15−25, 25−35 35−45, 45−55, 55−65	5.0
Sample 3	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50, 60	5−15, 15−25, 25−35 35−45, 45−55, 55−65	7.5
Sample 4	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50, 60	5−15, 15−25, 25−35 35−45, 45−55, 55−65	10.0
Sample 5	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50, 60	5−15, 15−25, 25−35 35−45, 45−55, 55−65	12.5
Sample 6	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50, 60	5−15, 15−25, 25−35 35−45, 45−55, 55−65	15.0
Sample 7	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50, 60	5−15, 15−25, 25−35 35−45, 45−55, 55−65	17.5
Sample 8	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	10, 20, 30, 40, 50, 60	5−15, 15−25, 25−35 35−45, 45−55, 55−65	20.0

First, we considered the scenario of normal WSV and an assumed MR difference of 0.025. This MR difference indicates that the products do not initially have any known differences (see Figure 4.2). The same colors indicate the same BSV, and the same type of line indicates the same WSV. We had six possible values of BSV and for this case only three possible values of WSV – 0.1, 0.2, 0.3. Notice that under the same BSV values, the BE results were similar for all three values of WSV. For small values of BSV (0.1 and 0.2), the difference in the WSV had a greater effect than for higher values of BSV. We see here that when we had reasonable BSV (less than 0.3), the sample size increase definitely improved the chances of passing at least 80% of the simulated samples under those conditions. Hence, if a pharmaceutical company suggests an increase of the sample size of its trial in order to increase the possibility of meeting the BE criteria, it would have to first ensure that it has a normal-variability drug and the tested subjects are highly homogenous. The inter-

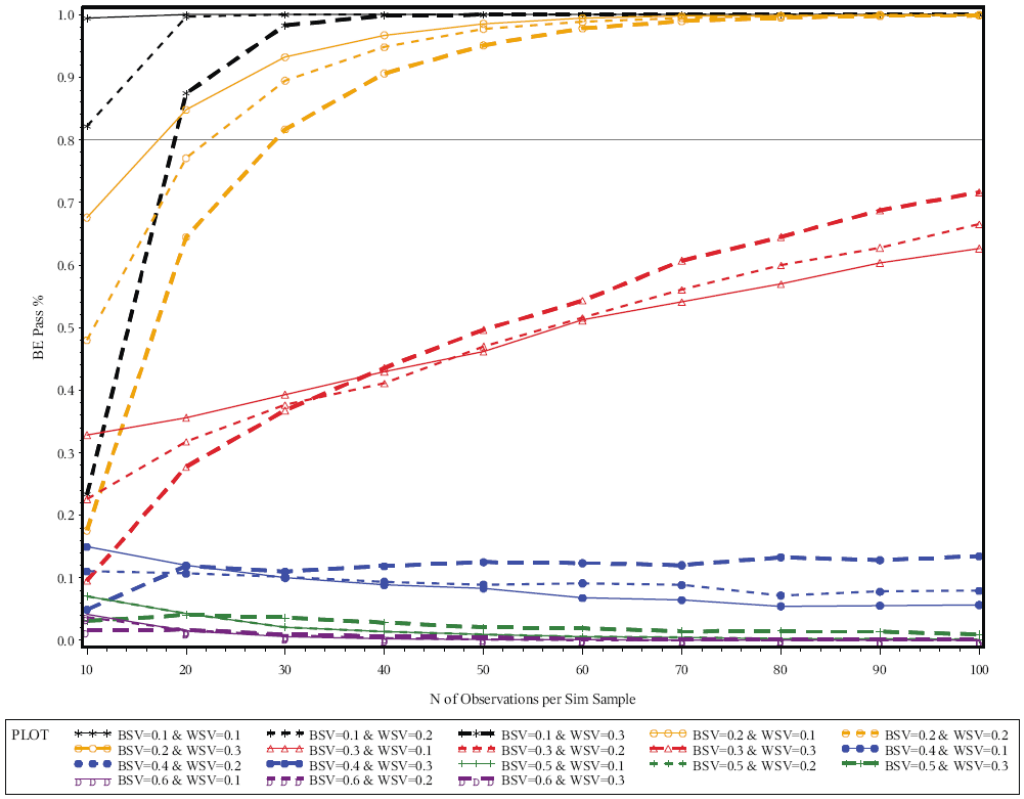

Figure 4.2: Percent studies passing the BE criteria under normal WSV and an assumed MR difference of 0.025

esting discovery here is that when the BSV went to values greater than 30%, the increase in sample size made no difference in meeting the BE criteria. Again, this fact emphasizes the importance of the similarity in the tested groups. In fact, notice that under high BSV with an increase in N, we introduce so much variability that it becomes even harder for the BE criterion to be met. Recall that this discussion is only for the case with MR difference of 0.025.

Figure 4.3 shows the results for the next scenario, where we allowed for a larger difference between the test and reference products. In this case, even when we had a BSV of 0.2, the variability was so large that increasing the sample size made no difference on the chances of meeting the BE requirements. Interestingly, under the right conditions – low WSV, BSV of 0.1 – 0.6 of the simulated samples met the BE

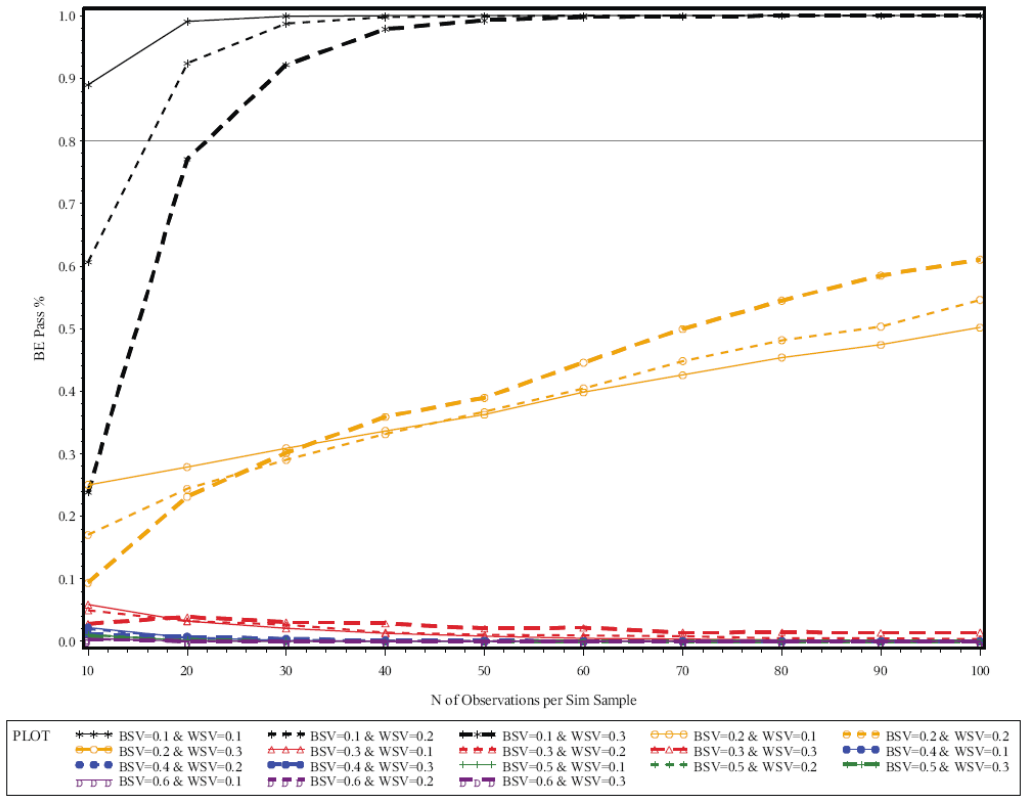

Figure 4.3: Percent studies passing the BE criteria under normal WSV and an assumed MR difference of 0.05

criteria for sample sizes near 20. Hence, the increase in sample size to a larger value of around 30 made no difference. A good drug product performed well even with a small sample size. Nevertheless, the homogeneity of the tested group should be controlled for with caution.

Next, consider the results shown on Figure 4.4. Here we allowed for more assumed variability between the test and reference product – MR difference of 0.075. A greater MR difference made it more difficult to detect equivalence, even with very similar products. Also, the BE criterion was more sensitive to the other types of variability. Notice in Figure 4.4 that the BE is met only under the conditions in the previous example. However, the BE is met at a larger number of subjects. Hence, by allowing for a greater difference between the two products, we must ensure that

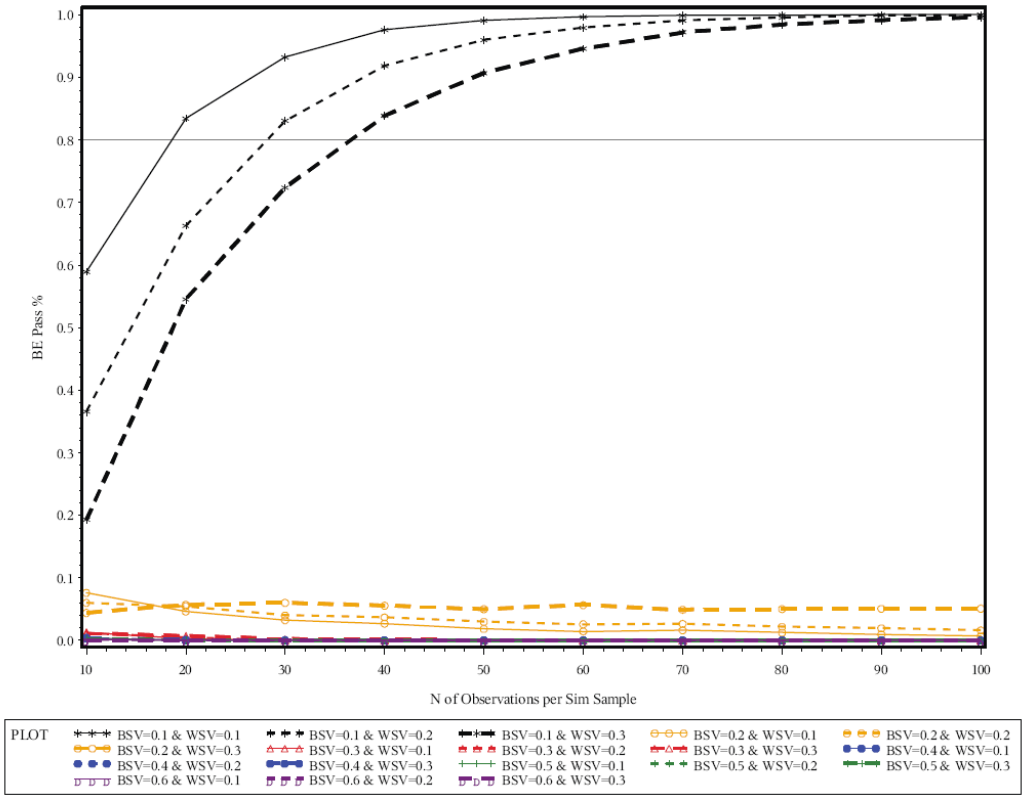

Figure 4.4: Percent studies passing the BE criteria under normal WSV and an assumed MR difference of 0.075

we have a larger sample size to meet the BE criteria. The rest of the figures show the results under the different MR values. However, because they are similar, they are included in Appendix C.1. When we allowed for an MR difference greater than 0.125 (Figure 4.5), a very small and insignificant number of simulated samples met the BE criteria. This result is somewhat consistent with the fact that these products were most likely too different to be found equivalent.

Overall, we conclude that the BSV does impact the outcome of the BE test and should be closely monitored. Also, if two drugs have a reasonable amount of variability, then they would be declared bioequivalent even with the small sample sizes. Hence, the currently used sample size of around 30 patients for these studies is reasonable. Increasing the sample size when the products already have a substantial

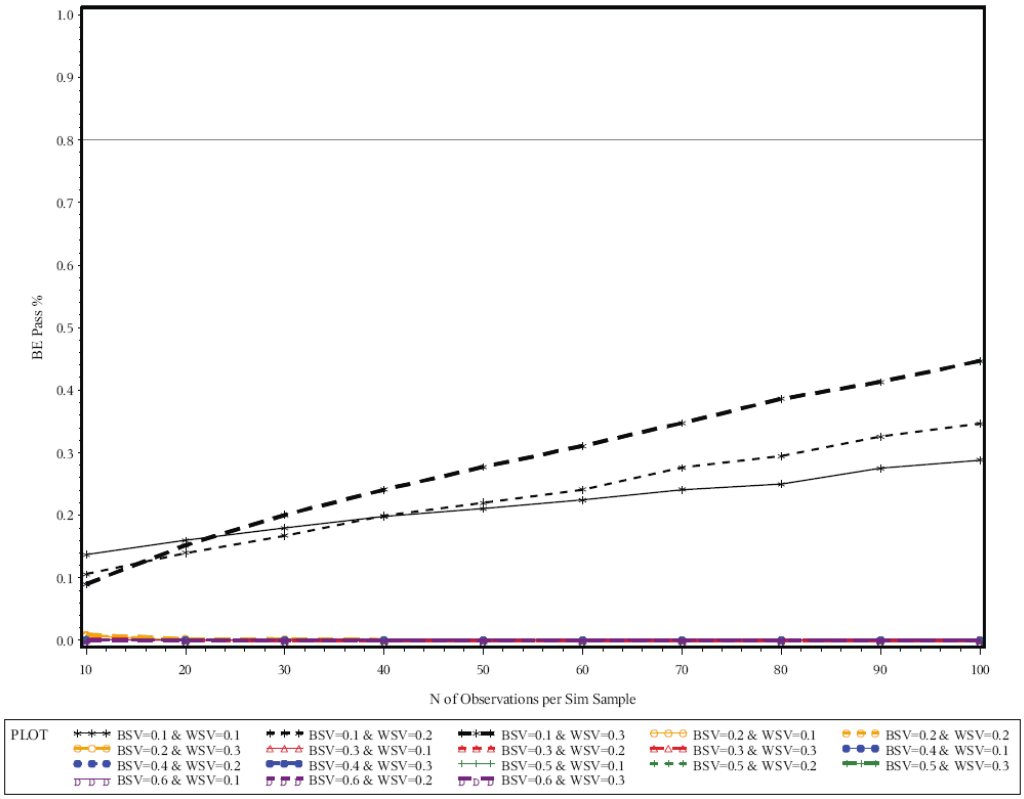

Figure 4.5: Percent studies passing the BE criteria under normal WSV and an assumed MR difference of 0.125

amount of variability between them did not make a significant difference on the BE outcome in the case of normal within-subject variability.

Next, we considered the scenario of highly variable drugs. Recall that these drugs have a WSV greater than 30% (FDA (2003)). First, we considered the case of almost no MR difference assumed, i.e. $MR = 0.025$, (see Figure 4.6). We remark that significant differences occurred in the overall shape of the BE passing curves. The curves with the same BSV and different WSV values had significant differences whereas with a normal WSV, they were very close. Hence, in this scenario even a small change in the WSV decreased the chance of meeting the BE criteria significantly. For example, for the case when BDV was 0.2 and WSV was 0.4, we found that about 30 subjects in the BE trial would be sufficient to meet the BE criteria

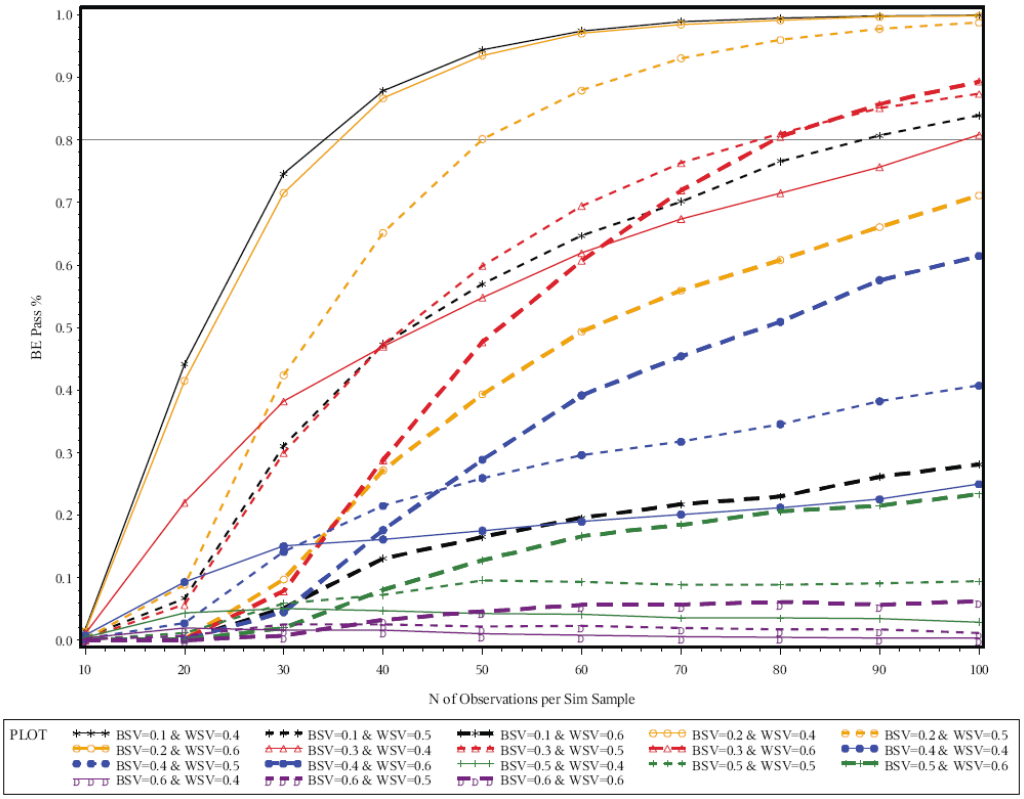

Figure 4.6: Percent studies passing the BE criteria under high WSV and an assumed MR difference of 0.025

at least 80% of the time. However, if we chose a WSV value of 0.5, this number jumped to 50 subjects, and with a WSV value of 0.6, no subjects reached declaration of BE. Also, the fact that under some larger values of BSV, the increase of WSV and sample size actually improved the chances of meeting the BE criteria. For instance, consider the case of BSV of 0.4. When the subjects increased to 40 and above, the studies with larger values of WSV (0.5 and 0.6) had an improved percent passing rate. This is consistent with our expectations that an increased sample size would minimize the effect of the WSV. Nevertheless, not enough improvement existed to meet the BE criteria in at least 80% of the studies.

Next, the conclusions from Figure 4.7 are similar to those above except that the larger assumed MR difference made any case with a BSV value of greater than 0.3

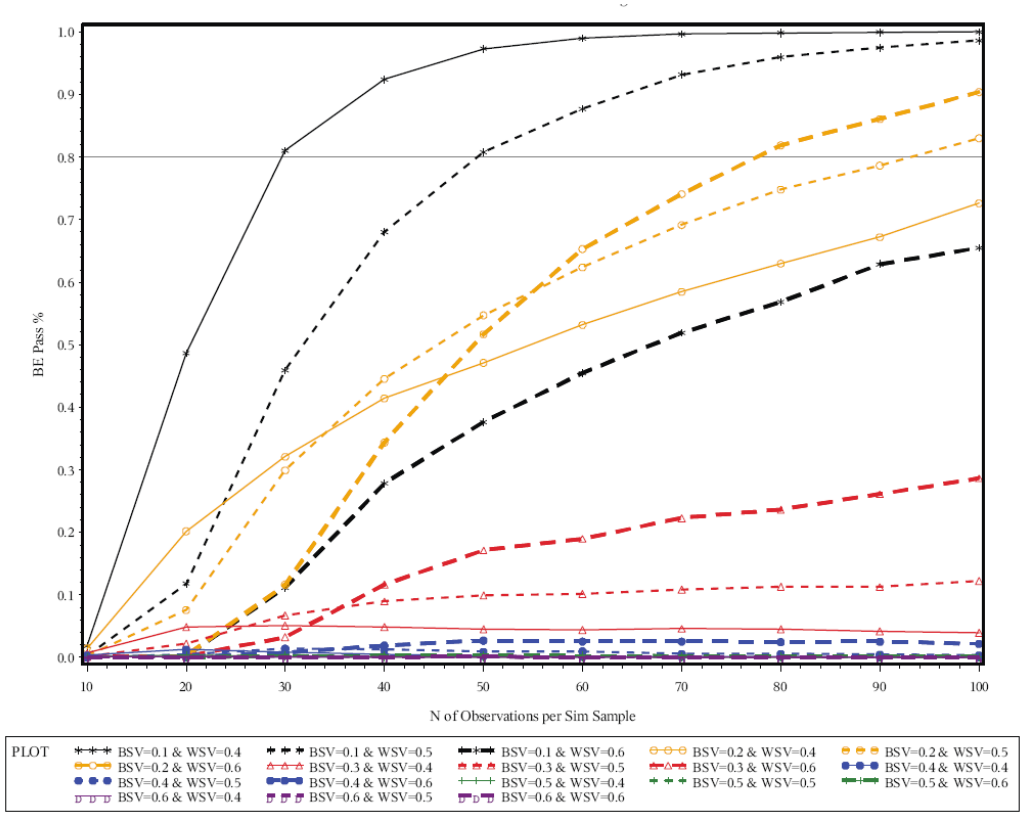

Figure 4.7: Percent studies passing the BE criteria under high WSV and an assumed MR difference of 0.05

almost impossible to pass the BE requirement. However, the results shown in Figure 4.5 are very interesting because the plots showed the significance of the BSV. Also, this figure is a great example of our earlier discussion about the relationship between the increase in WSV and sample size. Clearly, we see that in the case of a BSV of 0.1 and an MR difference of 0.125, and when we have a large WSV of 0.6, the increase in sample size of up to 90 subjects for the BE trial improved the chances of meeting the BE criteria. Again, this conclusion is somewhat consistent with the pharmaceutical companies' suggestion of increasing the sample size to yield BE. However, we remark here that in the case of normal variability drugs, the increased sample size did not make any difference. The plots with larger values of MR difference are included in Appendix C.2 because they showed the same results and indicated only that when

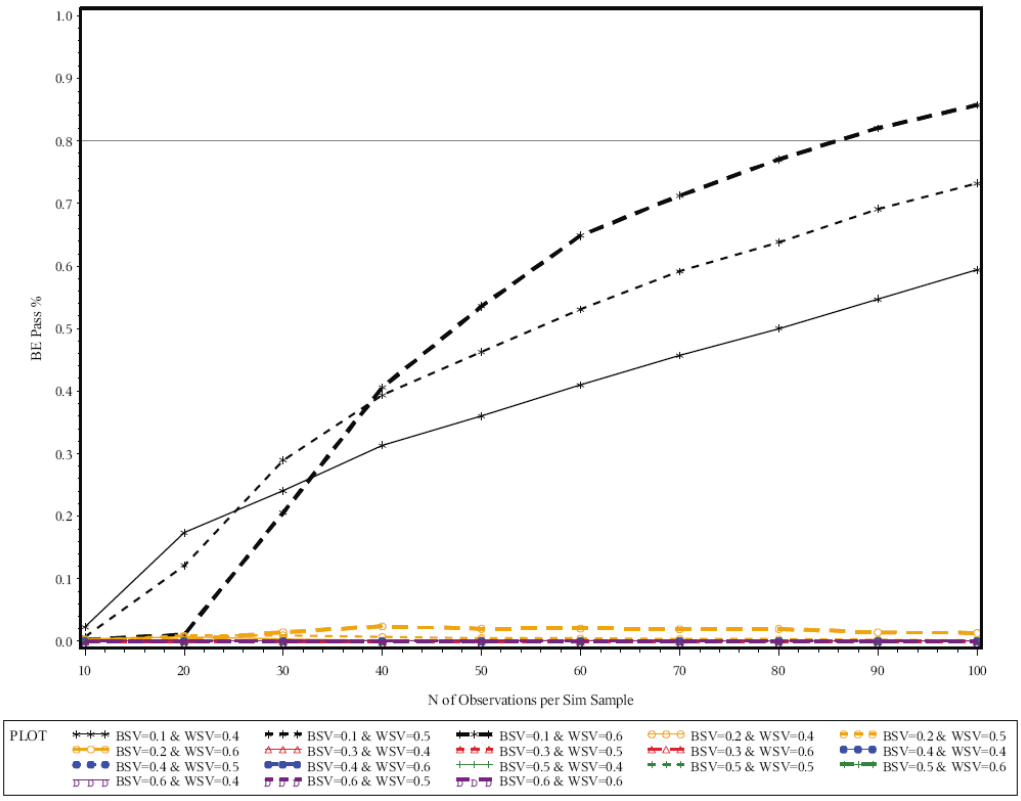

Figure 4.8: Percent studies passing the BE criteria under high WSV and an assumed MR difference of 0.125

the MR difference is large, the two drugs cannot be BE under any scenario of sample size, BSV, or WSV.

Overall, in the case of large variability drugs, we see that in fact an increased sample size decreased the width of the confidence intervals and helped two highly variable drugs meet the BE criteria. However, a large sample size is not always attractive to pharmaceutical companies because of time and cost issues. Also, we see that the BSV should be closely monitored and should not be ignored as a source of variability that affects the outcome of any BE trial. Fortunately, this defect is corrected when one uses the crossover design. Nevertheless, if not accounted for, it could cause a trial to fail the BE test.

4.5 Comments

In the previous section we discussed the results of our extensive simulation trial designed to assess the sources of variabilities affecting the outcome of a BE trial. The overall conclusions indicated that a significant difference existed between the normal and highly variable drugs. In both scenarios we noticed that the BSV is a parameter that pharmaceutical companies should monitor. Also, in the case of normal WSV drugs, the increase of sample size did not make any difference on the outcomes. Hence, when the products were "similar," BE was detected even with the small sample sizes. Again, this fact confirms the current regulatory requirements and designs for these particular studies and the use of a small sample size of subjects. However, we saw that in the case of highly variable drugs, the increase of sample size improved the chances of the two drugs tested being declared BE.

This fact brings us to a different topic that is of future interest. Because the FDA has considered the highly variable products as a separate group, many methods have been suggested for analyzing such trials using different BE criteria. The most common suggestion is using scaling of the regular interval endpoints to accommodate for the extra variability and the increased interval width of the BE test. Haidar, Makhlouf, Schuirmann, Hyslop, Davit, Conner, and Yu (2008c) have evaluated this approach. Also, Kytariolos et al. (2006) have presented novel scaled BE limits for the highly variable drugs. Haidar, Davit, Chen, Conner, Lee, Li, Lionberger, Makhlouf, Patel, Schuirmann, and Yu (2008a) as well as Tothfalusi, Endrenyi, Midha, Rawson, and Hubbard (2001) have given a good overall picture of the problems associated with such products. Last but not least, Tothfalusi and Endrenyi (2003) have well summarized several scaled approaches and how the CI limits should be adjusted for the BE comparison of highly variable drugs. When dealing with a highly variable drug, a company should very carefully consider the design and BE limits calculation for its trial. We will investigate this topic in our future work.

APPENDICES

APPENDIX A

Derivations for Chapter Two

A.1 Maximum Likelihood Estimators for ψ and $\boldsymbol{\eta}$

1.1.1 Tenenbein (1970)'s Re-parameterizations

To derive the MLEs, we re-parameterize so that

$$\begin{aligned}
\alpha_k &= \pi_k S_k + (1-\pi_k)(1-C_k), \\
\beta_k &= \frac{\pi_k S_k}{\alpha_k}, \\
\gamma_k &= \frac{\pi_k (1-S_k)}{1-\alpha_k}, \\
1-\beta_k &= \frac{(1-\pi_k)(1-C_k)}{\alpha_k},
\end{aligned}$$

and

$$1-\gamma_k = \frac{(1-\pi_k)C_k}{1-\alpha_k},$$

where $k = 0, 1$ indicates the disease group of the participant, $D = 0$ or 1 for diseased and non-diseased, respectively.

1.1.2 Likelihood and Log-Likelihood Functions in Terms of α, β, and γ

The likelihood function is

$$\begin{aligned}
L \propto\ & [\pi_1 S_1 + (1-\pi_1)(1-C_1)]^{W_{11}} [1-\pi_1 S_1 - (1-\pi_1)(1-C_1)]^{N_1 - W_{11}} \\
& \times [\pi_0 S_0 + (1-\pi_0)(1-C_0)]^{W_{10}} [1-\pi_0 S_0 - (1-\pi_0)(1-C_0)]^{N_0 - W_{10}} \\
& \times (\pi_1)^{T_1} (1-\pi_1)^{M_1 - T_1} (S_1)^{V_{111}} (1-S_1)^{T_1 - V_{111}} \\
& \times (\pi_0)^{T_0} (1-\pi_0)^{M_0 - T_0} (S_0)^{V_{110}} (1-S_0)^{T_0 - V_{110}} \\
& \times (C_1)^{V_{001}} (1-C_1)^{(M_1 - T_1) - V_{001}} \\
& \times (C_0)^{V_{000}} (1-C_0)^{(M_0 - T_0) - V_{000}}.
\end{aligned} \tag{A.1}$$

For ease of derivation, we consider each part of the likelihood that comes from the cases ($D = 1$) or controls ($D = 0$) separately. With substitution of the parameters defined in Appendix A.1, Section (1.1.1), and some algebraic manipulation, we see that the full likelihood is a product of binomials so that

$$\begin{aligned} L \propto\ & \alpha_1^{W_{11}+V_{111}+V_{101}}(1-\alpha_1)^{W_{01}+V_{011}+V_{001}}\beta_1^{V_{111}}(1-\beta_1)^{V_{101}}\gamma_1^{V_{011}}(1-\gamma_1)^{V_{001}} \\ & \times \alpha_0^{W_{10}+V_{110}+V_{100}}(1-\alpha_0)^{W_{00}+V_{010}+V_{000}}\beta_0^{V_{110}}(1-\beta_0)^{V_{100}}\gamma_0^{V_{010}}(1-\gamma_0)^{V_{000}}. \end{aligned}$$

Then, the log-likelihood is

$$\begin{aligned} \ell \propto\ & (W_{11}+V_{111}+V_{101})\ln(\alpha_1) + (W_{01}+V_{011}+V_{001})\ln(1-\alpha_1) \\ & + V_{111}\ln(\beta_1) + V_{101}\ln(1-\beta_1) + V_{011}\ln(\gamma_1) + V_{001}\ln(1-\gamma_1) \\ & + (W_{10}+V_{110}+V_{100})\ln(\alpha_0) + (W_{00}+V_{010}+V_{000})\ln(1-\alpha_0) \\ & + V_{110}\ln(\beta_0) + V_{100}\ln(1-\beta_0) + V_{010}\ln(\gamma_0) + V_{000}\ln(1-\gamma_0). \end{aligned} \tag{A.2}$$

1.1.3 Partial First Derivatives and MLEs for α, β, and γ

To determine the MLEs for α_k, β_k, and γ_k $(k = 0, 1)$, we first find the partial first derivatives of the log-likelihood function defined in (A.2) with respect to each of the parameters so that

$$\begin{aligned} \frac{\partial \ln L}{\partial \alpha_k} &= \frac{W_{1k}+V_{11k}+V_{10k}}{\alpha_k} - \frac{W_{0k}+V_{01k}+V_{00k}}{1-\alpha_k}, \\ \frac{\partial \ln L}{\partial \beta_k} &= \frac{V_{11k}}{\beta_k} - \frac{V_{10k}}{1-\beta_k}, \\ \frac{\partial \ln L}{\partial \gamma_k} &= \frac{V_{01k}}{\gamma_k} - \frac{V_{00k}}{1-\gamma_k}, \end{aligned}$$

where $k = 0$ indicates non-diseased status and $k = 1$ indicates a diseased status. Therefore, the MLEs are

$$\hat{\alpha}_k = \frac{V_{10k}+V_{11k}+W_{1k}}{M_k+N_k}$$

$$\hat{\beta}_k = \frac{V_{11k}}{V_{10k} + V_{11k}}$$
$$\hat{\gamma}_k = \frac{V_{01k}}{V_{00k} + V_{01k}}.$$

Now, going back to our original notation and using the invariance property of the MLEs to derive $\hat{\psi}$, we get the results given in Section 2.3.

A.2 Observed Information Matrix

To calculate the confidence intervals of interest, we obtain the observed information matrix. Recall from Chapter 2 that the matrix has the form

$$\boldsymbol{J}(\psi, \boldsymbol{\eta}) = -\begin{bmatrix} \frac{\partial^2 \ell}{\partial \psi^2} & \frac{\partial^2 \ell}{\partial \psi \partial \pi_1} & \frac{\partial^2 \ell}{\partial \psi \partial S_0} & \frac{\partial^2 \ell}{\partial \psi \partial S_1} & \frac{\partial^2 \ell}{\partial \psi \partial C_0} & \frac{\partial^2 \ell}{\partial \psi \partial C_1} \\ \cdot & \frac{\partial^2 \ell}{\partial \pi_1^2} & \frac{\partial^2 \ell}{\partial \pi_1 \partial S_0} & \frac{\partial^2 \ell}{\partial \pi_1 \partial S_1} & \frac{\partial^2 \ell}{\partial \pi_1 \partial C_0} & \frac{\partial^2 \ell}{\partial \pi_1 \partial C_1} \\ \cdot & \cdot & \frac{\partial^2 \ell}{\partial S_0^2} & \frac{\partial^2 \ell}{\partial S_0 \partial S_1} & \frac{\partial^2 \ell}{\partial S_0 \partial C_0} & \frac{\partial^2 \ell}{\partial S_0 \partial C_1} \\ \cdot & \cdot & \cdot & \frac{\partial^2 \ell}{\partial S_1^2} & \frac{\partial^2 \ell}{\partial S_1 \partial C_0} & \frac{\partial^2 \ell}{\partial S_1 \partial C_1} \\ \cdot & \cdot & \cdot & \cdot & \frac{\partial^2 \ell}{\partial C_0^2} & \frac{\partial^2 \ell}{\partial C_0 \partial C_1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \frac{\partial^2 \ell}{\partial C_1^2} \end{bmatrix}.$$

Now, we derive each term on the upper diagonal for the symmetric matrix

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \psi^2} = \Bigg[& \frac{\pi_1(\pi_1 - 1)^2(S_0 + C_0 - 1)W_{10}\left[2(C_0 - 1)\psi(\pi_1 - 1) + \pi_1(S_0 - C_0 + 1)\right]}{\left[(C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0\right]^2} \\ & + \frac{\pi_1(\pi_1 - 1)^2(S_0 + C_0 - 1)W_{00}\left[2C_0\psi(\pi_1 - 1) + \pi_1(S_0 - C_0 - 1)\right]}{\left[C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1)\right]^2} \\ & + \frac{\pi_1(M_0 - T_0)\left(2\psi(\pi_1 - 1) - \pi_1\right)}{\psi^2} + (\pi_1 - 1)^2 T_0 \Bigg] \Big/ \left(\psi + \pi_1 - \psi\pi_1\right)^2, \end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \psi \partial \pi_1} &= \frac{\pi_1(T_0 - M_0)}{\psi(\pi_1 - 1)(-\psi - \pi_1 + \psi\pi_1)} + \frac{\pi_1(T_0 - M_0)}{\psi(\pi_1 - 1)(\psi + \pi_1 - \psi\pi_1)^2} \\
&+ \frac{T_0(\pi_1 - 1)(\psi - 1)}{(\psi + \pi_1 - \psi\pi_1)^2} + \frac{T_0}{\psi + \pi_1 - \psi\pi_1} \\
&+ \frac{2(M_0 - T_0)\pi_1(\psi - 1)}{\psi(\psi + \pi_1 - \psi\pi_1)^2} + \frac{M_0 - T_0}{\psi(\psi + \pi_1 - \psi\pi_1)} \\
&+ \frac{(-\psi + \pi_1 + \psi\pi_1)(S_0 + C_0 - 1)(N_0 - W_{10})}{(\psi + \pi_1 - \psi\pi_1)^2(C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1))} \\
&- \frac{\psi\pi_1(\pi_1 - 1)(S_0 + C_0 - 1)^2(N_0 - W_{10})}{(\psi + \pi_1 - \psi\pi_1)^2(C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1))^2} \\
&- \frac{\psi\pi_1(\pi_1 - 1)(S_0 + C_0 - 1)^2 W_{10}}{(\psi + \pi_1 - \psi\pi_1)^2((C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0)^2} \\
&+ \frac{(\psi(\pi_1 - 1) + \pi_1)(S_0 + C_0 - 1)W_{10}}{(\psi + \pi_1 - \psi\pi_1)^2((C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0)},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \psi \partial S_0} = \Bigg[& \; C_0^2\psi(1 - \pi_1)\left[2N_0(\psi(\pi_1 - 1) - \pi_1 S_0) + W_{10}(\psi + 2\pi_1 - \psi\pi_1)\right] \\
& + C_0^3 N_0 \psi^2(\pi_1 - 1)^2 + C_0 N_0(\psi - \psi\pi_1 + \pi_1 S_0)^2 - \pi_1^2(S_0 - 1)^2 W_{10} \\
& - C_0 W_{10}\left[\psi^2(\pi_1 - 1)^2 - 2\psi\pi_1(\pi_1 - 1) + \pi_1^2(2S_0 - 1)\right] \Bigg] \\
& \times \frac{\pi_1(1 - \pi_1)}{[C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1)]^2[(C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0]^2},
\end{aligned}
$$

$$
\frac{\partial^2 \ell}{\partial \psi \partial S_1} = 0,
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \psi \partial C_0} = & \left[\frac{\psi(\pi_1 - 1)(S_0 + C_0 - 1)(N_0 - W_{10})}{(C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1))^2} + \frac{\psi(\pi_1 - 1)(S_0 + C_0 - 1)W_{10}}{((C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0)^2}\right. \\
& \left. - \frac{W_{10}}{(C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0} + \frac{-N_0 + W_{10}}{C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1)}\right] \\
& \times \frac{(\pi_1 - 1)\pi_1}{\psi(\pi_1 - 1) - \pi_1},
\end{aligned}
$$

$$
\frac{\partial^2 \ell}{\partial \psi \partial C_1} = 0,
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \pi_1^2} = & \frac{2(M_0 - T_0)(\psi - 1)}{(\pi_1 - 1)(\psi + \pi_1 - \psi\pi_1)^2} + \frac{\psi T_0}{(-\psi - \pi_1 + \psi\pi_1)\pi_1^2} + \frac{\psi(\psi - 1)T}{\pi_1(\psi + \pi_1 - \psi\pi_1)^2} \\
& - \frac{T_1}{\pi_1^2} + \frac{-M_1 + T_1}{(\pi_1 - 1)^2} + \frac{2\psi(\psi - 1)(S_0 + C_0 - 1)(N_0 - W_{10})}{(\psi + \pi_1 - \psi\pi_1)^2(C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1))} \\
& + \frac{T_0 - M_0}{(\pi_1 - 1)^2(\psi + \pi_1 - \psi\pi_1)^2} - \frac{(S_1 + C_1 - 1)^2 W_{11}}{(1 + C_1(\pi_1 - 1) + \pi_1(S_1 - 1))^2} \\
& - \frac{\psi^2(S_0 + C_0 - 1)^2(N_0 - W_{10})}{(\psi + \pi_1 - \psi\pi_1)^2(C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1))^2} \\
& - \frac{\psi^2(S_0 + C_0 - 1)^2 W_{10}}{(\psi + \pi_1 - \psi\pi_1)^2((C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0)^2} \\
& + \frac{2\psi(\psi - 1)(S_0 + C_0 - 1)W_{10}}{(\psi + \pi_1 - \psi\pi_1)^2((C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0)} \\
& - \frac{(S_1 + C_1 - 1)^2(N_1 - W_{11})}{(C_1(\pi_1 - 1) + \pi_1(S_1 - 1))^2},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \pi_1 \partial S_0} = & \Big[\; C_0^2\psi(1 - \pi_1)\left[2N_0(\psi(\pi_1 - 1) - \pi_1 S_0) + W_{10}(\psi + 2\pi_1 - \psi\pi_1)\right] \\
& + C_0^3 N_0 \psi^2(\pi_1 - 1)^2 - \pi_1^2(S_0 - 1)^2 W_{10} + C_0 N_0(\psi - \psi\pi_1 + \pi_1 S_0)^2 \\
& - C_0 W_{10}\left[\psi^2(\pi_1 - 1)^2 - 2\psi\pi_1(\pi_1 - 1) + \pi_1^2(2S_0 - 1)\right] \; \Big] \\
& \times \frac{-\psi}{[C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1)]^2[(C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0]^2},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \pi_1 \partial S_1} = & \Big[\; C_1^2(\pi_1 - 1)[2N_1(1 + \pi_1(S_1 - 1)) - W_{11}(\pi_1 + 1)] \\
& + C_1[N_1(1 + \pi_1(S_1 - 1))^2 + W_{11}(-1 - 2\pi_1^2(S_1 - 1))] \\
& + C_1^3 N_1(\pi_1 - 1)^2 - \pi_1^2(S_1 - 1)^2 W_{11} \; \Big] \\
& \times \frac{-1}{[C_1(\pi_1 - 1) + \pi_1(S_1 - 1)]^2[1 + C_1(\pi_1 - 1) + \pi_1(S_1 - 1)]^2},
\end{aligned}
$$

$$\frac{\partial^2 \ell}{\partial \pi_1 \partial C_0} = \Big[\; 2W_{10}\psi\pi_1(\pi_1-1)(S_0-1)S_0 - W_{10}\pi_1^2(S_0-1)S_0$$
$$+ W_{10}\psi^2(\pi_1-1)^2\left[1+C_0^2+2C_0(S_0-1)-S_0\right]$$
$$+ N_0(S_0-1)[(C_0-1)\psi(\pi_1-1)+\pi_1 S_0]^2 \quad \Big]$$
$$\times \frac{\psi}{[C_0\psi(\pi_1-1)+\pi_1(S_0-1)]^2[(C_0-1)\psi(\pi_1-1)+\pi_1 S_0]^2},$$

$$\frac{\partial^2 \ell}{\partial \pi_1 \partial C_1} = \frac{W_{11}}{1+C_1(\pi_1-1)+\pi_1(S_1-1)} - \frac{(\pi_1-1)(S_1+C_1-1)(N_1-W_{11})}{(C_1(\pi_1-1)+\pi_1(S_1-1))^2}$$
$$+ \frac{N_1-W_{11}}{C_1(\pi_1-1)+\pi_1(S_1-1)} - \frac{(\pi_1-1)(S_1+C_1-1)W_{11}}{(1+C_1(\pi_1-1)+\pi_1(S_1-1))^2},$$

$$\frac{\partial^2 \ell}{\partial S_0^2} = \frac{V_{110}-T_0}{(S_0-1)^2} - \frac{V_{110}}{S_0^2} - \frac{\pi_1^2(N_0-W_{10})}{(C_0\psi(\pi_1-1)+\pi_1(S_0-1))^2}$$
$$- \frac{\pi_1^2 W_{10}}{((C_0-1)\psi(\pi_1-1)+\pi_1 S_0)^2},$$

$$\frac{\partial^2 \ell}{\partial S_0 \partial S_1} = 0,$$

$$\frac{\partial^2 \ell}{\partial S_0 \partial C_0} = \left[\frac{N_0-W_{10}}{(C_0\psi(\pi_1-1)+\pi_1(S_0-1))^2} + \frac{W_{10}}{((C_0-1)\psi(\pi_1-1)+\pi_1 S_0)^2}\right]$$
$$\times \frac{\psi\pi_1(\pi_1-1)(\psi(\pi_1-1)-\pi_1)}{\psi+\pi_1-\psi\pi_1},$$

$$\frac{\partial^2 \ell}{\partial S_0 \partial C_1} = 0,$$

$$\frac{\partial^2 \ell}{\partial S_1^2} = -\frac{V_{111}}{S_1^2} + \frac{V_{111}-T_1}{(S_1-1)^2} - \frac{\pi_1^2(N_1-W_{11})}{(C_1(\pi_1-1)+\pi_1(S_1-1))^2}$$
$$- \frac{\pi_1^2 W_{11}}{(1+C_1(\pi_1-1)+\pi_1(S_1-1))^2},$$

$$\frac{\partial^2 \ell}{\partial S_1 \partial C_0} = 0,$$

$$\frac{\partial^2 \ell}{\partial S_1 \partial C_1} = -\frac{W_{11}\pi_1(\pi_1 - 1)}{(1 + C_1(\pi_1 - 1) + \pi_1(S_1 - 1))^2} + \frac{(-N_1 + W_{11})\pi_1(\pi_1 - 1)}{(C_1(\pi_1 - 1) + \pi_1(S_1 - 1))^2},$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial C_0^2} = &-\frac{V_{000}}{C_0^2} + \frac{V_{000} - M_0 + T_0}{(C_0 - 1)^2} - \frac{\psi^2(\pi_1 - 1)^2(N_0 - W_{10})}{(C_0\psi(\pi_1 - 1) + \pi_1(S_0 - 1))^2} \\ &- \frac{\psi^2(\pi_1 - 1)^2 W_{10}}{((C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0)^2},\end{aligned}$$

$$\frac{\partial^2 \ell}{\partial C_0 \partial C_1} = 0,$$

and

$$\begin{aligned}\frac{\partial^2 \ell}{\partial C_1^2} = &-\frac{V_{001}}{C_1^2} + \frac{V_{001} - M_1 + T_1}{(C_1 - 1)^2} - \frac{(\pi_1 - 1)^2(N_1 - W_{11})}{(C_1(\pi_1 - 1) + \pi_1(S_1 - 1))^2} \\ &- \frac{(\pi_1 - 1)^2 W_{11}}{(1 + C_1(\pi_1 - 1) + \pi_1(S_1 - 1))^2},\end{aligned}$$

where the reminder of the terms under the diagonal are found by the symmetry property of the information matrix.

APPENDIX B

Derivations for Chapter Three

B.1 Restricted Maximum Likelihood Estimation

First, let L_U be the full likelihood function. Then,

$$
\begin{aligned}
L_U = &\begin{pmatrix} U_{111}+U_{011} \\ U_{111} \end{pmatrix} S^{U_{111}}(1-S)^{U_{011}} \begin{pmatrix} U_{110}+U_{010} \\ U_{110} \end{pmatrix} S^{U_{110}}(1-S)^{U_{010}} \\
&\times \begin{pmatrix} N_1 \\ U_{111}+U_{011} \end{pmatrix} \pi_1^{U_{111}+U_{011}}(1-\pi_1)^{N_1-(U_{111}+U_{011})} \\
&\times \begin{pmatrix} N_0 \\ U_{110}+U_{010} \end{pmatrix} \pi_0^{U_{110}+U_{010}}(1-\pi_0)^{N_0-(U_{110}+U_{010})} \\
&\times \begin{pmatrix} N_1-(U_{111}+U_{011}) \\ W_{01}-U_{011} \end{pmatrix} C^{W_{01}-U_{011}}(1-C)^{W_{11}-U_{111}} \\
&\times \begin{pmatrix} N_0-(U_{110}+U_{010}) \\ W_{00}-U_{010} \end{pmatrix} C^{W_{00}-U_{010}}(1-C)^{W_{10}-U_{110}} \\
&\times \begin{pmatrix} T_1 \\ V_{111} \end{pmatrix} S^{V_{111}}(1-S)^{T_1-V_{111}} \begin{pmatrix} T_0 \\ V_{110} \end{pmatrix} S^{V_{110}}(1-S)^{T_0-V_{110}} \\
&\times \begin{pmatrix} M_1 \\ T_1 \end{pmatrix} \pi_1^{T_1}(1-\pi_1)^{M_1-T_1} \begin{pmatrix} M_0 \\ T_0 \end{pmatrix} \pi_0^{T_0}(1-\pi_0)^{M_0-T_0} \\
&\times \begin{pmatrix} M_1-T_1 \\ V_{001} \end{pmatrix} C^{V_{001}}(1-C)^{(M_1-T_1)-V_{001}} \\
&\times \begin{pmatrix} M_0-T_0 \\ V_{000} \end{pmatrix} C^{V_{000}}(1-C)^{(M_0-T_0)-V_{000}},
\end{aligned}
$$

where the function is in terms of the complete observed and unobserved data, and the last four lines come from the complete data. Then, we rewrite this equation with the intention of grouping for each of the latent variables, U_{ijk}, so that

$$
\begin{aligned}
L_U = & \begin{pmatrix} N_1 \\ U_{111}+U_{011} \end{pmatrix} \begin{pmatrix} U_{111}+U_{011} \\ U_{111} \end{pmatrix} \begin{pmatrix} N_1-(U_{111}+U_{011}) \\ W_{01}-U_{011} \end{pmatrix} \\
& \times \begin{pmatrix} N_0 \\ U_{110}+U_{010} \end{pmatrix} \begin{pmatrix} U_{110}+U_{010} \\ U_{110} \end{pmatrix} \begin{pmatrix} N_0-(U_{110}+U_{010}) \\ W_{00}-U_{010} \end{pmatrix} \\
& \times (\pi_1 S)^{U_{111}}[\pi_1(1-S)]^{U_{011}}[(1-\pi_1)(1-C)]^{W_{11}-U_{111}}[(1-\pi_1)C]^{W_{01}-U_{011}} \\
& \times (\pi_0 S)^{U_{110}}[\pi_0(1-S)]^{U_{010}}[(1-\pi_0)(1-C)]^{W_{10}-U_{110}}[(1-\pi_0)C]^{W_{00}-U_{010}} \\
& \times \begin{pmatrix} T_1 \\ V_{111} \end{pmatrix} S^{V_{111}}(1-S)^{T_1-V_{111}} \begin{pmatrix} T_0 \\ V_{110} \end{pmatrix} S^{V_{110}}(1-S)^{T_0-V_{110}} \\
& \times \begin{pmatrix} M_1 \\ T_1 \end{pmatrix} \pi_1^{T_1}(1-\pi_1)^{M_1-T_1} \begin{pmatrix} M_0 \\ T_0 \end{pmatrix} \pi_0^{T_0}(1-\pi_0)^{M_0-T_0} \\
& \times \begin{pmatrix} M_1-T_1 \\ V_{001} \end{pmatrix} C^{V_{001}}(1-C)^{(M_1-T_1)-V_{001}} \\
& \times \begin{pmatrix} M_0-T_0 \\ V_{000} \end{pmatrix} C^{V_{000}}(1-C)^{(M_0-T_0)-V_{000}},
\end{aligned}
$$

where $\pi_0 = \frac{\pi_1}{\pi_1-\pi_1\psi+\psi}$. Define the mixture of binomial distributions for the complete data as $f_1(T_0)$, $f_2(T_1)$, $f_3(V_{111}|T_1)$, $f_4(V_{110}|T_0)$, $f_5(V_{001}|T_1)$, and $f_6(V_{000}|T_0)$. Then, after several adjustments and regrouping as described by Joseph et al. (1995), we get

$$
\begin{aligned}
L_U \propto & \frac{W_{11}!}{U_{111}!(W_{11}-U_{111})!}\frac{W_{10}!}{U_{110}!(W_{10}-U_{110})!}\frac{W_{01}!}{U_{011}!(W_{01}-U_{011})!}\frac{W_{00}!}{U_{010}!(W_{00}-U_{010})!} \\
& \times \left[\frac{\pi_1 S}{\pi_1 S + (1-\pi_1)(1-C)}\right]^{U_{111}} \left[\frac{(1-\pi_1)(1-C)}{\pi_1 S + (1-\pi_1)(1-C)}\right]^{W_{11}-U_{111}} \\
& \times \left[\frac{\pi_0 S}{\pi_0 S + (1-\pi_0)(1-C)}\right]^{U_{110}} \left[\frac{(1-\pi_0)(1-C)}{\pi_0 S + (1-\pi_0)(1-C)}\right]^{W_{10}-U_{110}} \\
& \times \left[\frac{\pi_1 (1-S)}{\pi_1 (1-S) + (1-\pi_1)C}\right]^{U_{011}} \left[\frac{(1-\pi_1)C}{\pi_1 (1-S) + (1-\pi_1)C}\right]^{W_{01}-U_{011}} \\
& \times \left[\frac{\pi_0 (1-S)}{\pi_0 (1-S) + (1-\pi_0)C}\right]^{U_{010}} \left[\frac{(1-\pi_0)C}{\pi_0 (1-S) + (1-\pi_0)C}\right]^{W_{00}-U_{010}} \\
& \times f_1(T_0) f_2(T_1) f_3(V_{111}|T_1) f_4(V_{110}|T_0) f_5(V_{001}|T_1) f_6(V_{000}|T_0).
\end{aligned}
$$

Substituting for $\pi_0 = \frac{\pi_1}{\pi_1 - \pi_1\psi + \psi}$ gives us

$$
\begin{aligned}
L_U \propto & \frac{W_{11}!}{U_{111}!(W_{11}-U_{111})!}\frac{W_{10}!}{U_{110}!(W_{10}-U_{110})!}\frac{W_{01}!}{U_{011}!(W_{01}-U_{011})!}\frac{W_{00}!}{U_{010}!(W_{00}-U_{010})!} \\
& \times \left[\frac{\pi_1 S}{\pi_1 S + (1-\pi_1)(1-C)}\right]^{U_{111}} \left[\frac{(1-\pi_1)(1-C)}{\pi_1 S + (1-\pi_1)(1-C)}\right]^{W_{11}-U_{111}} \\
& \times \left[\frac{\pi_1 S}{\pi_1 S + \psi(1-\pi_1)(1-C)}\right]^{U_{110}} \left[\frac{\psi(1-\pi_1)(1-C)}{\pi_1 S + \psi(1-\pi_1)(1-C)}\right]^{W_{10}-U_{110}} \\
& \times \left[\frac{\pi_1 (1-S)}{\pi_1 (1-S) + (1-\pi_1)C}\right]^{U_{011}} \left[\frac{(1-\pi_1)C}{\pi_1 (1-S) + (1-\pi_1)C}\right]^{W_{01}-U_{011}} \\
& \times \left[\frac{\pi_1 (1-S)}{\pi_1 (1-S) + \psi(1-\pi_1)C}\right]^{U_{010}} \left[\frac{\psi(1-\pi_1)C}{\pi_1 (1-S) + \psi(1-\pi_1)C}\right]^{W_{00}-U_{010}} \\
& \times f_1(T_0) f_2(T_1) f_3(V_{111}|T_1) f_4(V_{110}|T_0) f_5(V_{001}|T_1) f_6(V_{000}|T_0).
\end{aligned}
$$

Clearly, the conditional distributions of the latent variables are binomial with parameters as described in Chapter 3, Section 3.4.

B.2 Observed Information Matrix

To calculate the confidence intervals of interest, we obtain the observed information matrix. Recall from Chapter 3 that the matrix has the form

$$\boldsymbol{J}(\psi,\boldsymbol{\eta}) = -\begin{bmatrix} \frac{\partial^2\ell}{\partial\psi^2} & \frac{\partial^2\ell}{\partial\psi\partial\pi_1} & \frac{\partial^2\ell}{\partial\psi\partial S} & \frac{\partial^2\ell}{\partial\psi\partial C} \\ \cdot & \frac{\partial^2\ell}{\partial\pi_1^2} & \frac{\partial^2\ell}{\partial\pi_1\partial S} & \frac{\partial^2\ell}{\partial\pi_1\partial C} \\ \cdot & \cdot & \frac{\partial^2\ell}{\partial S^2} & \frac{\partial^2\ell}{\partial S\partial C} \\ \cdot & \cdot & \cdot & \frac{\partial^2\ell}{\partial C^2} \end{bmatrix}.$$

Now, we derive each term on the upper diagonal for the symmetric $\boldsymbol{J}(\psi,\boldsymbol{\eta})$ matrix. Hence,

$$\begin{aligned}\frac{\partial^2\ell}{\partial\psi^2} &= \frac{(-1+\pi_1)\pi_1(V_{000}+V_{100})}{\psi(\pi_1+\psi-\pi_1\psi)^2} - \frac{\pi_1(V_{000}+V_{100})}{\psi^2(\pi_1+\psi-\pi_1\psi)} + \frac{(-1+\pi_1)^2(V_{010}+V_{110})}{(\pi_1+\psi-\pi_1\psi)^2} \\ &+ \frac{C(-1+\pi_1)^2\pi_1(-1+C+S)W_{00}}{(\pi_1(-1+\psi)-\psi)(C\psi-\pi_1(-1+C\psi+S))^2} \\ &+ \frac{(-1+\pi_1)^2\pi_1(-1+C+S)W_{00}}{(\pi_1+\psi-\pi_1\psi)^2(-C\psi+\pi_1(-1+C\psi+S))} \\ &+ \frac{(-1+C)(-1+\pi_1)^2\pi_1(-1+C+S)W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1S)^2} \\ &+ \frac{(-1+\pi_1)^2\pi_1(-1+C+S)W_{10}}{(\pi_1+\psi-\pi_1\psi)^2((-1+C)(-1+\pi_1)\psi+\pi_1S)},\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial\psi\partial\pi_1} &= \frac{\pi_1(-1+\psi)(V_{000}+V_{100})}{\psi(\pi_1+\psi-\pi_1\psi)^2} + \frac{V_{000}+V_{100}}{\psi(\pi_1+\psi-\pi_1\psi)} \\
&+ \frac{(-1+\pi_1)(-1+\psi)(V_{010}+V_{110})}{(\pi_1+\psi-\pi_1\psi)^2} + \frac{V_{010}+V_{110}}{\pi_1+\psi-\pi_1\psi} \\
&+ \frac{(-1+\pi_1)\pi_1(-1+C+S)(-1+C\psi+S)W_{00}}{(\pi_1(-1+\psi)-\psi)(C\psi-\pi_1(-1+C\psi+S))^2} \\
&- \frac{(-1+\pi_1)(-1+C+S)W_{00}}{(\pi_1(-1+\psi)-\psi)(-C\psi+\pi_1(-1+C\psi+S))} \\
&- \frac{\pi_1(-1+C+S)W_{00}}{(\pi_1(-1+\psi)-\psi)(-C\psi+\pi_1(-1+C\psi+S))} \\
&+ \frac{(-1+\pi_1)\pi_1(-1+\psi)(-1+C+S)W_{00}}{(\pi_1+\psi-\pi_1\psi)^2(-C\psi+\pi_1(-1+C\psi+S))} \\
&+ \frac{(-1+\pi_1)\pi_1(-1+C+S)((-1+C)\psi+S)W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1 S)^2} \\
&- \frac{(-1+\pi_1)(-1+C+S)W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1 S)} \\
&- \frac{\pi_1(-1+C+S)W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1 S)} \\
&+ \frac{(-1+\pi_1)\pi_1(-1+\psi)(-1+C+S)W_{10}}{(\pi_1+\psi-\pi_1\psi)^2((-1+C)(-1+\pi_1)\psi+\pi_1 S)},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial\psi\partial S} &= \frac{(-1+\pi_1)\pi_1}{\pi_1(-1+\psi)\psi}\left[\frac{W_{00}}{C\psi-\pi_1(-1+C\psi+S)} - \frac{W_{10}}{(-1+C)(-1+\pi_1)\psi+\pi_1 S}\right. \\
&+ \pi_1(-1+C+S)\left(\frac{W_{00}}{(C\psi-\pi_1(-1+C\psi+S))^2}\right. \\
&\left.\left. + \frac{W_{10}}{((-1+C)(-1+\pi_1)\psi+\pi_1 S)^2}\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial\psi\partial C} &= \frac{1}{\pi_1(-1+\psi)-\psi}(-1+\pi_1)\pi_1 \\
&\times \left[\frac{(-1+\pi_1)\psi(-1+C+S)W_{00}}{(C\psi-\pi_1(-1+C\psi+S))^2} + \frac{W_{00}}{C\psi-\pi_1(-1+C\psi+S)}\right. \\
&+ \left.\frac{(-1+\pi_1)\psi(-1+C+S)W_{10}}{((-1+C)(-1+\pi_1)\psi+\pi_1 S)^2} - \frac{W_{10}}{(-1+C)(-1+\pi_1)\psi+\pi_1 S}\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \pi_1^2} &= \frac{V_{000}+V_{100}}{(-1+\pi_1)^2(\pi_1(-1+\psi)-\psi)} + \frac{(-1+\psi)(V_{000}+V_{100})}{(-1+\pi_1)(\pi_1+\psi-\pi_1\psi)^2} \\
&- \frac{V_{001}+V_{101}}{(-1+\pi_1)^2} + \frac{(-1+\psi)\psi(V_{010}+V_{110})}{\pi_1(\pi_1+\psi-\pi_1\psi)^2} - \frac{\psi(V_{010}+V_{110})}{\pi_1^2(\pi_1+\psi-\pi_1\psi)} \\
&- \frac{V_{011}+V_{111}}{\pi_1^2} + \frac{\psi(-1+C+S)(-1+C\psi+S)W_{00}}{(\pi_1(-1+\psi)-\psi)(C\psi-\pi_1(-1+C\psi+S))^2} \\
&+ \frac{(-1+\psi)\psi(-1+C+S)W_{00}}{(\pi_1+\psi-\pi_1\psi)^2(-C\psi+\pi_1(-1+C\psi+S))} - \frac{(-1+C+S)^2W_{01}}{(C(-1+\pi_1)+\pi_1(-1+S))^2} \\
&+ \frac{\psi(-1+C+S)((-1+C)\psi+S)W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1S)^2} \\
&+ \frac{(-1+\psi)\psi(-1+C+S)W_{10}}{(\pi_1+\psi-\pi_1\psi)^2((-1+C)(-1+\pi_1)\psi+\pi_1S)} \\
&- \frac{(-1+C+S)^2W_{11}}{(1+C(-1+\pi_1)+\pi_1(-1+S))^2},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \pi_1 \partial S} &= \frac{\pi_1\psi(-1+C+S)W_{00}}{(\pi_1(-1+\psi)-\psi)(C\psi-\pi_1(-1+C\psi+S))^2} \\
&- \frac{\psi W_{00}}{(\pi_1(-1+\psi)-\psi)(-C\psi+\pi_1(-1+C\psi+S))} \\
&+ \frac{W_{01}}{C(-1+\pi_1)+\pi_1(-1+S)} - \frac{\pi_1(-1+C+S)W_{01}}{(C(-1+\pi_1)+\pi_1(-1+S))^2} \\
&+ \frac{\pi_1\psi(-1+C+S)W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1S)^2} \\
&- \frac{\psi W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1S)} \\
&+ \frac{W_{11}}{1+C(-1+\pi_1)+\pi_1(-1+S)} - \frac{\pi_1(-1+C+S)W_{11}}{(1+C(-1+\pi_1)+\pi_1(-1+S))^2},
\end{aligned}
$$

$$\begin{aligned}\frac{\partial^2\ell}{\partial\pi_1\partial C} &= \frac{(-1+\pi_1)\psi^2(-1+C+S)W_{00}}{(\pi_1(-1+\psi)-\psi)(C\psi-\pi_1(-1+C\psi+S))^2}\\ &- \frac{\psi W_{00}}{(\pi_1(-1+\psi)-\psi)(-C\psi+\pi_1(-1+C\psi+S))}\\ &+ \frac{W_{01}}{C(-1+\pi_1)+\pi_1(-1+S)} - \frac{(-1+\pi_1)(-1+C+S)W_{01}}{(C(-1+\pi_1)+\pi_1(-1+S))^2}\\ &+ \frac{(-1+\pi_1)\psi^2(-1+C+S)W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1S)^2}\\ &- \frac{\psi W_{10}}{(\pi_1(-1+\psi)-\psi)((-1+C)(-1+\pi_1)\psi+\pi_1S)}\\ &+ \frac{W_{11}}{1+C(-1+\pi_1)+\pi_1(-1+S)} - \frac{(-1+\pi_1)(-1+C+S)W_{11}}{(1+C(-1+\pi_1)+\pi_1(-1+S))^2},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\ell}{\partial S^2} &= -\frac{V_{010}+V_{011}}{(-1+S)^2} - \frac{V_{110}+V_{111}}{S^2}\\ &- \frac{\pi_1^2W_{00}}{(C\psi-\pi_1(-1+C\psi+S))^2} - \frac{\pi_1^2W_{01}}{(C(-1+\pi_1)+\pi_1(-1+S))^2}\\ &- \frac{\pi_1^2W_{10}}{((-1+C)(-1+\pi_1)\psi+\pi_1S)^2} - \frac{\pi_1^2W_{11}}{(1+C(-1+\pi_1)+\pi_1(-1+S))^2},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\ell}{\partial S\partial C} &= (-1+\pi_1)\pi_1\left(-\frac{\psi W_{00}}{(C\psi-\pi_1(-1+C\psi+S))^2} - \frac{W_{01}}{(C(-1+\pi_1)+\pi_1(-1+S))^2}\right.\\ &\left.- \frac{\psi W_{10}}{((-1+C)(-1+\pi_1)\psi+\pi_1S)^2} - \frac{W_{11}}{(1+C(-1+\pi_1)+\pi_1(-1+S))^2}\right),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2\ell}{\partial C^2} &= -\frac{V_{000}+V_{001}}{C^2} - \frac{V_{100}+V_{101}}{(-1+C)^2} - \frac{(-1+\pi_1)^2\psi^2W_{00}}{(C\psi-\pi_1(-1+C\psi+S))^2}\\ &- \frac{(-1+\pi_1)^2W_{01}}{(C(-1+\pi_1)+\pi_1(-1+S))^2} - \frac{(-1+\pi_1)^2\psi^2W_{10}}{((-1+C)(-1+\pi_1)\psi+\pi_1S)^2}\\ &- \frac{(-1+\pi_1)^2W_{11}}{(1+C(-1+\pi_1)+\pi_1(-1+S))^2}.\end{aligned}$$

APPENDIX C

Additional Plots For Chapter Four

C.1 Additional Plots with Assumed Normal Within-subject Variability

In this section we include the plots that were not discussed in the results section of Chapter Four in the case of normal variability case.

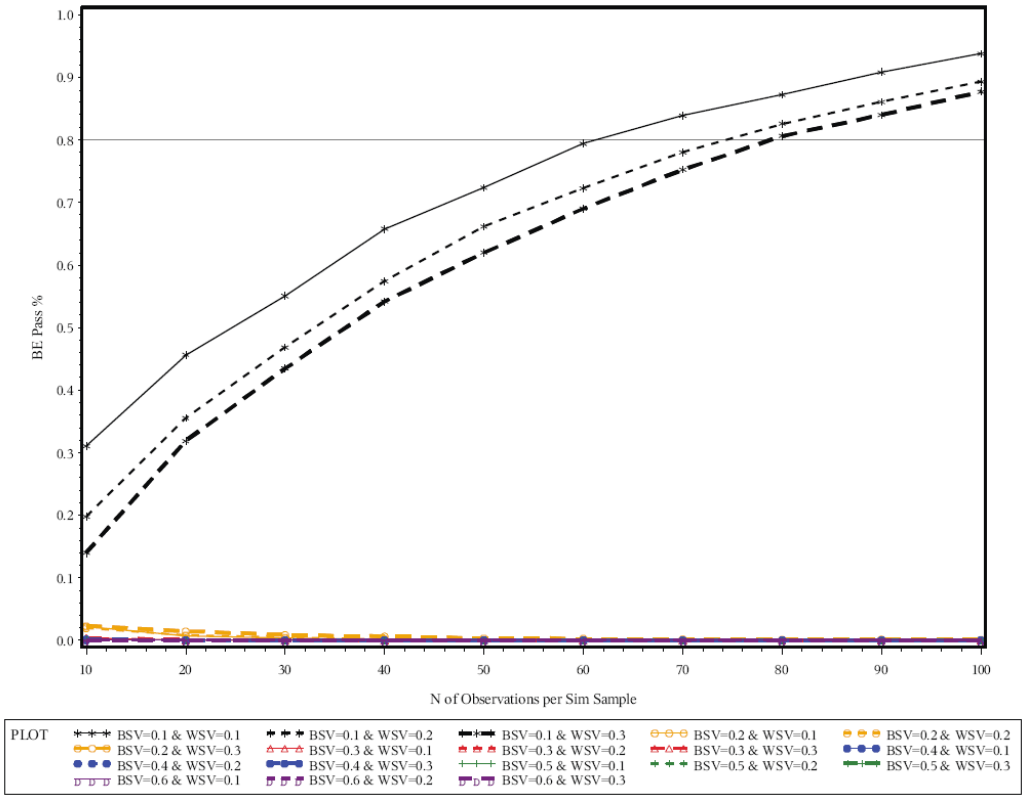

Figure C.1: Percent studies passing the BE criteria under normal WSV and an assumed MR difference of 0.10

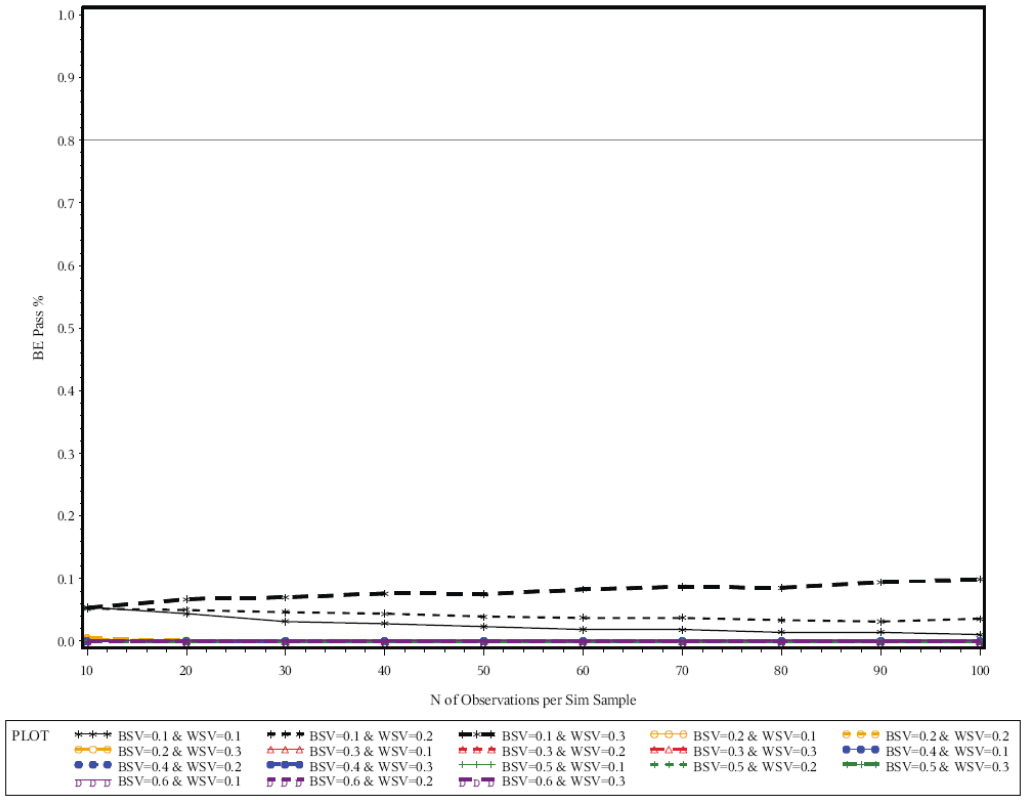

Figure C.2: Percent studies passing the BE criteria under normal WSV and an assumed MR difference of 0.15

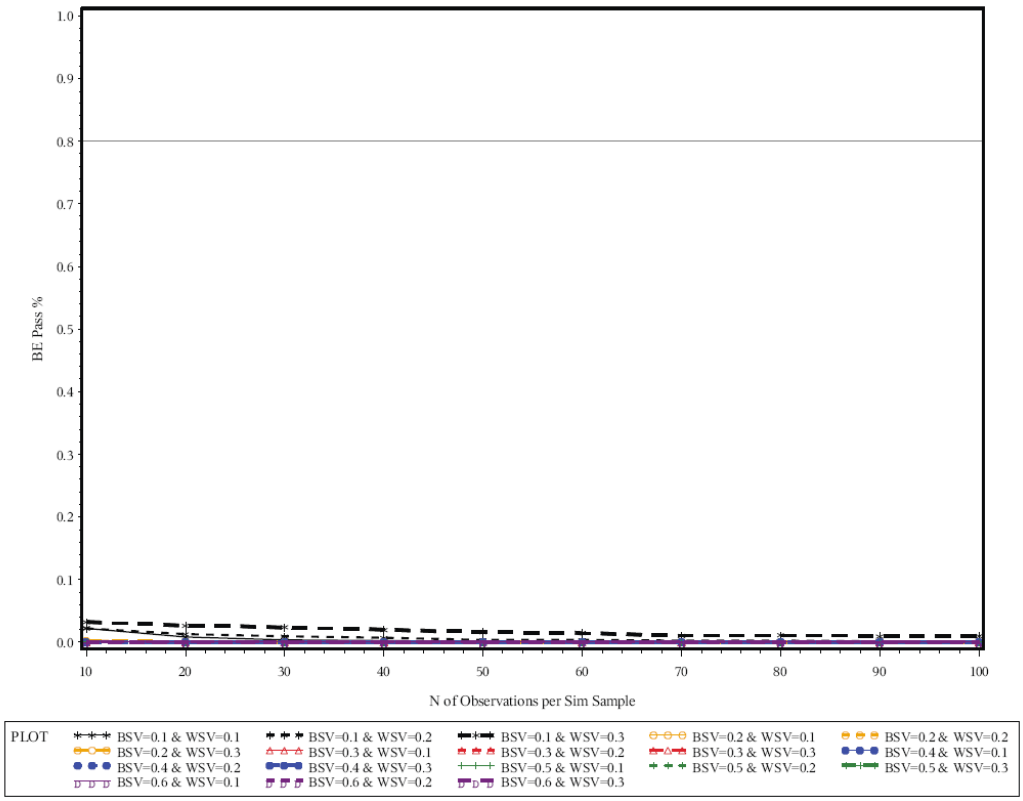

Figure C.3: Percent studies passing the BE criteria under normal WSV and an assumed MR difference of 0.175

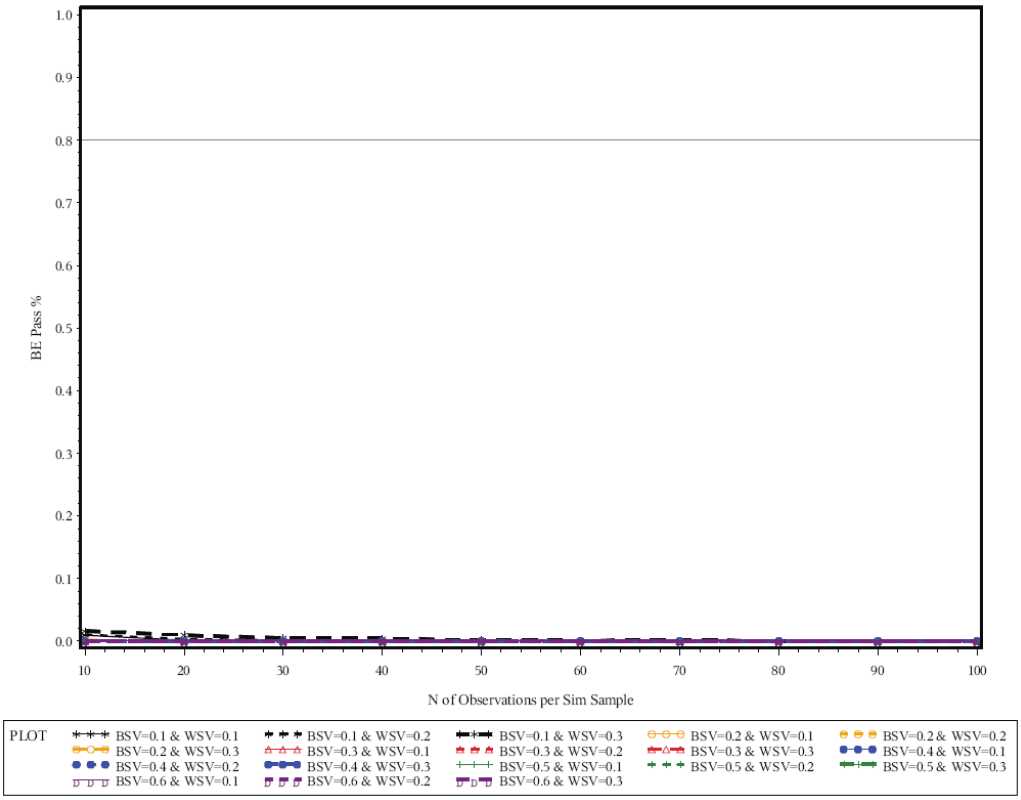

Figure C.4: Percent studies passing the BE criteria under normal WSV and an assumed MR difference of 0.20

C.2 Additional Plots with Assumed High Within-subject Variability

In this section we include the plots that were not discussed in the results section of Chapter Four in the case of highly variable drugs.

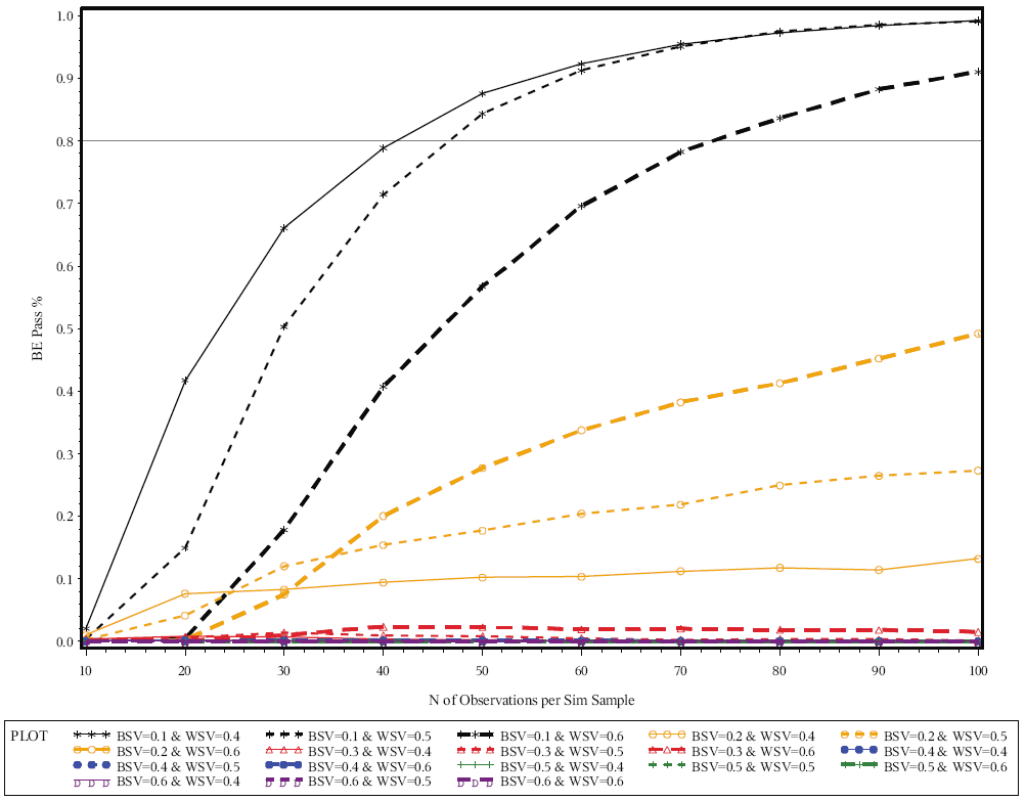

Figure C.5: Percent studies passing the BE criteria under high WSV and an assumed MR difference of 0.075

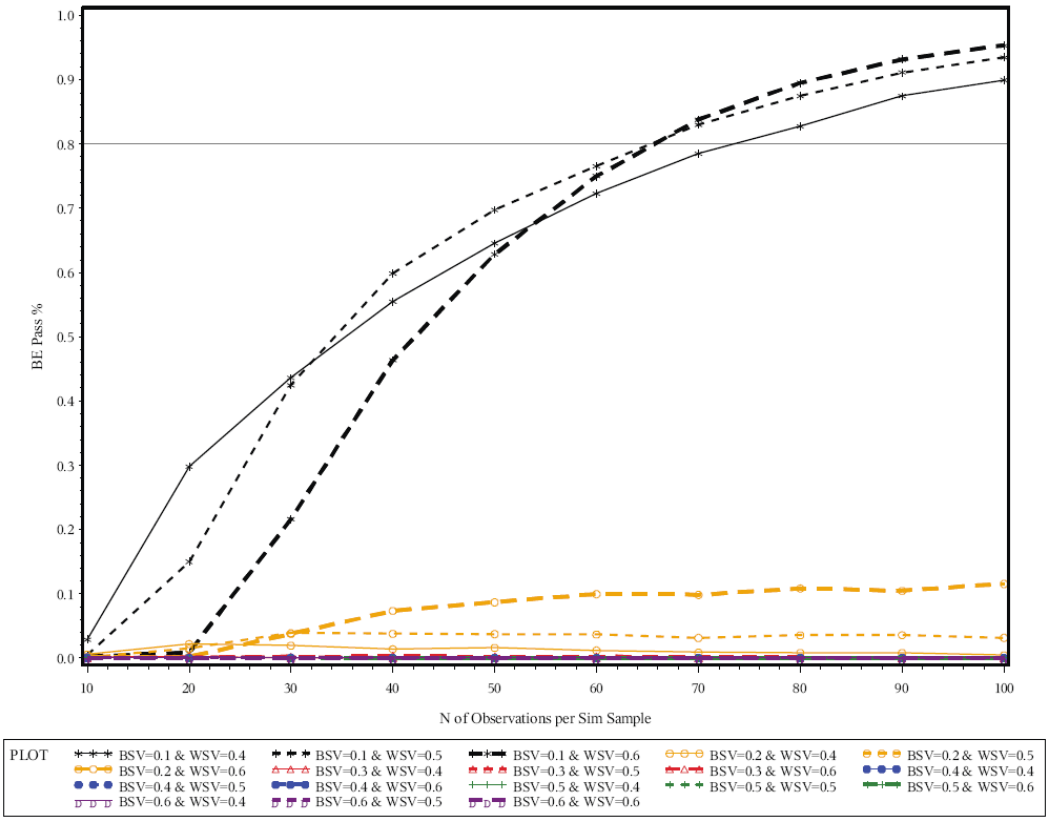

Figure C.6: Percent studies passing the BE criteria under high WSV and an assumed MR difference of 0.10

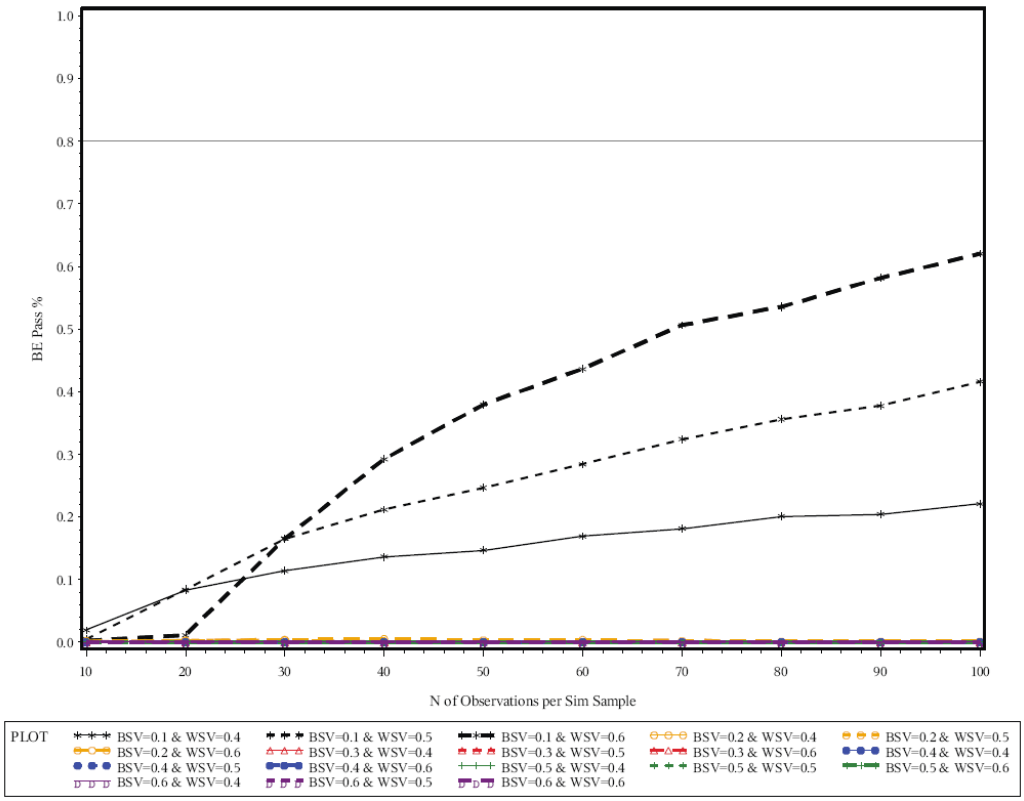

Figure C.7: Percent studies passing the BE criteria under high WSV and an assumed MR difference of 0.15

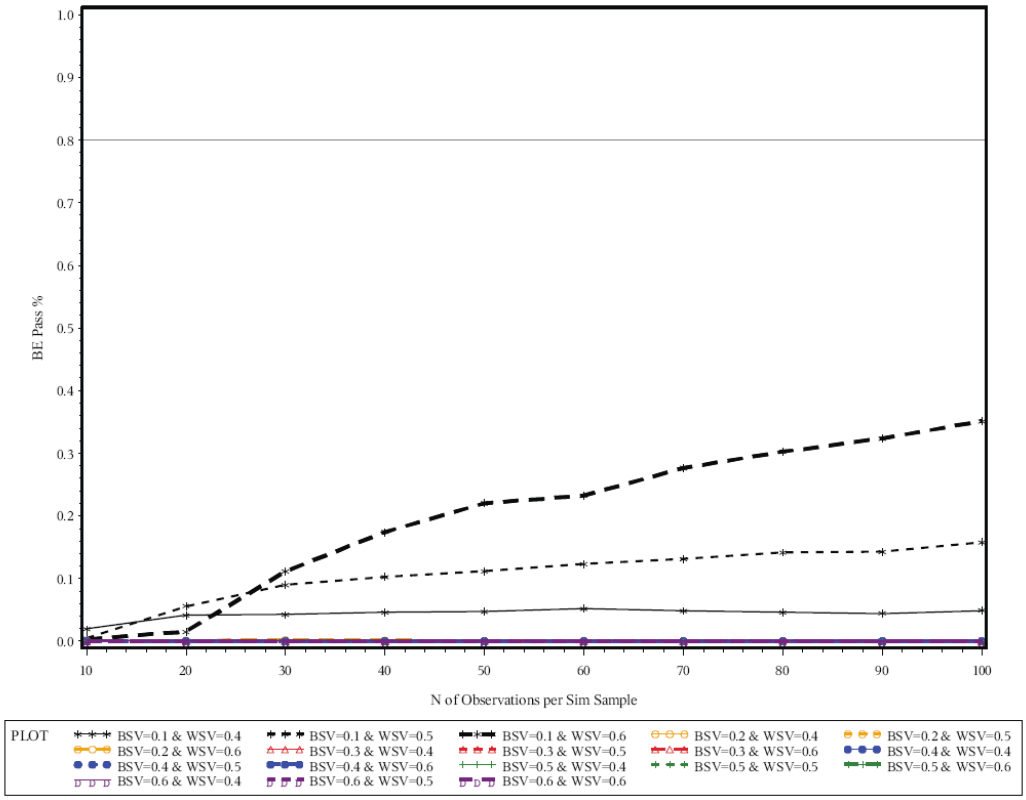

Figure C.8: Percent studies passing the BE criteria under high WSV and an assumed MR difference of 1.075

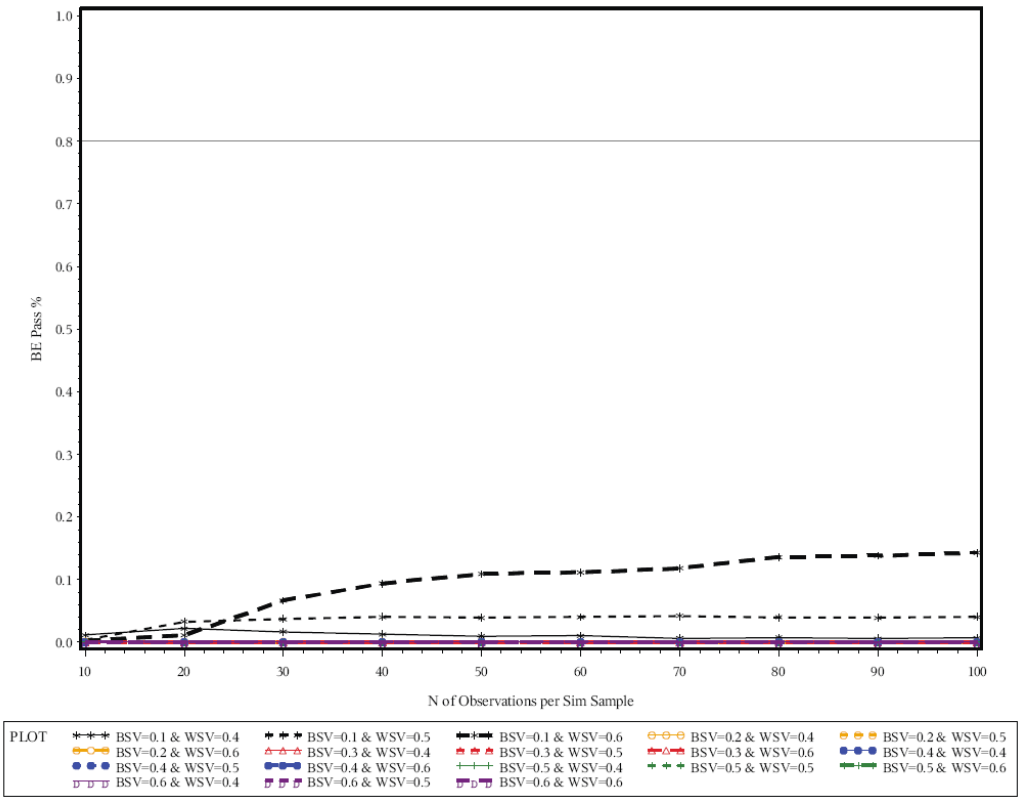

Figure C.9: Percent studies passing the BE criteria under high WSV and an assumed MR difference of 0.20

BIBLIOGRAPHY

Agresti, A. and Coull, B. A. (1998), "Approximate Is Better than 'Exact' for Interval Estimation of Binomial Proportions," *The American Statistician*, 52, 119-126.

Asagiri, M., Hirai, T., Kunigami, T., Kamano, S., Gober, H.-J., Okamoto, K., Nishikawa, K., Latz, E., Golenbock, D. T., Aoki, K., Ohya, K., Imai, Y., Morishita, Y., Miyazono, K., Kato, S., Saftig, P., and Takayanagi, H. (2008), "Cathepsin K-Dependent Toll-Like Receptor 9 Signaling Revealed in Experimental Arthritis," *Science*, 319, 624-627.

Balthasar, J. P. (1999), "Bioequivalence and Bioequivalency Testing," *American Journal of Pharmaceutical Education*, 63, 194-198.

Bapuji, A. T., Ravikiran, H. L. V., Nagesh, M., Syedba, S., Ramaraju, D., Reddy, C. R. J., Ravinder, S., Yeruva, R. R., and Roy, S. (2010), "Bioequivalence Testing – Industry Perspective," *Journal of Bioequivelence and Bioavailability*, 2, 98-101.

Basu, D. (1977), "On the Elimination of Nuisance Parameters," *Journal of the American Statistical Association*, 72, 355-366.

Berger, J. O., Liseo, B., and Wolpert, R. L. (1999), "Integrated Likelihood Methods for Eliminating Nuisance Parameters," *Statistical Science*, 14, 1-22.

Birkett, D. J. (2003), "Generics – Equal or Not?" *Australian Prescriber*, 26, 85-87.

Boese, D. (2005), "Likelihood-based Confidence Intervals for Proportion Parameters with Binary Data Subject to Misclassification," Ph.D. thesis, Baylor University.

Boese, D. H., Young, D. M., and Stamey, J. D. (2006), "Confidence Intervals for a Binomial Parameter Based on Binary Data Subject to False-positive Misclassification," *Computational Statistics & Data Analysis*, 50, 3369-3385.

Brenner, H. (1992), "Use and Limitations of Dual Measurements in Correcting for Nondifferential Exposure Misclassification," *Epidemiology*, 3, 216-222.

— (1996), "Correcting for Exposure Misclassification Using an Alloyed Gold Standard," *Epidemiology*, 7, 406-410.

Breslow, N. E. (1996), "Statistics in Epidemiology: The Case-Control Study," *Journal of the American Statistical Association*, 91, 14-28.

Bross, I. (1954), "Misclassification in 2×2 Tables," *Biometrics*, 10, 478-486.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001), "Interval Estimation for a Binomial Proportion," *Statistical Science*, 16, 101-133.

Chen, M.-L., Shah, V., Patnaik, R., Adams, W., Hussain, A., Conner, D., Mehta, M., Malinowski, H., Lazor, J., Huang, S.-M., Hare, D., Lesko, L., Sporn, D., and Williams, R. (2001), "Bioavailability and Bioequivalence: An FDA Regulatory Overview," *Pharmaceutical Research*, 18, 1645-1650.

Chow, S.-C. and Wang, H. (2001), "On Sample Size Calculation in Bioequivalence Trials," *Journal of Pharmacokinetics and Pharmacodynamics*, 28, 155-169.

Cornfield, J. (1951), "A Method of Estimating Comparative Rates from Clinical Data. Applications to Cancer of the Lung, Breast, and Cervix." *Journal of the National Cancer Institute*, 11, 1269-1275.

Correa-Villaseor, A., Stewart, W. F., Franco-Marina, F., and Seacat, H. (1995), "Bias from Nondifferential Misclassification in Case-Control Studies with Three Exposure Levels," *Epidemiology*, 6, 276-281.

Dahm, P. F., Gail, M. H., Rosenberg, P. S., and Pee, D. (1995), "Determining the Value of Additional Surrogate Exposure Data for Improving the Estimate of an Odds Ratio," *Statistics in Medicine*, 14, 2581-2598.

Davit, B., Conner, D., Fabian-Fritsch, B., Haidar, S., Jiang, X., Patel, D., Seo, P., Suh, K., Thompson, C., and Yu, L. (2008), "Highly Variable Drugs: Observations from Bioequivalence Data Submitted to the FDA for New Generic Drug Applications," *The AAPS Journal*, 10, 148-156, 10.1208/s12248-008-9015-x.

Davit, B. M., Nwakama, P. E., Buehler, G. J., Conner, D. P., Haidar, S. H., Patel, D. T., Yang, Y., Yu, L. X., and Woodcock, J. (2009), "Comparing Generic and Innovator Drugs: A Review of 12 Years of Bioequivalence Data from the United States Food and Drug Administration," *The Annals of Pharmacotherapy*, 43, 1583-1596.

Drews, C. D., Kraus, J. F., and Greenland, S. (1990), "Recall Bias in a Case-Control Study of Sudden Infant Death Syndrome," *International Journal of Epidemiology*, 19, 405-411.

El-Tahtawy, A. A., Jackson, A. J., and Ludden, T. M. (1995), "Evaluation of Bioequivalence of Highly Variable Drugs Using Monte Carlo Simulations. I. Estimation of Rate of Absorption for Single and Multiple Dose Trials Using Cmax," *Pharmaceutical Research*, 12, 1634-1641, 10.1023/A:1016288916410.

Endrenyi, L., Amidon, G. L., Midha, K. K., and Skelly, J. P. (1998), "Individual Bioequivalence: Attractive in Principle, Difficult in Practice," *Pharmaceutical Research*, 15, 1321-1325, 10.1023/A:1011972732530.

Espeland, M. A. and Hui, S. L. (1987), "A General Approach to Analyzing Epidemiologic Data that Contain Misclassification Errors," *Biometrics*, 43, 1001-1012.

FDA (2003), *Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products – General Considerations*, U.S. Food and Drug Administration, Rockville, MD.

Forbes, A. B. and Santner, T. J. (1995), "Estimators of Odds Ratio Regression Parameters in Matched Case-Control Studies with Covariate Measurement Error," *Journal of the American Statistical Association*, 90, 1075-1084.

Greenhouse, S. W. (1982), "Jerome Cornfield's Contributions to Epidemiology," *Biometrics*, 38, 33-45.

Greenland, S. (1988), "Statistical Uncertainty Due to Misclassification: Implications for Validation Substudies," *Journal of Clinical Epidemiology*, 41, 1167-1174.

— (1994), "Modelling Risk Ratios from Matched Cohort Data: An Estimating Equation Approach," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43, 223-232.

— (2008), "Maximum-Likelihood and Closed-form Estimators of Epidemiologic Measures Under Misclassification," *Journal of Statistical Planning and Inference*, 138, 528-538, special Issue: Statistical Design and Analysis in the Health Sciences.

Greer, B. A. (2008), "Bayesian and Pseudo-Likelihood Interval Estimation for Comparing Two Poisson Rate Parameters Using Under-Reported Data," Ph.D. thesis, Baylor University.

Gustafson, P., Le, N. D., and Saskin, R. (2001), "Case-Control Analysis with Partial Knowledge of Exposure Misclassification Probabilities," *Biometrics*, 57, 598-609.

Haidar, S., Davit, B., Chen, M.-L., Conner, D., Lee, L., Li, Q., Lionberger, R., Makhlouf, F., Patel, D., Schuirmann, D., and Yu, L. (2008a), "Bioequivalence Approaches for Highly Variable Drugs and Drug Products," *Pharmaceutical Research*, 25, 237-241, 10.1007/s11095-007-9434-x.

Haidar, S., Kwon, H., Lionberger, R., and Yu, L. (2008b), *Bioavailability and Bioequivalence*, Springer US.

Haidar, S., Makhlouf, F., Schuirmann, D., Hyslop, T., Davit, B., Conner, D., and Yu, L. (2008c), "Evaluation of a Scaling Approach for the Bioequivalence of Highly Variable Drugs," *The AAPS Journal*, 10, 450-454, 10.1208/s12248-008-9053-4.

Haynes, J. D. (1981), "Statistical Simulation Study of New Proposed Uniformity Requirement for Bioequivalency Studies," *Journal of Pharmaceutical Sciences*, 70, 673-675.

Joseph, L., Gyorkos, T. W., and Coupal, L. (1995), "Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard," *Am. J. Epidemiol.*, 141, 263-272.

Jurek, A. M., Greenland, S., Maldonado, G., and Church, T. R. (2005), "Proper Interpretation of Non-Differential Misclassification Effects: Expectations Vs Observations," *Int. J. Epidemiol.*, 34, 680-687.

Karalis, V., Macheras, P., and Symillides, M. (2005), "Geometric Mean Ratio-Dependent Scaled Bioequivalence Limits with Leveling-Off Properties," *Eur J Pharm Sci.*, 26, 54-61.

Karalis, V., Symillides, M., and Macheras, P. (2004), "Novel Scaled Average Bioequivalence Limits Based on GMR and Variability Considerations," *Pharmaceutical Research*, 21, 1933-1942, 10.1023/B:PHAM.0000045249.83899.ae.

Karunaratne, B. P. M. (1991), "Estimating the Odds Ratio Under Double Sampling," Ph.D. thesis, Texas AM University.

Kytariolos, J., Karalis, V., Macheras, P., and Symillides, M. (2006), "Novel Scaled Bioequivalence Limits with Leveling-off Properties," *Pharmaceutical Research*, 23, 2657-2664, 10.1007/s11095-006-9107-1.

Langholz, B. (2010), "Case-Control Studies = Odds Ratios: Blame the Retrospective Model," *Epidemiology*, 21, 10-12.

Lee, S.-C. and Byun, J.-S. (2008), "A Bayesian Approach to Obtain Confidence Intervals for Binomial Proportion in a Double Sampling Scheme Subject to False-Positive Misclassification," *Journal of the Korean Statistical Society*, 37, 393-403.

Leu, M., Czene, K., and Reilly, M. (2010), "Bias Correction of Estimates of Familial Risk from Population-Based Cohort Studies," *Int. J. Epidemiol.*, 39, 80-88.

Lie, R. T., Heuch, I., and Irgens, L. M. (1994), "Maximum Likelihood Estimation of the Proportion of Congenital Malformations Using Double Registration Systems," *Biometrics*, 50, 433-444.

Lyles, R. H. (2002), "A Note on Estimating Crude Odds Ratios in Case-Control Studies with Differentially Misclassified Exposure," *Biometrics*, 58, 1034-1037.

Magder, L. S. (2003), "Simple Approaches to Assess the Possible Impact of Missing Outcome Information on Estimates of Risk Ratios, Odds Ratios, and Risk Differences," *Controlled Clinical Trials*, 24, 411-421.

Midha, K. K., Rawson, M. J., and Hubbard, J. W. (1997), "Bioequivalence: Issues and Options," *Journal of Pharmacokinetics and Pharmacodynamics*, 25, 743-752, 10.1023/A:1025785902230.

Morrissey, M. J. and Spiegelman, D. (1999), "Matrix Methods for Estimating Odds Ratios with Misclassified Exposure Data: Extensions and Comparisons," *Biometrics*, 55, 338-344.

Nurminen, M. (1995), "To Use or Not to Use the Odds Ratio in Epidemiologic Analyses?" *European Journal of Epidemiology*, 11, 365-371.

— (2003), "Evolution Of Epidemiologic Methodology From The Statistical Perspective," *The Internet Journal of Epidemiology*, 1.

Paul, S. R. and Thedchanamoorthy, S. (1997), "Likelihood-Based Confidence Limits for the Common Odds Ratio," *Journal of Statistical Planning and Inference*, 64, 83-92.

Pawitan, Y. (2001), *In All Likelihood: Statistical Modeling and Inference Using Likelihood*, Oxford University Press Inc., New York.

Prescott, G. J. and Garthwaite, P. H. (2002), "A Simple Bayesian Analysis of Misclassified Binary Data with a Validation Substudy," *Biometrics*, 58, 454-458.

Rani, S. and Pargal, A. (2004), "Bioequivalence: An Overview of Statistical Concepts," *Indian Journal of Pharmacology*, 36, 209-216.

Reiczigel, J., Abonyi-Tth, Z., and Singer, J. (2008), "An Exact Confidence Set for Two Binomial Proportions and Exact Unconditional Confidence Intervals for the Difference and Ratio of Proportions," *Computational Statistics & Data Analysis*, 52, 5046-5053.

Riggs, K. E. (2006), "Maximum-Likelihood-Based Confidence Regions and Hypothesis Tests for Selected Statistical Models," Ph.D. thesis, Baylor University.

Rothman, K. J. (1986), *Modern Epidemiology*, Boston: Little, Brown, 1st ed.

Satten, G. A. and Kupper, L. L. (1990), "Sample Size Determination for Pair-Matched Case-Control Studies Where the Goal is Interval Estimation of the Odds Ratio," *Journal of Clinical Epidemiology*, 43, 55-59.

Schuirmann, D. J. (1987), "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Avarage Bioavailability," *J. Pharmacokinet. Biopharm.*, 15, 657-680.

Severini, T. A. (1998), "Likelihood Functions for Inference in the Presence of a Nuisance Parameter," *Biometrika*, 85, 507-522.

— (2000), *Likelihood Methods in Statistics*, no. 22 in Oxford Statistical Science Series, Oxford University Press Inc., New York.

Stamey, J. D., Boese, D. H., and Young, D. M. (2008), "Confidence Intervals for Parameters of Two Diagnostic Tests in the Absence of a Gold Standard," *Computational Statistics & Data Analysis*, 52, 1335-1346.

Tenenbein, A. (1970), "A Double Sampling Scheme for Estimating from Binomial Data With Misclassifications," *JASA*, 65, 1350-1361.

Thomas, D., Stram, D., and Dwyer, J. (1993), "Exposure Measurement Error: Influence on Exposure-Disease Relationships and Methods of Correction," *Annual Review of Public Health*, 14, 69-93.

Tothfalusi, L. and Endrenyi, L. (2003), "Limits for the Scaled Average Bioequivalence of Highly Variable Drugs and Drug Products," *Pharmaceutical Research*, 20, 382-389, 10.1023/A:1022695819135.

Tothfalusi, L., Endrenyi, L., Midha, K. K., Rawson, M. J., and Hubbard, J. W. (2001), "Evaluation of the Bioequivalence of Highly-Variable Drugs and Drug Products," *Pharmaceutical Research*, 18, 728-733, 10.1023/A:1011015924429.

Troendle, J. F. and Frank, J. (2001), "Unbiased Confidence Intervals for the Odds Ratio of Two Independent Binomial Samples with Application to Case-Control Data," *Biometrics*, 57, 484-489.

Walter, S. D. and Irwig, L. M. (1988), "Estimation of Test Error Rates, Disease Prevalence and Relative Risk from Misclassified Data: A Review," *Journal of Clinical Epidemiology*, 41, 923-937.

Westlake, W. J. (1972), "Use of Confidence Intervals in Analysis of Comparative Bioavailability Trials," *Journal of Pharmaceutical Sciences*, 61, 1340-1341.