ABSTRACT

Topics in Interval Estimation for Two Problems Using Double Sampling

Linda Njoh, Ph.D.

Chairperson: Dean M. Young, Ph.D.

This dissertation addresses two distinct topics. The first considers interval estimation methods of the odds ratio parameter in 2×2 cohort studies with misclassified data. That is, we derive two first-order likelihood-based confidence intervals and two pseudo-likelihood-based confidence intervals for the odds ratio in a 2×2 cohort study subject to differential misclassification and non-differential misclassification using a double-sampling paradigm for binary data. Specifically, we derive the Wald, score, profile likelihood, and approximate integrated likelihood-based confidence intervals for the odds ratio of a 2×2 cohort study. We then compare coverage properties and median interval widths of the newly derived confidence intervals via a Monte Carlo simulation. Our simulation results reveal the consistent superiority of the approximate integrated likelihood confidence interval, especially when the degree of misclassification is high.

The second topic is concerned with interval estimation methods of a Poisson rate parameter in the presence of count misclassification. More specifically, we derive multiple first-order asymptotic confidence intervals for estimating a Poisson rate parameter using a double sample for data containing false-negative and false-positive observations in one case and for data with only false-negative observations in another case. We compare the new confidence intervals in terms of coverage probability

and median interval width via a simulation experiment. We then apply our derived confidence intervals to real-data examples. Over the parameter configurations and observation-opportunity sizes considered here, our investigation demonstrates that the Wald interval is the best omnibus interval estimator for a Poisson rate parameter using data subject to over-and under-counts. Also, the profile log-likelihood-based confidence interval is the best omnibus confidence interval for a Poisson rate parameter using data subject to visibility bias.

Topics in Interval Estimation for Two Problems Using Double Sampling

by

Linda Njoh, 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

_____

Jack D. Tubbs, Ph.D.

_____

James D. Stamey, Ph.D.

_____

Jane L. Harvill, Ph.D.

_____

Joe C. Yelderman, Ph.D.

Accepted by the Graduate School
December 2013

_____

J. Larry Lyon, Ph.D., Dean

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

To Laure P. Fotso, Ph.D. – My mother

CHAPTER ONE

Introduction and Background

1.1 Overview

A well-known problem in statistical science is how to account for or to eliminate nuisance parameters from a model. While some parameters are relevant, others are merely required to complete a model. Pawitan (2001) states that nuisance parameters create most of the complications in likelihood theory. They often appear in problems as a natural consequence of our effort to use bigger and better models. Because nuisance parameters can dramatically impact the inference of the parameters of interest, accounting for nuisance parameters is important. Even if we are interested in all model parameters, our inability to view a multidimensional likelihood forces us to view individual parameters in isolation. While we consider one parameter, the remaining parameters become a nuisance. Consequently, to access the information for the parameter of interest, one must account for or eliminate the nuisance parameters.

Binary or binomial data are frequently encountered in a wide range of applications, including survey analysis, criminology, clinical medicine, and information technology. One also finds such data in epidemiology, which is the study of the distribution of health-related states and events in populations. The primary objective of an epidemiology study is to obtain a valid and precise estimate of the effect of an exposure on the occurrence of a disease in the source population of the study. The parameter of interest is the odds ratio, which remains perhaps the most popular relative measure of the exposure-disease relation in epidemiology to date (Nurminen, 1995). However, misclassification errors in the sampling and measurement of subjects in a study can cause systematic errors in the estimator of the odds ratio.

Count data or data with an underlying Poisson distribution can also be found in the areas of epidemiology, market research, and criminal justice. Researchers often propose Poisson models to compare the rates of certain events for different populations. For instance, one might use Poisson models to compare mortality rates for different diseases. Market researchers use such models to analyze purchasing trends, and investigators in criminal justice studies use them to compare crime rates for different neighborhoods. In all three examples, the counts used to estimate the Poisson rate of interest could be subject to error; that is, inferences on rates using discrete counts can be inaccurate due to misclassification (Stamey, Young, and Stephens, 2005).

Bross (1954) has shown that when misclassification is present, the sample proportion is a biased estimator of the population proportion p and that the bias, which is a function of the amount of misclassification in the data, can be substantial. Furthermore, Whittemore and Gong (1991) and Sposto, Preston, Shimizu, and Mabuchi (1992) have shown that misclassification in the data collection process can lead to incorrect counts that adversely affect our inferences. That is, misclassification of count occurrences causes a biased estimator of the Poisson rate of interest. In order to correct for the bias, Tenenbein (1970) has developed a double-sampling plan for obtaining an unbiased estimate of the population proportion of binary data with misclassification. His double-sampling scheme is used in the following chapters to derive nearly unbiased estimators and approximate confidence intervals (CIs) for the odds ratio and the Poisson rate. However, Tenenbein's scheme produces one or more nuisance parameters that must be accounted for or eliminated.

In this dissertation, we use maximum likelihood and pseudo-likelihood estimations to derive the Wald, score, profile likelihood (PL), integrated likelihood (IL), and approximate integrated likelihood (AIL) CIs. In Chapters Two and Three, we compare three interval estimation methods for the odds ratio that measures the as-

sociation between disease and exposure levels under a double-sampling procedure for binomial data with two types of misclassification. In Chapters Four and Five, we derive and compare likelihood and pseudo-likelihood confidence intervals for the Poisson rate parameter using a double sample of infallible data and fallible data subject to misclassification in the form of false-negative and/or false-positive counts.

The remainder of Chapter One is organized as follows. Cohort studies are defined in Section 1.2. In Section 1.3 Tenebein's double-sampling plan is presented. In Section 1.4 we define the different types of misclassification used in this work. Section 1.5 briefly introduces the four confidence interval methods we utilize in Chapters Two through Five. We provide an outline of the dissertation organization in Section 1.6.

1.2 Cohort Studies

A cohort study is an observational study in which a sample or a cohort is selected and information is obtained to determine which subjects have a particular characteristic (e.g., blood group A) that is suspected of being related to the development of the disease under investigation or have been exposed to a possible etiological agent (e.g., cigarette smoking). The entire study sample is then followed up in time, and the incidence of the disease in the exposed individuals is compared with the incidence in those not exposed. Two main types of cohort studies are defined according to the point of time when information on exposure was collected: prospective cohort studies or retrospective cohort studies. In prospective cohort studies, data on exposure is collected once the study population has been defined. The main disadvantage of this type of cohort study, is that the time from exposure to onset of disease may be too long. The alternative, particularly useful for conditions with long induction periods, is to rely on exposure measurements made many years before the study was set up. These measurements may be available from medical, employment, or other

personal records. By using data from existing records, we can reduce or even eliminate waiting for the exposure to affect the risk of disease. This type of cohort study is called a retrospective cohort study. One of the main limitations of retrospective cohort studies is that the exposure data available in past records are generally less accurate and detailed than those collected prospectively.

The main advantages of cohort studies are that the exposure is measured before onset of disease and is therefore likely to yield an unbiased viewpoint in terms of disease development; rare exposures can be examined by appropriate selection of study cohorts; multiple outcomes (diseases) can be studied for any one exposure; and incidence of disease can be measured in the exposed and unexposed groups. The main disadvantages of this type of study are that they can be very expensive and time-consuming, particularly if conducted prospectively; changes in exposure status and in diagnostic criteria over time can affect the classification of individuals according to exposure and disease status; ascertainment of outcome may be influenced by knowledge of the subject's exposure status (information bias); and losses to follow-up may introduce selection bias.

In Chapters Two and Three, we investigate estimation of the odds ratio in a cohort study using the double-sampling scheme for binary data subject to differential and non-differential misclassification.

1.3 Double Sampling

Binary data are usually obtained when experimental units are classified into two mutually exclusive categories. Generally, a statistical classifier is not perfect. Hence, misclassified binary data can occur. Many researchers have demonstrated that classical estimators that ignore misclassification are biased when applied to misclassified binary data (Rahardja and Young, 2011). In particular, Bross (1954) has shown that, when misclassification occurs, the sample proportion is a biased

estimate of the parameter p and that the bias, which is a function of the amount of misclassification in the data, can be substantial. In order to adjust for this bias, we must gain some knowledge of the amount of misclassification in the data. Tenenbein (1970) states that one method of obtaining information on the extent of misclassification is to compare results obtained by two or more measuring devices involving the same group of sampling units.

Suppose that two measuring devices are available to classify experimental units into one of two mutually exclusive categories. Suppose the use of the first device is an expensive procedure that classifies units correctly. While the use of the second device is cheaper, it tends to misclassify units. Tenenbein (1970) has proposed a double-sampling scheme to obtain an unbiased estimator for the population proportion π of binary data with misclassification.

Suppose we conduct a test that allows us to obtain a disease status on a large sample of participants. However, such an instrument, although fast, inexpensive, and perhaps non-invasive, can be fallible. Hence, the counts we observe will have errors due to misclassification, thus causing a biased estimator of the parameter of interest, the odds ratio from a 2×2 cohort study. Another example could be one in which a researcher compares the colon cancer mortality rates for obese and non-obese populations. Misclassification could occur in the data if a subject is assigned an incorrect cause of death, which leads to a biased estimator of the Poisson rate of interest.

To calculate the misclassification rate and account for the induced 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 on which we use boh tests 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, Gail, Rosenberg, and Pee (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.

1.4 Differential and Nondifferential Misclassification

When analyzing misclassified data, we focus on whether the misclassification rates between the fallible classifier or error-prone test and the infallible classifier or "gold-standard" test depend on the error-free disease status of the patient (Tenenbein, 1970). Thus, we define the test sensitivity of an exposure measurement method as the probability that an individual who is truly exposed is classified as exposed by the method and the test specificity of an exposure measurement method as the probability that someone who is truly unexposed will be classified as unexposed. Hence, high levels of specificity and sensitivity indicate low misclassification rates.

Misclassification errors are non-differential if the classification between fallible exposure level and "gold-standard" exposure level is independent of the disease status. More simply, non-differential misclassification arises when measurement errors with regard to exposure are independent of disease status. Hence, for non-differential misclassification, we assume that the specificity and sensitivity are independent of the disease status. The assumption of non-differential misclassification is usually plausible for prospective cohort studies in which one determines the exposure status before the disease status is measured (Dahm, Gail, Rosenberg, and Pee, 1995).

Differential misclassification occurs when the error rate or probability of being misclassified differs across groups of study subjects. For example, if the exposed

group in a cohort study is more likely to be mistakenly classified as having developed the disease than the exposed group, then the misclassification is differential. In such instances, the sensitivity and specificity between diseased and non-diseased groups are assumed to be different.

1.5 Four First-Order Asymptotic Confidence Interval Methods

Let $\mathbf{x} = (x_1, x_2, ..., x_n)'$ be a random vector of n iid observations from the probability density $f(\boldsymbol{\theta}|\mathbf{x})$. Here, $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)'$ is a parameter vector of dimension p contained in the parameter space $\boldsymbol{\Theta} \in \Re_p$. The likelihood is a function of $\boldsymbol{\theta}$ for the given observation vector $\mathbf{x}$ and, by the likelihood principle, contains all the information concerning the experiment of interest (Bjornstad (1996)). Thus, inferences about $\boldsymbol{\theta}$ should depend on the random variable only through the likelihood function defined as

$$L(\boldsymbol{\theta}|\mathbf{x}) = f(\boldsymbol{\theta}|\mathbf{x}).$$

Berger, Liseo, and Wolpert (1999) notice that because $\boldsymbol{\theta}$ is a vector of parameters, one might be interested in only one parameter or a subset of parameters. Thus, we partition $\boldsymbol{\theta}$ into the parameter of interest, the odds ratio, ψ or the Poisson rate, λ, and $\boldsymbol{\eta}$, the set of nuisance parameters. For inferences in the following chapters, we propose and describe five confidence interval (CI) methods for estimating the parameter of interests, the odds ratio, and the Poisson rate. After comparing the intervals in terms of coverage probability and average width, we consider and define in subsequent chapters the Wald and score CI as part of the maximum likelihood-based CI as well as the profile likelihood, and the approximate integrated likelihood, which are pseudo-likelihood-based CIs.

In Chapters Two and Three, we conduct Monte-Carlo simulations to compare confidence intervals for the odds ratio using double sampling for cohort studies under both differential and non-differential misclassification. We use the abbrevi-

ation CODIFF for cohort studies with differential misclassification and CONDIFF for cohort studies with non-differential misclassification.

1.6 Dissertation Outline

In this dissertation we cover two distinct topics. We first consider the interval estimation of the odds ratio parameter in cohort studies with data subject to misclassification. Second, we dwell on the interval estimation of the Poisson rate using data subject to misclassification. Each chapter of the dissertation is an independent unit with individual literature reviews and conclusions. In Chapters Two and Three we restrict ourselves to cohort studies subject to differential and non-differential misclassification, respectively. In Chapter Four, we consider the interval estimation of a Poisson rate using data subject to misclassification in the form of false-negative and false-positive counts. Finally, in Chapter Five, we derive interval estimation methods for a Poisson rate using data subject to visibility bias or under-reporting. In all chapters, we propose a model and derive numerically maximum likelihood estimators (MLEs) and the restricted maximum (RMLEs) likelihood estimators. We then use the MLEs to derive multiple CIs for the odds ratio parameter or the Poisson rate parameter of interest. In the last part of each chapter, we conduct a Monte Carlo simulation study to compare the proposed intervals. In Chapters Four and Five, we also apply the newly derived CIs to real data examples.

CHAPTER TWO

Three Approximate Confidence Intervals for the Odds Ratio in a 2 × 2 Cohort Study with Differential Misclassification

2.1 Abstract

We derive two first-order likelihood-based confidence intervals (CIs) and one pseudo-likelihood-based CI for the odds ratio in a 2 × 2 cohort study subject to differential misclassification using the double sampling paradigm for binary data. Specifically, we derive the Wald, score, the approximate integrated likelihood (AIL)-based CIs. We compare average coverage properties and median interval widths of the newly derived CIs via a Monte Carlo simulation. We conclude that the AIL interval is superior in terms of average coverage and interval width properties to the Wald and score CIs for the parameter configurations we study.

2.2 Introduction

The typical goal of a cohort study is to compare the incidence of disease in one or more study cohorts. If there are two cohorts in the study, then the one with individuals who have experienced a putative causal event or condition is referred as the exposed cohort, and the other is referred to as the unexposed, or reference, cohort (Rothman, Greenland, and Lash, 2008). The odds ratio (OR) is the odds that an outcome will occur given a particular exposure, divided by the odds of the outcome occurring in the absence of that exposure. Odds ratios are most commonly used in case-control studies but may also be used in cohort study designs as well (Szumilas, 2010).

Differential misclassification is a problem that can arise in many cohort studies. This systematic error leads to an over- or under-estimation of the actual magnitude of the OR that can produce considerable bias in the OR estimator, which depends

on the proportion of subjects that are misclassified. Epidemiologists have long recognized that measurement errors have been among the major weaknesses of their studies and have gone to great lengths to try to assess the magnitude of such errors and their likely impact on the conclusions. This problem has spurred much research, initially with respect to understanding the effects of measurement error on exposure-response relationships and more recently on developing methods to correct such errors. For instance, Gustafson, Le, and Saskin (2001) have considered a case-control analysis with a dichotomous exposure variable that is subject to misclassification. Their research shows that if the classification probabilities are known, methods are available to adjust odds-ratio estimates in light of misclassification. Walter and Irwig (1988) have reviewed methods for the analysis of categorical, clinical, and epidemiological data, in which the observations are subject to misclassification. They have found that under certain conditions, one can estimate error parameters such as sensitivity, specificity, relative risk, or predictive value, even though no gold standard is available. Thomas, Stram, and Dwyer (1993) have reviewed a variety of methods that have been suggested for adjusting exposure-response relationships for measurement error and their relationships.

A large research literature is available on binary data subject to misclassification that provides point and interval estimation methods for various functions of the proportion parameters. For one-sample problems, several researchers have considered the case in which only one type of error is present. For example, Lie, Heuch, and Irgens (1994) have used a maximum likelihood approach, where false-negative errors were corrected using fallible classifiers. York, Madigan, Heuch, and Lie (1995) have considered this same problem from a Bayesian perspective. When data are obtained using a double-sampling scheme, Moors, Van Der Genugten, and Strijbosch (2000) have discussed the method of moment and maximum likelihood estimation, in addition to one-sided interval estimation. Boese, Young, and Stamey

(2006) have derived several likelihood and pseudo-likelihood-based CIs for a single proportion parameter, while Lee and Byun (2008) have provided Bayesian credible sets using non-informative priors for the same problem.

Moreover, several researchers have also studied one-sample problems with both types of misclassification error. In conjunction with double sampling, Tenenbein (1970) has proposed a maximum likelihood estimator (MLE) for a proportion parameter and has derived an expresssion for the asymptotic variance. For the case when training data is unavailable in one-sample problems, Gaba and Winkler (1992) and Viana, Ramakrishnan, and Levy (1993) have developed Bayesian approaches using sufficiently informative priors.

Here, we derive three approximate CIs for the OR parameter in a 2×2 cohort study that uses double sampling. The Wald and score intervals are likelihood based, and the AIL interval is based on the integrated or marginal likelihood. We consider the case of differential misclassification for both fallible and infallible samples; 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 and then derive the three CIs mentioned above.

This chapter is organized as follows. We present the model and the double-sampling scheme in Section 2.3. In Section 2.4, we derive the MLES of the parameter of interest and the model nuisance parameters, which are used to derive the Wald CI. To derive the other two CIs, we define the restricted maximum likelihood estimators (RMLEs) of the nuisance parameters and describe their derivation in Section 2.5. In Section 2.6, we derive the observed Fisher information (OFIM) matrix and then derive the Wald, score, and AIL CIs. In Section 2.7, we utilize Monte Carlo simulation methods to compare the coverage and interval width properties of the CIs. Finally, we comment on the simulation results in Section 2.8.

2.3 The Model

The individuals in a 2×2 cohort study either have been exposed or have not been exposed to a certain medical condition. We define the binary random variable D as the disease level ($D = 1$ for diseased, $D = 0$ for not diseased) for each individual in the population. Following the double-sampling scheme introduced by Tenenbein (1970), we assume that two testing procedures are available to determine the exposure level of each participant. In the first stage, we classify all individuals in the study using solely a fallible procedure. Let Z denote the fallible exposure level, where $Z = 1$ represents an individual who is classified by the fallible test as exposed and $Z = 0$ represents an individual who is classified by the fallible procedure as not exposed. Such a procedure is usually fast, inexpensive, and non-invasive 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 gold standard exposure level, where $X = 1$ represents an individual who is classified as exposed by the gold standard test, and $X = 0$ represents an individual who is classified as not exposed by the gold standard test. This test is often so expensive, time-consuming, and invasive that it is performed on only a small group of individuals. Because we know that the gold standard test is absolutely accurate, we use the parameter estimates from this sub-sample to correct for the estimator bias (Tenenbein, 1970).

Let $Z = j$, $X = k$, and $D = i$ represent the fallible test outcome, gold standard test outcome, and actual disease outcome, respectively, where $i, j, k = 0, 1$. The cell count for the fallible data in the cell with $D = i$, and $Z = j$ is denoted by W_{ij}, and the cell count for the validation data or double sampled-data for the cell with $X = k$, $D = i$ and $Z = j$ will be denoted by V_{kij}.

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 $M_k + N_i$ for each $i, k = 0, 1$. For a visual representation of the above notations, refer to Table 2.1.

Table 2.1: Counts for a Study with Misclassified Exposure Data

	Validation Study (Complete)				fallible Study (Incomplete)	
Fallible	$X=1$	$X=0$	$X=1$	$X=0$		
(Z)	$D=1$	$D=1$	$D=0$	$D=0$	$D=1$	$D=0$
$Z=1$	V_{111}	V_{011}	V_{101}	V_{001}	W_{11}	W_{01}
$Z=0$	V_{110}	V_{010}	V_{100}	V_{000}	W_{10}	W_{00}
	M_1		M_0		N_1	N_0

For the infallible classifier study, we denote the joint probability of exposure and the disease level categories.

$$\pi_i = Pr(X = 1, D = i), \tag{2.1}$$

where $X = 1$ for the i^{th} group $(D = i)$ with $i = 0, 1$. Also, for the fallible test or classifier, we define the sensitivity, S_i, as the probability that an individual tests positive for the fallible test $(Z = 1)$ and the same individual tests positive on the gold standard procedure $(X=1)$ for each of the disease-level categories $D = i, i = 0, 1$. Thus,

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

where $i = 0,\ 1$. We denote the probability that an individual does not have the disease according to the fallible test $(Z = 0)$ and has a negative result on the gold standard procedure (X=0) for disease categories $D = i$, $i = 0, 1$, as

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

where $i = 0, 1$. We remark that we assume distinct values for the specificity and sensitivity for each of the disease-level group, which we refer to as differential misclassification.

Based on (2.1) - (2.3) and the derivations given in Prescott and Garthwaite (2002), the distribution on the observable counts for the validation and fallible studies are

$$(V_{1i1}, V_{1i0}, V_{0i1}, V_{0i0}) \sim Multi(\pi_i S_i, \pi_i(1-S_i), (1-\pi_i)(1-C_i), (1-\pi_i)C_i) \quad (2.4)$$

and

$$W_{i1} \sim Bin(N_i, \pi_i S_i + (1-\pi_i)(1-C_i)),$$

respectively, where $i = 0, 1$. The OR ψ that associates the gold standard exposure level X with the disease outcome D is

$$\psi \equiv \frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}. \quad (2.5)$$

Dahm, Gail, Rosenberg, and Pee (1995) have provided an example of a prospective cohort study of the prognostic value of a renal biopsy (X) on a five-year survival rate ($D = 1$, if death is within five years). One might wish to perform renal biopsy studies on no more than 100 patients because of the possible complications and discomfort. Instead of a renal biopsy, one might wish to use a non-invasive test (Z) to obtain additional information on the prognostic significance of a renal biopsy by studying an additional sample of individuals using non-invasive tests only. Also, Pepe (1992) has reported a cohort study that investigated whether aplastic anemia patients who had undergone bone marrow transplants would develop graft-versus-host disease (GVHD). The data sets consisted of 179 subjects in two groups: those aged below 20 and those above 20. Two classifiers were used to classify the patients. The first was a gold standard test (X), referred to as "Chronic GVHD," an expensive test that obtained responses via the long-term follow-up of each patient. The other was the fallible classifier (Z), referred to as "Acute GVHD," which could be measured instantly and was adopted as a sensible surrogate for "Chronic GVHD."

2.4 Maximum-Likelihood Estimators

Let $\boldsymbol{\eta} \equiv (\pi_1, C_1, C_0, S_1, S_0)'$ be the nuisance parameter vector and let $\boldsymbol{d} \equiv (W_{0i}, W_{1i}, V_{1i1}, V_{0i1}, V_{1i0}, V_{0i0})'$, where $i = 0, 1$ (0 for diseased, 1 for non-diseased), denote the observed counts for both fallible and validation studies. If we let $L_o \equiv L_o(\pi_1, \pi_0, C_1, C_0, S_1, S_0|\boldsymbol{d})$, then the concentrated observed-likelihood function is

$$L_o \propto \prod_{i=0}^{1}[\pi_i S_i + (1-\pi_i)(1-C_i)]^{W_{i1}}[1-\pi_i S_i - (1-\pi_i)(1-C_i)]^{N_i - W_{i1}} \times$$
$$(\pi_i S_i)^{V_{1i1}}[\pi_i(1-S_i)]^{V_{1i0}}[(1-\pi_i)(1-C_i)]^{V_{0i1}} \times [(1-\pi_i)C_i]^{[M_i-(V_{1i1}+V_{1i0}+V_{0i1})]}.$$

Because our interest is the estimation of ψ, using (2.5), we solve for π_0 and then substitute for its expression into the likelihood function L_o. The transformed concentrated log-likelihood is

$$\begin{aligned}
\ell_\psi \propto\ & W_{11}\ln[\pi_1 S_1 + (1-\pi_1)(1-C_1)] + W_{01}\ln\left[\frac{(1-C_0)(1-\pi_1)\psi + \pi_1 S_0}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + (N_1 - W_{11})\ln[1-\pi_1 S_1 - (1-\pi_1)(1-C_1)] + V_{101}\ln\left[\frac{\pi_1 S_0}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + (N_0 - W_{01})\ln\left[\frac{C_0\psi + \pi_1 - C_0\psi\pi_1 - \pi_1 S_0}{\psi + \pi_1 - \psi\pi_1}\right] + V_{100}\ln\left[\frac{(\pi_1 - \pi_1 S_0)}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + V_{111}\ln(\pi_1 S_1) + V_{110}\ln[\pi_1(1-S_1)] + V_{011}\ln[(1-\pi_1)(1-C_1)] \\
& + [M_1 - (V_{111} + V_{110} + V_{011})]\ln[(1-\pi_1)C_1] + V_{001}\ln\left[\frac{(1-C_0)(1-\pi_1)\psi}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + [M_0 - (V_{101} + V_{100} + V_{001})]\ln\left[\frac{C_0\psi - \pi_1\psi C_0}{\psi + \pi_1 - \psi\pi_1}\right].
\end{aligned} \tag{2.6}$$

We next use (2.6) and calculations from Joseph et al. (1995) and Prescott and Garthwaite (2002) to obtain the MLEs

$$\hat{\pi}_i = \frac{(V_{0i0} + V_{1i0} + W_{i0})(V_{1i0}V_{0i1} + V_{1i0}V_{1i1})}{(V_{0i0} + V_{0i1} + V_{1i0} + V_{1i1} + W_{i0} + W_{i1})(V_{0i0} + V_{1i0})(V_{0i1} + V_{1i1})}$$
$$+ \frac{(V_{0i1} + V_{1i1} + W_{i1})(V_{1i1}V_{0i0} + V_{1i1}V_{1i0})}{(V_{0i0} + V_{0i1} + V_{1i0} + V_{1i1} + W_{i0} + W_{i1})(V_{0i0} + V_{1i0})(V_{0i1} + V_{1i1})},$$

$$\hat{S}_i = \frac{(V_{0i0} + V_{1i0})V_{1i1}(V_{0i1} + V_{1i1} + W_{i1})}{(V_{0i1} + V_{1i1})\left[(V_{0i0} + V_{1i0})(V_{1i0} + V_{1i1}) + V_{1i0}W_{i0}\right] + (V_{0i0} + V_{1i0})V_{1i1}W_{i1}},$$

$$\hat{C}_i = \frac{V_{0i0}(V_{0i1}+V_{1i1})(V_{0i0}+V_{1i0}+W_{i0})}{(V_{0i1}+V_{1i1})\left[(V_{0i0}+V_{1i0})(V_{0i0}+V_{0i1})+V_{0i0}W_{i0}\right]+(V_{0i0}+V_{1i0})V_{0i1}W_{i1}},$$

where $i = 0, 1$. By the invariance property of the MLEs, we have that

$$\hat{\psi} = \frac{(V_{001}+V_{101})\left[(V_{000}+V_{100})(V_{000}+V_{001})+V_{000}W_{00}\right]+(V_{000}+V_{100})V_{001}W_{01}}{(V_{001}+V_{101})\left[(V_{000}+V_{100})(V_{100}+V_{101})+V_{100}W_{00}\right]+(V_{000}+V_{100})V_{101}W_{01}}$$

$$\times \frac{(V_{011}+V_{111})[(V_{010}+V_{110})(V_{110}+V_{111})+V_{110}W_{10}]+(V_{010}+V_{110})V_{111}W_{11}}{(V_{011}+V_{111})\left[(V_{010}+V_{110})(V_{010}+V_{011})+V_{010}W_{10}\right]+(V_{010}+V_{110})V_{011}W_{11}}. \tag{2.7}$$

One can find a more detailed description of the above MLE derivations in Appendix A.1.

2.5 Restricted Maximum-Likelihood Estimators

Two of the CIs we consider here require evaluations of RMLEs for the nuisance parameters. To calculate the score and AIL CIs, we must evaluate the likelihood function using the RMLEs to eliminate the nuisance parameters. We next describe how we obtain the RMLEs of the nuisance parameters.

First, consider the cell counts for the validation study (refer to Table 2.1) and recall that these counts are subject to misclassification. We define a set of new latent variables, U_{1ij}, for the unobserved misclassified data counts under the assumption that an infallible test was performed. The subscripts $i, j = 0, 1$, correspond to the outcomes of the true disease condition (D) and the fallible test (Z), respectively, for each patient in the study. In Table 2.2 we show that U_{1ij} are the unobserved portions of the observed counts W_{ij}. Also, from previous derivations and the assumption that

Table 2.2: Counts for the Fallible Study with Unobserved, Misclassified Data

	Fallible Study (Incomplete)			
Fallible	$X = 1$	$X = 0$	$X = 1$	$X = 0$
(Z)	$D = 1$	$D = 1$	$D = 0$	$D = 0$
$Z = 1$	U_{111}	$W_{11} - U_{111}$	U_{101}	$W_{01} - U_{101}$
$Z = 0$	U_{110}	$W_{10} - U_{110}$	U_{100}	$W_{00} - U_{100}$
	$U_{111} + U_{110}$	$N_1 - (U_{111} + U_{110})$	$U_{101} + U_{100}$	$N_0 - (U_{101} + U_{100})$

the validation study is a sub-sample of the main study, we have

$$(U_{1i1} + U_{1i0}) \sim Bin(N_i, \pi_i),$$

$$U_{1i1}|(U_{1i1} + U_{1i0}) \sim Bin(U_{1i1} + U_{1i0}, S_i),$$

and

$$(W_{i0} - U_{1i0})|(U_{1i1} + U_{1i0}) \sim Bin(N_i - (U_{1i1} + U_{1i0}), C_i),$$

where $i = 0, 1$. See Joseph, Gyorkos, and Coupal (1995) for more details. Let $\boldsymbol{d}^{full} \equiv (W_{0i}, W_{1i}, V_{1i1}, V_{0i1}, V_{1i0}, V_{0i0}, U_{i11}, U_{i10})'$, where $i = 0, 1$ represents the full data vector. The complete-data likelihood function $L_U \equiv L_U(\pi_1, \pi_0, C_1, C_0, S_1, S_0|\boldsymbol{d}^{full})$ is detailed in Appendix A.2. Next, we construct an EM algorithm to determine the RMLEs for a fixed value of ψ because no closed-form solutions for the RMLEs exist. The EM algorithm steps are outlined in Appendix 1.2.1.

2.6 Three Approximate Confidence Intervals for ψ

In this section we derive three approximate CIs for the OR ψ. All three CIs utilize the observed Fisher information matrix(OFIM) described below.

2.6.1 The Observed Fisher Information Matrix

We use the inverse of the OFIM to estimate the covariance matrix of the MLEs and then construct the likelihood and pseudo-likelihood-based CIs for ψ. Recall that we have five nuisance parameters because we assume differential misclassification and have distinct values for the specificity and sensitivity for each of the disease-level group. Let $\boldsymbol{\theta} = (\psi, \boldsymbol{\eta}')'$ represent our vector of parameters, where ψ is the OR.

Then, the OFIM is

$$\mathbf{J}(\psi,\boldsymbol{\eta}) \equiv -\begin{bmatrix} \frac{\partial^2\ell_\psi}{\partial\psi^2} & \frac{\partial^2\ell_\psi}{\partial\psi\partial\pi_1} & \frac{\partial^2\ell_\psi}{\partial\psi\partial S_0} & \frac{\partial^2\ell_\psi}{\partial\psi\partial S_1} & \frac{\partial^2\ell_\psi}{\partial\psi\partial C_0} & \frac{\partial^2\ell_\psi}{\partial\psi\partial C_1} \\ \cdot & \frac{\partial^2\ell_\psi}{\partial\pi_1^2} & \frac{\partial^2\ell_\psi}{\partial\pi_1\partial S_0} & \frac{\partial^2\ell_\psi}{\partial\pi_1\partial S_1} & \frac{\partial^2\ell_\psi}{\partial\pi_1\partial C_0} & \frac{\partial^2\ell_\psi}{\partial\pi_1\partial C_1} \\ \cdot & \cdot & \frac{\partial^2\ell_\psi}{\partial S_0^2} & \frac{\partial^2\ell_\psi}{\partial S_0\partial S_1} & \frac{\partial^2\ell_\psi}{\partial S_0\partial C_0} & \frac{\partial^2\ell_\psi}{\partial S_0\partial C_1} \\ \cdot & \cdot & \cdot & \frac{\partial^2\ell_\psi}{\partial S_1^2} & \frac{\partial^2\ell_\psi}{\partial S_1\partial C_0} & \frac{\partial^2\ell_\psi}{\partial S_1\partial C_1} \\ \cdot & \cdot & \cdot & \cdot & \frac{\partial^2\ell_\psi}{\partial C_0^2} & \frac{\partial^2\ell_\psi}{\partial C_0\partial C_1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \frac{\partial^2\ell_\psi}{\partial C_1^2} \end{bmatrix}, \tag{2.8}$$

where ℓ_ψ is given in (2.6). In Appendix A.3, we give expressions for each of the elements in (2.8). Also, we partition (2.8) so that

$$\mathbf{J}\left(\psi,\boldsymbol{\eta}\right) \equiv \begin{bmatrix} J_{11} & \boldsymbol{J}_{12} \\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22} \end{bmatrix}, \tag{2.9}$$

where $J_{11} = J_\psi$ is a scalar. We use (2.9) in Section 2.6.2 below to develop the score and AIL CIs for ψ.

2.6.2 A Wald Confidence Interval for ψ

We first describe a Wald CI for ψ. The Wald statistic for ψ with nuisance vector $\boldsymbol{\eta}$ 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 CI for the parameter ψ consists of the values of ψ that satisfy

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

where $\chi_1^2(1-\alpha)$ denotes the $(1-\alpha)$ quantile of a central chi-squared distribution with one degree of freedom. Thus, an approximate $(1-\alpha)\%$ Wald CI for ψ under the double-sampling paradigm with differential misclassification is

$$\hat{\psi} - \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\psi},\hat{\boldsymbol{\eta}})]} < \psi < \hat{\psi} + \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\psi},\hat{\boldsymbol{\eta}})]} \tag{2.10}$$

2.6.3 A Score Confidence Interval for ψ

Second, we describe a score CI for the OR ψ, which is a likelihood related interval. The score interval with nuisance parameters, based on Rao's score statistic, is

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

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

$$\left[S(\psi, \hat{\boldsymbol{\eta}}_\psi)\right]^2 \left[J^{11}(\psi, \hat{\boldsymbol{\eta}}_\psi)\right] < \chi^2_1(1-\alpha), \tag{2.11}$$

where $\hat{\boldsymbol{\eta}}_\psi$ is the restricted MLE (RMLE) of the nuisance vector, $\boldsymbol{\eta}$, evaluated at a fixed ψ. We evaluate the OFIM at the RMLEs of the nuisance parameters for each fixed value of ψ. Because one cannot directly solve (2.11) for ψ, we use an EM algorithm described in Appendix A.2 and a bisectional root-finding method to numerically determine the CI (2.11).

2.6.4 An Approximate Integrated Likelihood Confidence Interval for ψ

Last, we describe a CI based on the approximate integrated likelihood. In most cases, integrating out the nuisance parameters, as in the integrated likelihood, can be computationally intensive or even intractable. We therefore employ the Laplace approximation to obtain a more computationally feasible likelihood and express the approximate integrated likelihood function as

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

where $c = (2\pi)^{\nu/2}$, ν is the dimension of the nuisance parameter and

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

where $\hat{\boldsymbol{\eta}}_\psi$ is the restricted or profile MLE of $\boldsymbol{\eta}$ and the profile likelihood function. Notice that $\boldsymbol{J}_{\boldsymbol{\eta}}$ is the nuisance parameter sub-matrix of the OFIM. Thus, $\hat{\boldsymbol{J}}_{\boldsymbol{\eta}} \equiv \boldsymbol{J}_{22}$

in (2.9) evaluated at the nuisance parameter MLEs. We use the OFIM elements given in Appendix A.3, the MLEs shown in Section 2.4 and an iterative bisectional root-finding method to determine a CI composed of the values of ψ that satisfy the inequality

$$-2\left[\ell_{AI}(\psi) - \ell_{AI}(\hat{\psi}_{AI})\right] < \chi_1^2(1-\alpha), \tag{2.12}$$

where $\ell_{AI}(\psi)$ is the approximate integrated log likelihood evaluated at ψ.

2.7 A Monte Carlo Simulation

We compared the performance of the three proposed CIs for ψ based on coverage probability and median interval width properties via a Monte Carlo simulation. Specifically, we examined the effect of the sample sizes of both cohorts, M_i and N_i, respectively, for $i = 0, 1$, the effect of the disease probability π_k, where $k = 0, 1$, for the gold standard test outcome $X = 0, 1$, and the magnitude of ψ, on the coverage probability and median interval width of the CIs defined in (2.10), (2.11) and (2.12).

2.7.1 Simulation Parameter and Sample-Size Configurations

In this simulation study, we assume two values for the odds ratio ψ: $\psi = 2$ and $\psi = 4$. Also, $\pi_i = Pr(X = 1, D = i)$, where $i = 0, 1$, represents the joint probabilities of exposure and disease status. Here, $X = 1$ indicates that the participant was classified as exposed by the gold standard and $D = 0$ or 1 indicated the condition of the participant as not-diseased or diseased, respectively. Next, we defined the sensitivity and specificity probabilities for the simulation in this cohort study with differential misclassification (CODIFF). Recall that the sensitivities and specificities are $S_i = Pr(Z = 1|X = 1, D = i)$ and $C_i = Pr(Z = 0|X = 0, D = i)$, where $i = 0, 1$, respectively. Due to the high degree of complexity involved in the simulation over the parameter space, we considered only low and high misclassification with the corresponding sensitivity and specificity probability values displayed in Table 2.3. We examined the CIs for eight different sample sizes in each of the described situa-

Table 2.3: Parameter Configurations Used in the Simulation for CODIFF

Conf.	ψ	π_0	π_1	Misclassification	C_0	C_1	S_0	S_1
$C_{2.1}$	2	0.25	0.40	low	0.99	0.97	0.98	0.96
$C_{2.2}$	4	0.38	0.71	low	0.99	0.97	0.98	0.96
$C_{2.3}$	2	0.25	0.40	high	0.90	0.85	0.80	0.75
$C_{2.4}$	4	0.38	0.71	high	0.90	0.85	0.80	0.75

tions – low misclassification with $\psi = 2$ and high misclassification with $\psi = 4$. The sample sizes for each of the considered validation and fallible studies are displayed in Table 2.4. Refer to Table 2.1 for the definition of M_i and N_i, $i = 0, 1$. We re-

Table 2.4: Sample Sizes Used in the Simulation for CODIFF

	A1	A2	A3	A4	A5	A6	A7	A8
M_0	100	100	150	200	250	300	350	400
M_1	60	60	74	100	124	150	174	200
N_0	500	1000	1500	2000	2500	3000	3500	4000
N_1	240	500	740	1000	1240	1500	1740	2000

mark that the sample sizes for the validation studies are considerably smaller when compared to the corresponding sample sizes for the fallible studies. We generated 10,000 data sets for each parameter configuration given in Table 2.3 and calculated the three proposed approximate 95% confidence intervals for ψ for each data set.

2.7.2 Simulation Results

We compared the coverage properties and median interval widths for the Wald, score, and AIL CIs via a Monte Carlo simulation. The simulation results for parameter configurations $C_{2.1}$ and $C_{2.2}$ are displayed in Figures 2.1 - 2.4. Figures 2.1 and 2.3 depict coverage probabilities, and Figures 2.2 and 2.4 display median interval widths for the sample-size scenarios $A1$ - $A8$ given in Table 2.4, for $\psi = 2, 4$, and for our low differential misclassification scenario.

Under parameter configuration $C_{2.1}$, we see in Figure 2.1 that the Wald interval coverage probability was conservative and overcovered the nominal 95% confidence

level for all considered sample-size scenarios, $A1$ - $A8$. Although the score CI is often believed to be an improvement over the Wald CI, it yielded the smallest coverage probability for most of the sample-size scenarios considered here. In Figure 2.1, the AIL CI showed better coverage properties of ψ than the Wald and the score CIs. The AIL CI slightly under-covered the OR ψ for the configurations $A1$ - $A8$, but as the sample sizes increased, the AIL CI coverage probabilities approached the nominal 95% confidence level.

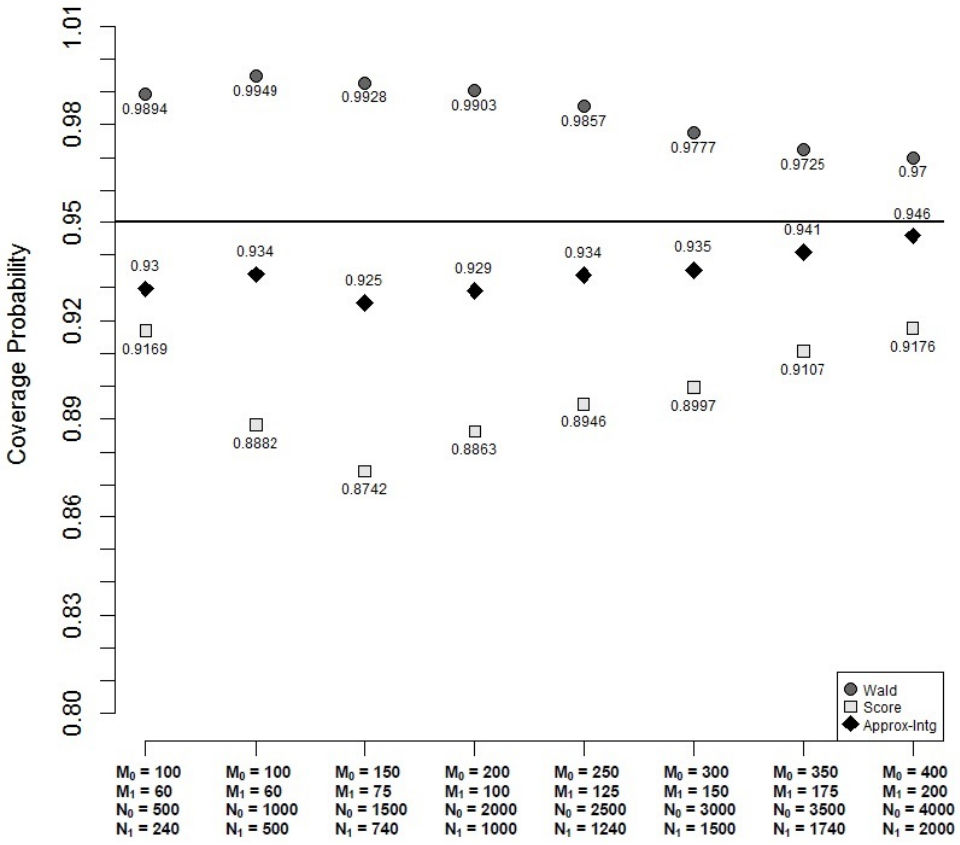

Figure 2.1: Coverage rates of W, S, and AIL CIs for a cohort study under $C_{2.1}$.

In Figure 2.2, we see that as the sample sizes increased, the median interval widths of the Wald CIs became consistently larger than the median interval widths of the AIL CIs. Hence, the actual parameter value for ψ is more likely to be contained in the Wald CIs, thus resulting in over-coverage for the Wald interval.

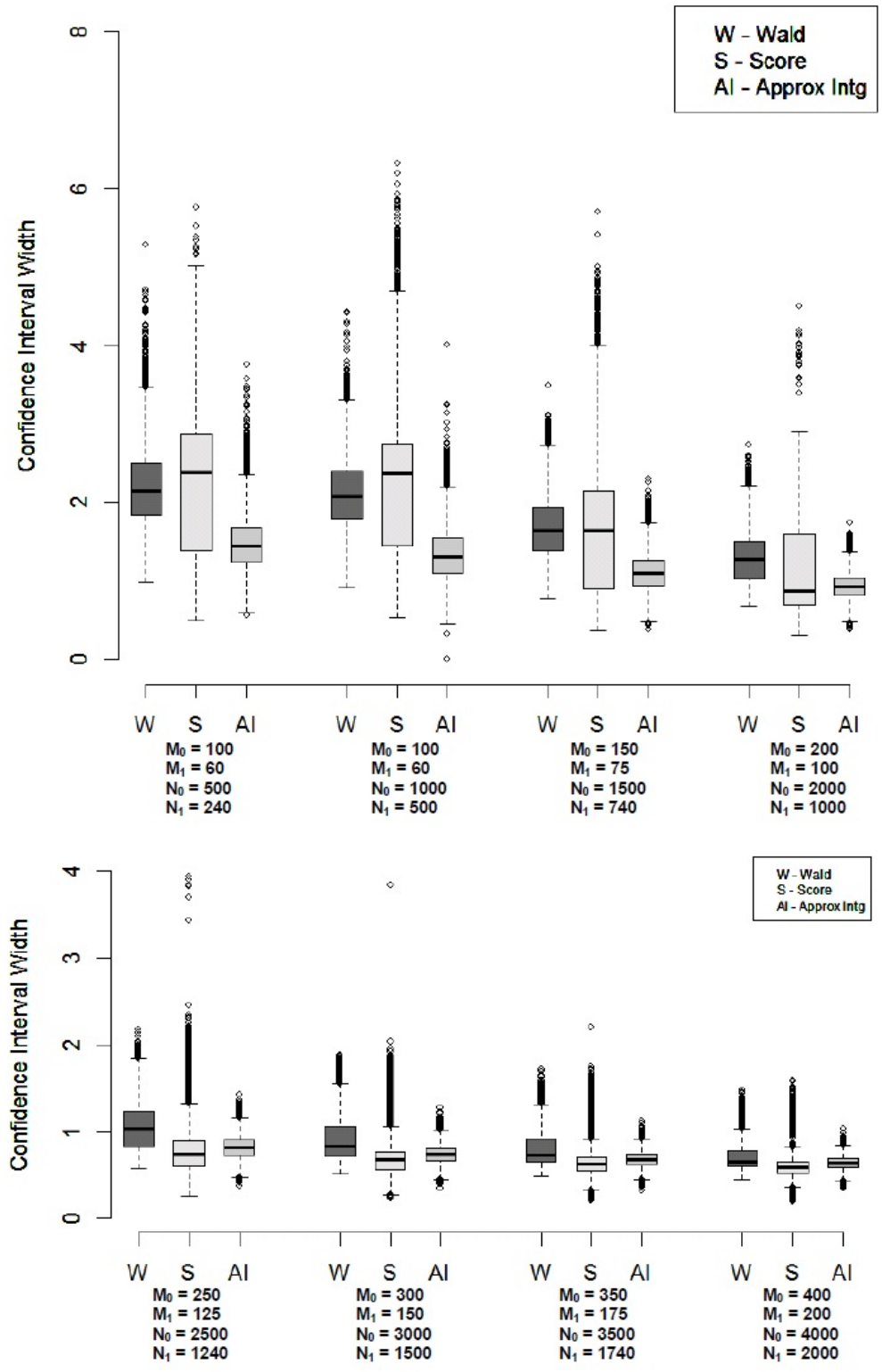

Figure 2.2: Interval widths of W, S, and AIL CIs for a cohort study for $C_{2.1}$.

The score CI yielded the greatest interquartile ranges for the sample-size scenarios $A1$ - $A8$ and the smallest coverage probabilities. The poor coverage properties for the score interval could be partially explained by the large variability in the score interval widths. The median interval widths of the score CIs for scenarios $A1$ - $A8$ were larger than those of the Wald and AIL CIs. The median AIL interval widths were much smaller than the Wald and score median interval widths and the validity in interval width was much less as well. Of the three CIs considered here, the AIL CI yielded the best overall coverage properties for the interval estimation of ψ. The simulation coverage probabilities and median interval width results obtained under parameter configuration $C_{2.2}$ are almost identical to the coverage and interval width properties for the case of low differential misclassification with $\psi = 2$.

We noticed in Figure 2.3 that the Wald CIs were very conservative and over-covered the nominal 95% confidence level for all of the sample-size scenarios in Table 2.4. The score CIs yielded the smallest average coverage probabilities and appeared to underestimate the nominal 95% confidence level for all considered sample-size scenarios. However, the AIL interval displayed better coverage properties of ψ than those of the Wald and the score CIs in that they displayed only a slight under-coverage. Figure 2.3 also showed that for the sample-size scenario A1, the AIL CI underestimated the coverage, but as the considered sample sizes increased, the AIL CIs approached the nominal 95% confidence level.

In Figure 2.4, we see that as the sample sizes increased, the Wald CIs had larger median widths compared to those of the other competing CIs and, as the sample sizes increased, the interval widths of all intervals decreased. The median interval widths of the score CIs for the considered sample-size scenarios $A1$ - $A8$ were considerably shorter than those of the Wald and AIL CI widths, thus perhaps explaining the under-coverage for the score interval displayed in Figure 2.3.

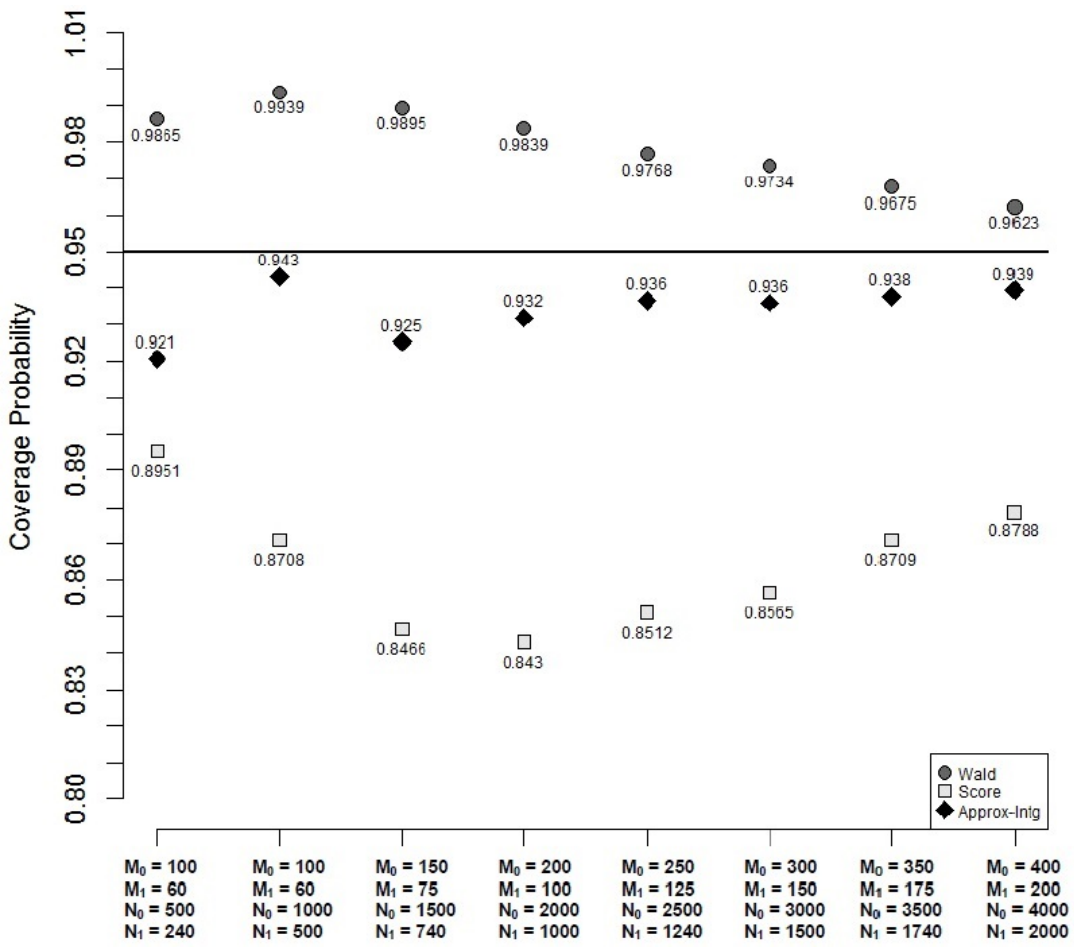

Figure 2.3: Coverage rates of W, S, and AIL CIs for a cohort study for $C_{2.2}$.

Of the three competing interval estimators, the AIL CI yielded the best coverage properties and displayed good interval-width properties for the interval estimation of ψ. Lastly, we compared the coverage probability properties and median interval widths for the Wald, score, and AIL interval estimators using the simulation results in Figures 2.5 - 2.8. Figures 2.5 and 2.7 depict estimated coverage probabilities and Figures 2.6 and 2.8 display the median interval widths across the different sample-size scenarios, $A1$ - $A8$, in Table 2.4 for $\psi = 2, 4$, under high differential misclassification.

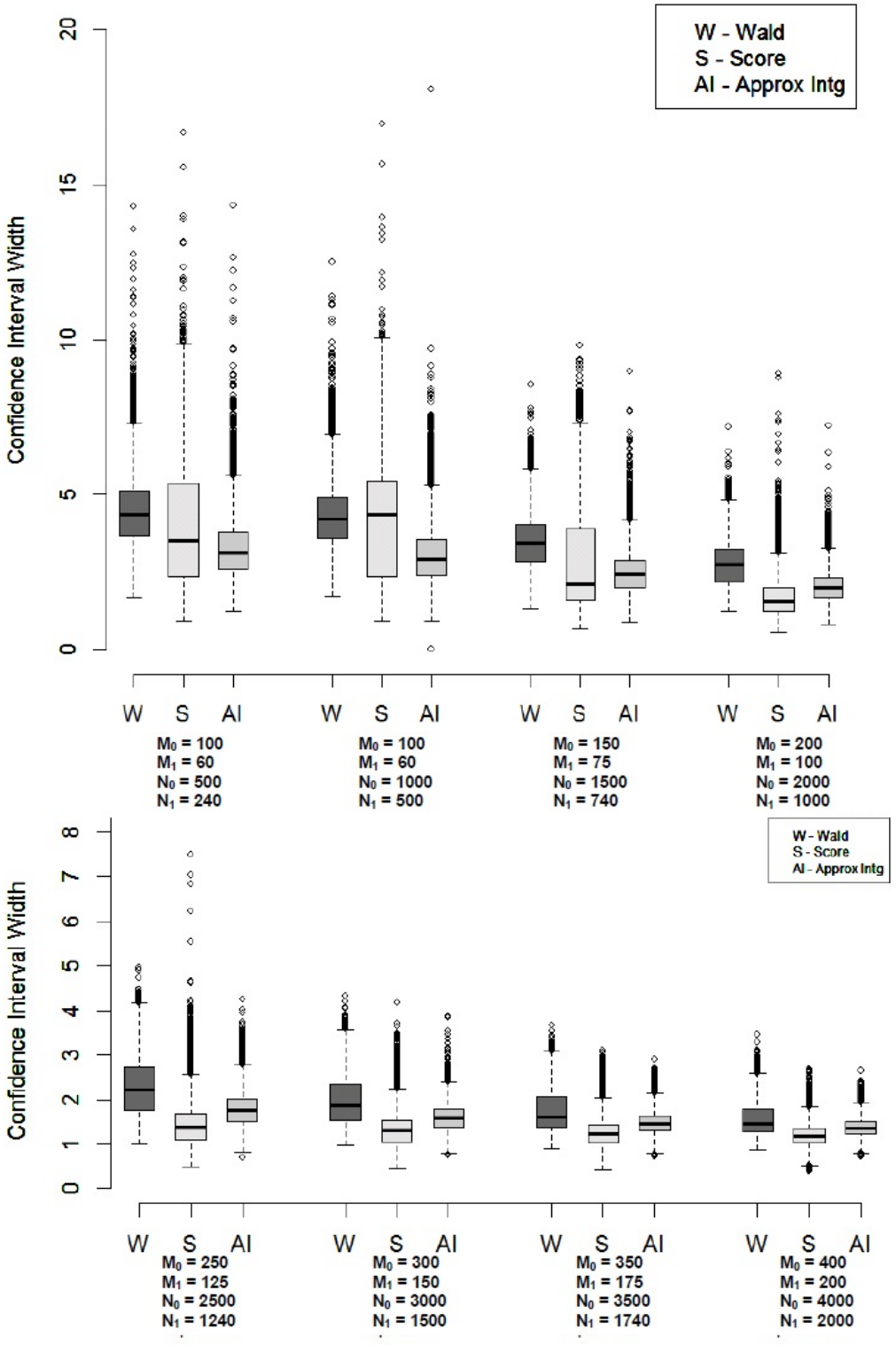

Figure 2.4: Interval widths of W, S, and AIL CIs for a cohort study for $C_{2.2}$.

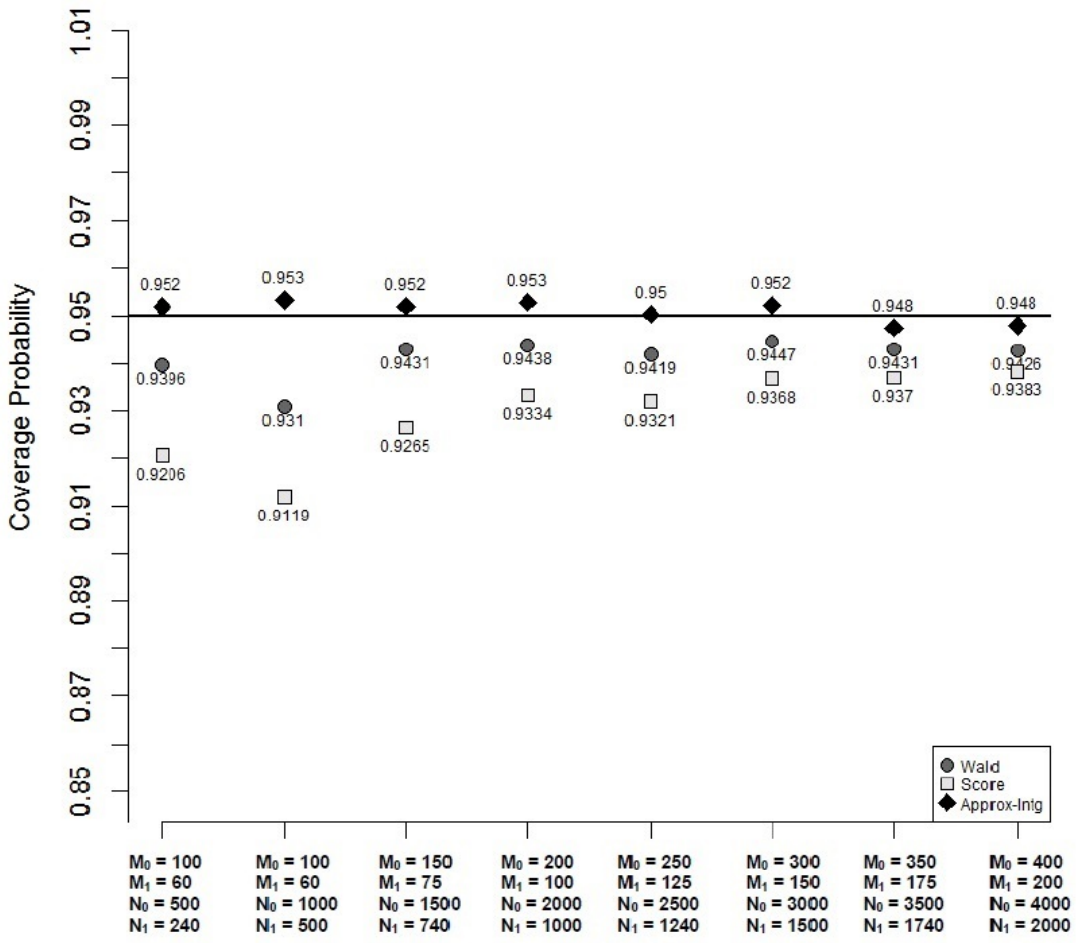

Figure 2.5: Coverage rates of W, S, and AIL CIs for a cohort study for $C_{2.3}$.

For the parameter configurations $C_{2.3}$ and $C_{2.4}$, we see in Figures 2.5 and 2.7, respectively, that although the Wald CI slightly under-covered for the nominal 95% confidence level for all considered sample-size scenarios, its coverage properties were good. However, the score interval yielded the lowest coverage probabilities. From Figures 2.5 and 2.7, we observed that the AIL CI yielded better coverage properties than the Wald and score CIs. Summarily, the AIL CI was an excellent CI for estimating ψ under parameter configurations $C_{2.3}$ and $C_{2.4}$.

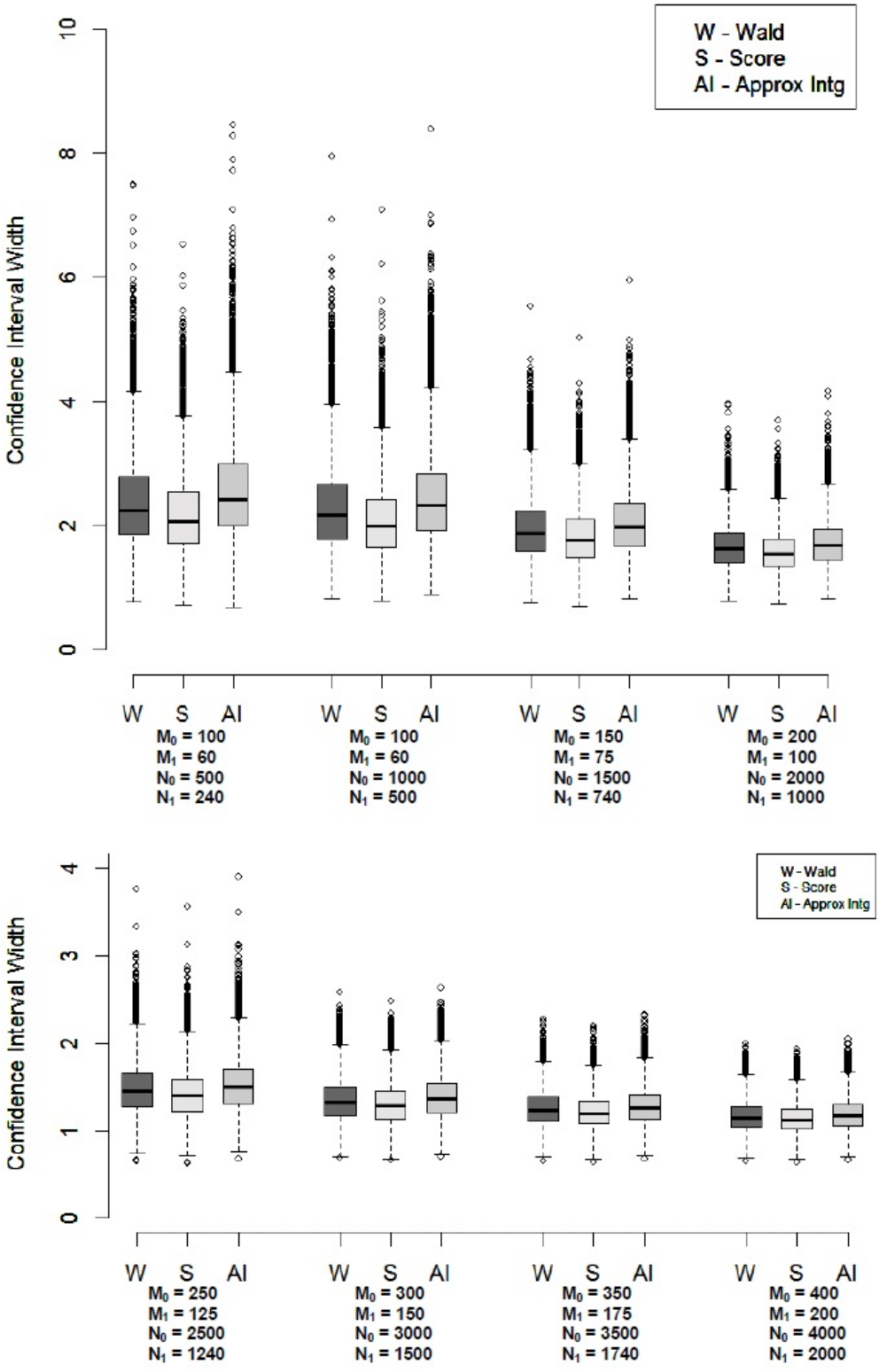

Figure 2.6: Interval widths of W, S, and AIL CIs for a cohort study for $C_{2.3}$.

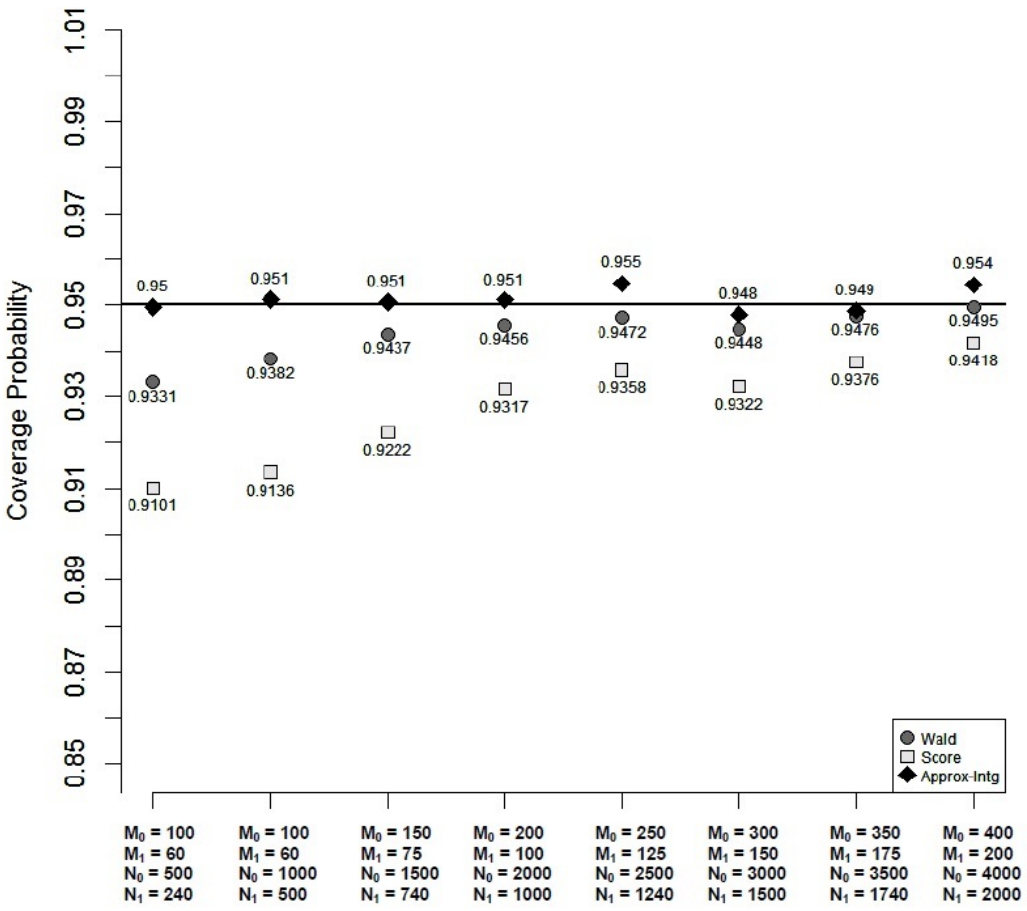

Figure 2.7: Coverage rates of W, S, and AIL CIs for a cohort study for $C_{2.4}$.

Figures 2.6 and 2.8 indicated that as the sample sizes increased, the median interval widths of the AIL interval were consistently larger than the corresponding Wald and score interval widths. Hence, it is highly probable that the actual parameter value is more frequently contained in the AIL CIs, therefore resulting in the best overall coverage properties confirmed in Figures 2.5 and 2.7. The median interval widths of the score interval were smaller than those of the competing CIs for all sample-size scenarios, A_1 - A_8, hence supporting the under-coverage seen in Figures 2.5 - 2.7. Overall, we have smaller variability for all CI widths in Figures 2.6 and 2.8. This fact could occur because we have more data to estimate the misclassification parameters S_i and $C_i, i = 0, 1$, under the high misclassification scenario.

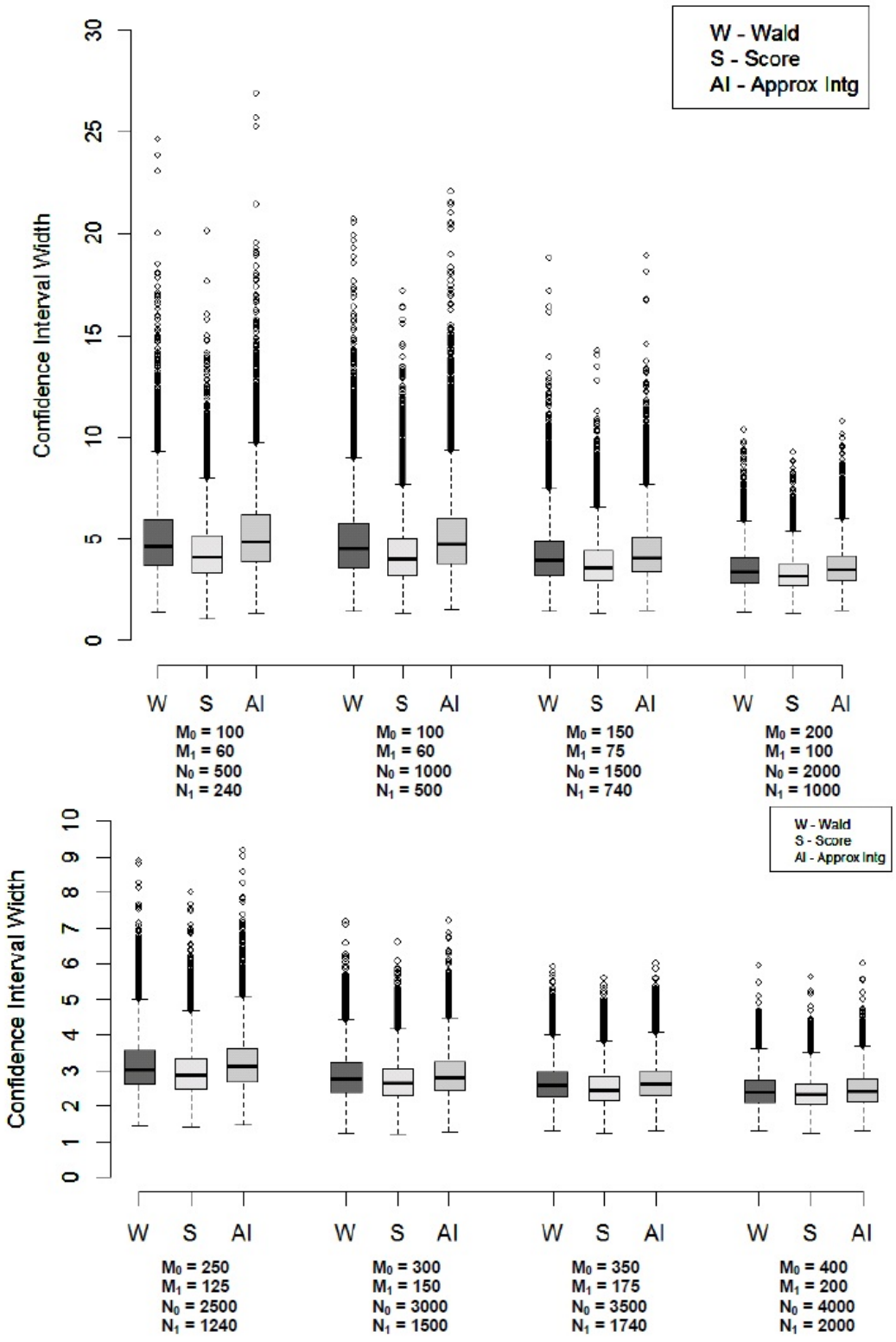

Figure 2.8: Interval widths of W, S, and AIL CIs for a cohort study for $C_{2.4}$.

2.8 Comments

In this paper, we have derived two likelihood-based and one pseudo-likelihood-based approximate CIs for the OR parameter for a cohort study using the double-sampling scheme for binary data subject to differential misclassification. Through our simulation results, we have concluded that the AIL interval yields coverage probabilities closest to the nominal confidence level regardless of the OR parameter values or the degree of differential misclassification. This result appears to be consistent with the findings of several authors who have examined the AIL pseudo-likelihood interval under different settings, (Boese (2005), Greer (2008), and Markova (2011)). The score CI performed the worst in our simulation study in terms of coverage probability and was also more computationally demanding than the Wald and AIL CIs.

CHAPTER THREE

Three First-Order Asymptotic Confidence Intervals for the Odds Ratio in a 2×2 Cohort Study with Non-Differential Misclassification

3.1 Abstract

We derive three first-order asymptotic confidence intervals for the odds ratio in 2×2 cohort studies using a double-sampling scheme under the assumption of non-differential misclassification. Specifically, we obtain Wald, profile likelihood (PL), and approximate integrated likelihood-based (AIL) confidence intervals (CIs). We examine average coverage probabilities and median interval width properties of the newly obtained CIs via a Monte Carlo simulation study. For a low degree of non-differential misclassification, the Wald CI performed the best among the competing intervals in terms of average coverage properties. However, for a high degree of non-differential misclassification, the AIL CI was superior to the Wald and PL CIs in terms of average probabilities coverage and median-interval-width properties for the parameter configurations considered here.

3.2 Introduction

A cohort study is a form of longitudinal study in which the population under investigation consists of individuals who are at risk of developing a specific disease or health outcome. These individuals are observed for a period of time in order to measure the frequency of occurrence of the disease among those exposed to the suspected causal agent as compared to those not exposed (Blumenthal, Fleisher, Esrey, and Peasey, 2001). Hence, the objective in all cohort studies is to partition the cohort into a group of exposed individuals and a group of non-exposed individuals and follow their disease status over time. A famous example of a cohort study is the Nurse's Health Study, in which cohorts of nurses were followed for over thirty years

to see how various factors, such as smoking, exercise, and hormone levels, affect their long-term health.

In most 2×2 cohort studies, a researcher uses the odds ratio (OR), which is the ratio of the probability of occurrence of an event to that of non-occurrence. The OR describes the strength of association between an exposure and a disease outcome. In most epidemiologically based studies, such as cohort studies, the OR may be poorly estimated due to non-differential misclassification of exposure. For example, for a binary exposure variable, some exposed subjects may be classified as non-exposed while some non-exposed subjects may be classified as exposed. Non-differential misclassification of exposure is present if, regardless of disease, all exposed and non-exposed subjects have the same probability of being misclassified (Sorahan and Gilthorpe, 1994).

Modern epidemiological techniques have been developed largely as a result of outbreak investigations of infectious disease during the nineteenth century (Blumenthal, Fleisher, Esrey, and Peasey, 2001). John Snow's study of cholera in London and its relationship to water supply is widely considered to be the first epidemiological study (Snow, 1851). A goal of many epidemiological studies is to assess the association between a binary exposure variable and the presence or absence of disease. A potential complication is that, for various reasons, the exposure variable may be misclassified (Prescott and Garthwaite, 2002). Data that includes misclassified data has initiated much research on comprehending the deleterious effects of misclassification on exposure-response relationships and developing procedures to rectify the misclassification-caused bias problems.

A considerable literature on the estimation of the OR under misclassification has accumulated. For instance, Walter and Irwig (1988) have examined methods to analyze categorical clinical and epidemiological data, in which the observations were subject to misclassification. They demonstrated that under certain conditions, it

is possible to estimate error parameters such as specificity, sensitivity, relative risk, or predictive value although no definitive classification (gold standard) is available. Thomas, Stram, and Dwyer (1993) have examined several proposed methods for adjusting exposure-response relationships for measurement error. Gustafson, Le, and Saskin (2001) have considered a case-control analysis with a dichotomous exposure variable that was subject to misclassification and showed that if the classification probabilities are known, then methods are available to adjust the OR for misclassified counts.

Several researchers have been concerned with point and interval estimation on functions of the population proportion obtained from misclassified data by using a double-sampling scheme. See Tenenbein (1970, 1972); Geng and Asano (1989); York, Madigan, Heuch, and Lie (1995); Moors, Van Der Genugten, and Strijbosch (2000); Barnett, Haworth, and Smith (2001); Boese, Young, and Stamey (2006); and Lee and Byun (2008).

In particular, Tenenbein (1970, 1972) has introduced a double-sampling scheme to estimate a population proportion parameter using misclassified binomial and multinomial data. York et al. (1995) have illustrated the advantage of a double-sampling scheme to estimate the proportion of infants born with Down's Syndrome, while Barnett et al. (2001) have presented a two-phase sampling scheme with applications to auditing. They have discussed methodologies for combining two data sets to produce optimum estimates of the proportion of financial transaction in error. Also, Boese et al. (2006) have approached the interval-estimation problem by deriving five asymptotic confidence intervals in the false-positive misclassification model. These intervals were based on certain combinations of pseudo-likelihood-based statistics, likelihood-based statistics, and differing Fisher-information types. Lee and Byun (2008) have provided a simple but effective interval estimator for the population proportion with double-sampled data subject to false-positive classifica-

tion relying on the Bayesian paradigm. In this paper we obtain one likelihood-based CI and two pseudo-likelihood-based CIs for the OR parameter in a cohort study using the double-sampling paradigm. Specifically, we derive and compare Wald, PL, and the AIL CIs to estimate the OR from a 2×2 cohort study.

We have organized the remainder of this paper as follows. We present the model, basic terms, and needed notation in Section 3.3. We examine the double-sampling scheme as well as the assumption of non-differential misclassification in Section 3.3. In Section 3.4 we derive maximum-likelihood-estimating equations for both the parameter of interest and the nuisance parameters. We use the Newton-Raphson method for determining the parameter MLEs for the model because we cannot determine closed-form solutions. We use the results of this section to develop a Wald CI for the OR. The other two intervals require the derivation of restricted maximum-likelihood estimators (RMLEs) for the nuisance parameters. We present an EM algorithm to determine the RMLEs in Section 3.5. Further, in Section 3.6 we obtain the observed Fisher information matrix (OFIM), which is used to derive the three CIs for the odds ratio: the Wald, PL, and AIL intervals. In Section 3.7 we describe a Monte Carlo simulation study comparing the coverage probabilities and median interval widths of each of the three intervals for two levels of misclassification–low and high. Finally, in Section 3.8 we give some brief concluding remarks.

3.3 A Double-Sampling Model

Recall that the cohort under investigation consists of individuals in a 2×2 cohort study who have been observed for a period of time and then tested for the disease of interest and exposure outcomes. The binary random variable D represents the disease level ($D = 1$ for diseased, $D = 0$ for not diseased) of each individual in the study. Researchers have found cohort studies advantageous for the study of

both relatively common outcomes and relatively rare exposures. However, because careful classification of exposures and outcomes is needed, as is the measurement and control of confounding factors, cohort studies are often complex and difficult to manage. The time span is usually at least a year and, consequently, cohort studies are expensive (Blumenthal et al. (2001)).

Employing the double-sampling scheme first proposed by Tenenbein (1970), we assume that there are two measuring devices used to classify participants as exposed or not exposed to a suspect causal agent. One is an inexpensive fallible device, and the other is an expensive infallible device. In stage one, the whole sample is classified by an inexpensive but fallible device, and in stage two, a smaller sub-sample is classified by a supplementary inerrant device usually referred to as a gold standard device. We therefore define two binary random variables: the gold standard exposure indicator X ($X = 1$ for exposed, $X = 0$ for not exposed) and the fallible exposure indicator Z ($Z = 1$ for exposed, $Z = 0$ for not exposed).

Let the subscripts k, i, and j represent the gold standard test outcome, $X = k$; the true disease status, $D = i$; and the fallible test outcome, $Z = j$. We denote the count for the incomplete data for the cell with $D = i$ and $Z = j$ by W_{ij}. We denote the cell count for the complete data for the cell with $X = k$, $D = i$ and $Z = j$ by V_{kij}, where $k, i, j = 0, 1$.

Also, we use the notation $M_k + N_i$ to represent the sample size from stage one and represent the sample size of stage two by M_k for $k, i = 0, 1$. The overall design of the cohort study with double sampling is shown in Table 3.1.

Table 3.1: Counts for a Study with Misclassified Exposure Data

	Validation Study (Complete)				Main Study (Incomplete)	
Fallible	$X = 1$	$X = 0$	$X = 1$	$X = 0$		
(Z)	$D = 1$	$D = 1$	$D = 0$	$D = 0$	$D = 1$	$D = 0$
$Z = 1$	V_{111}	V_{011}	V_{101}	V_{001}	W_{11}	W_{01}
$Z = 0$	V_{110}	V_{010}	V_{100}	V_{000}	W_{10}	W_{00}
	M_1		M_0		N_1	N_0

For the complete (validation, infallible) study, we denote the joint probability of exposure and the disease-level categories; that is, the probability for $X = 1$ for the i^{th} group is

$$\pi_i = Pr(X = 1, D = i), \tag{3.1}$$

where $i = 0, 1$, for not diseased and diseased, respectively. After introducing the fallible test, we define the specificity C_i as the probability that an individual does not have the disease according to the fallible test ($Z = 0$) and the individual has a negative result on the gold standard procedure (X=0) for each of the disease-level categories $D = 0, 1$. We also define the sensitivity, or true positive rate, S_i, as the probability that an individual tests positive under the fallible test ($Z = 1$) and the individual has a positive result on the gold standard procedure (X=1) for each of the disease-level categories $D = 0, 1$.

Under the assumption of non-differential misclassification, we know that the exposure status is independent of the disease outcome level, and we redefine the specificity and sensitivity probabilities, respectfully, as

$$S = Pr(Z = 1|X = 1, D = i), \quad i = 0, 1 \tag{3.2}$$

and

$$C = Pr(Z = 0|X = 0, D = i), \quad i = 0, 1. \tag{3.3}$$

Based on (3.1) - (3.3) and derivations in Prescott and Garthwaite (2002), we utilize the following multinomial distributions on the observable counts for the complete study. We assume

$$(V_{111}, V_{110}, V_{011}, V_{010}) \sim Multi(\pi_1 S, \pi_1(1 - S), (1 - \pi_1)(1 - C), (1 - \pi_1)C), \tag{3.4}$$

and

$$(V_{101}, V_{100}, V_{001}, V_{000}) \sim Multi(\pi_0 S, \pi_0(1 - S), (1 - \pi_0)(1 - C), (1 - \pi_0)C), \tag{3.5}$$

where $i = 0, 1$, indicates the actual disease status of the person in the study (0 for diseased, 1 for non-diseased). For the incomplete study, we assume a binomial distribution for the observable cell counts so that

$$W_{i1} \sim Bin(N_i, \pi_i S + (1 - \pi_i)(1 - C)),$$

where $i = 0, 1$. The OR, which associates the gold standard exposure level X to the disease outcome D, is

$$\psi \equiv \frac{\pi_1(1 - \pi_0)}{\pi_0(1 - \pi_1)}. \tag{3.6}$$

Many documented 2×2 cohort studies with misclassified data have appeared in the literature. For example, based on the work of Tenenbein (1970), York et al. (1995) have illustrated the advantage of the double-sampling scheme in a cohort study estimating the proportion of infants born with Down's Syndrome nationwide. From 1979 - 1984, for every birth in Norway, the midwife or obstetrician classified each child with Down's Syndrome based on a visual inspection (Z), and for a small sub-sample of births, an expensive but accurate cytogenetic test (X) was applied for the classification. Dahm, Gail, Rosenberg, and Pee (1995) gave another example of a prospective cohort study of the prognostic value of a renal biopsy (X) on 5-year survival ($D = 1$ if death within five years). Renal biopsy studies were performed on 100 or fewer patients because of the possible complications and discomfort, and instead of using a renal biopsy, researchers used a non-invasive test of renal function (Z) on the whole cohort.

3.4 Maximum-Likelihood Estimators

Let $\boldsymbol{\theta} \equiv (\psi, \boldsymbol{\eta}')'$ be the model parameter vector, where $\boldsymbol{\eta} \equiv (\pi_1, C, S)'$ is the vector of nuisance parameters. To derive the Wald, PL, and AIL CIs for ψ, we must obtain the MLE of ψ and MLEs of the nuisance parameters π_1, C, and S. Let $\boldsymbol{d} \equiv (W_{0i}, W_{1i}, V_{1i1}, V_{0i1}, V_{1i0}, V_{0i0})'$, where $i = 0, 1$, indicates the actual disease status (0 for diseased, 1 for non-diseased) for the observed data counts from both the main

(fallible) and the validation (infallible) studies. If we let $L_o \equiv L(\pi_1, \pi_0, C, S|\boldsymbol{d})$ be the observed likelihood function then

$$\begin{aligned}
L_o \propto\ & [\pi_1 S + (1-\pi_1)(1-C)]^{W_{11}}[1-\pi_1 S-(1-\pi_1)(1-C)]^{N_1-W_{11}} \\
& \times (\pi_1 S)^{V_{111}}[\pi_1(1-S)]^{V_{110}}[(1-\pi_1)(1-C)]^{V_{011}} \times [(1-\pi_1)C]^{[M_1-(V_{111}+V_{110}+V_{011})]}, \\
& \times [\pi_0 S + (1-\pi_0)(1-C)]^{W_{01}}[1-\pi_0 S-(1-\pi_0)(1-C)]^{N_0-W_{01}} \\
& \times (\pi_0 S)^{V_{101}}[\pi_0(1-S)]^{V_{100}}[(1-\pi_0)(1-C)]^{V_{001}} \times [(1-\pi_0)C]^{[M_0-(V_{101}+V_{100}+V_{001})]}.
\end{aligned}$$

Using (3.6), we next make the transformation

$$\pi_0 = \frac{\pi_1}{\pi_1 - \pi_1\psi + \psi}. \tag{3.7}$$

Now substituting the right-hand side of (3.7) for π_0 into the log-likelihood function, we get

$$\begin{aligned}
\ell_\psi \propto\ & W_{11}\ln[\pi_1 S + (1-\pi_1)(1-C)] + W_{01}\ln\left[\frac{(1-C)(1-\pi_1)\psi + \pi_1 S}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + (N_1 - W_{11})\ln[1-\pi_1 S-(1-\pi_1)(1-C)] + V_{101}\ln\left[\frac{\pi_1 S}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + (N_0 - W_{01})\ln\left[\frac{C\psi + \pi_1 - C\psi\pi_1 - \pi_1 S}{\psi + \pi_1 - \psi\pi_1}\right] + V_{100}\ln\left[\frac{(\pi_1 - \pi_1 S)}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + V_{111}\ln(\pi_1 S) + V_{110}\ln[\pi_1(1-S)] + V_{011}\ln[(1-\pi_1)(1-C)] \\
& + [M_1 - (V_{111} + V_{110} + V_{011})]\ln[(1-\pi_1)C] + V_{001}\ln\left[\frac{(1-C)(1-\pi_1)\psi}{\psi + \pi_1 - \psi\pi_1}\right] \\
& + [M_0 - (V_{101} + V_{100} + V_{001})]\ln\left[\frac{C\psi - \pi_1\psi C}{\psi + \pi_1 - \psi\pi_1}\right],
\end{aligned} \tag{3.8}$$

where $\boldsymbol{\eta}$ is the nuisance parameter vector and ℓ_ψ represents the log-likelihood in terms of ψ and $\boldsymbol{\eta}$, given the observed data counts $\boldsymbol{d}$. We remark that using a different parametrization, Karunaratne (1991) has observed that one cannot derive closed-form MLEs for any parameter in (3.8). Thus, we use a numerical method for determining the MLEs of ψ and $\boldsymbol{\eta}$.

3.5 Restricted Maximum-Likelihood Estimators

The PL and AIL CIs require evaluation not only at the MLEs, but also at the restricted MLEs (RMLEs) of the nuisance parameters. Hence, we next describe a method to determine the RMLE for $\boldsymbol{\eta}$ for a fixed value of ψ.

We assume that the cell counts from the validation sample (refer to Table 3.1) are subject to misclassification and define a new set of latent variables, U_{1ij}, as the unobserved misclassified counts. Recall that the subscripts, $i, j = 0, 1$, correspond to the outcomes of the gold standard test (X), the true disease condition (D), and the fallible test (Z), respectively, for each patient in the study. In Table 3.2 we display the latent variables across the main, or fallible, study. Hence, the counts U_{1ij} are the unobserved portions of the observed counts W_{ij}. From previous derivations and

Table 3.2: Counts for the Main Study with Unobserved, Misclassified Data

	Main Study (Incomplete)			
Fallible	$X = 1$	$X = 0$	$X = 1$	$X = 0$
(Z)	$D = 1$	$D = 1$	$D = 0$	$D = 0$
$Z = 1$	U_{111}	$W_{11} - U_{111}$	U_{101}	$W_{01} - U_{101}$
$Z = 0$	U_{110}	$W_{10} - U_{110}$	U_{100}	$W_{00} - U_{100}$
	$U_{111} + U_{110}$	$N_1 - (U_{111} + U_{110})$	$U_{101} + U_{100}$	$N_0 - (U_{101} + U_{100})$
	N_1		N_0	

the assumption that the main study is a sub-sample of the complete study including both the fallible and infallible data, we have that the distributions of the unobserved latent variables are

$$(U_{1i1} + U_{1i0}) \sim Bin(N_i, \pi_i),$$

$$U_{1i1}|(U_{1i1} + U_{1i0}) \sim Bin(U_{1i1} + U_{1i0}, S),$$

and

$$(W_{i0} - U_{1i0})|(U_{1i1} + U_{1i0}) \sim Bin(N_i - (U_{1i1} + U_{1i0}), C),$$

where $i = 0, 1$, indicates the non-diseased and the diseased categories, respectively. One should refer to Joseph, Gyorkos, and Coupal (1995) for more details. Let $\boldsymbol{d}^{full}$

represent the full-data vector including the unobserved latent variables. The full-data likelihood function $L_U \equiv L_U(\pi_1, \pi_0, C, S|\boldsymbol{d}^{full})$ is displayed in expression (B.1.1) in Appendix B.1. Because we cannot derive closed-form solutions for the RMLE for $\boldsymbol{\eta}$, we construct an EM algorithm to determine the RMLE for $\boldsymbol{\eta}$ given ψ. We outline the EM algorithm steps in detail in Appendix B.1.

3.6 Three First-Order Asymptotic Confidence Intervals for ψ

In this section, we derive three particular first-order asymptotic confidence intervals for the OR parameter ψ defined in (3.6). First, however, we give the OFIM whose inverse is used to obtain an estimate of the variance of $\hat{\psi}$, the MLE of ψ.

Recall that $\boldsymbol{\eta} = (\pi_1, S, C)'$ is the nuisance parameter vector where S and C represent the common sensitivity and the common specificity, respectively. Then

$$\boldsymbol{J}(\psi, \boldsymbol{\eta}) \equiv - \begin{bmatrix} \frac{\partial^2 \ell_\psi}{\partial \psi^2} & \frac{\partial^2 \ell_\psi}{\partial \psi \partial \pi_1} & \frac{\partial^2 \ell_\psi}{\partial \psi \partial S} & \frac{\partial^2 \ell_\psi}{\partial \psi \partial C} \\ \cdot & \frac{\partial^2 \ell_\psi}{\partial \pi_1^2} & \frac{\partial^2 \ell_\psi}{\partial \pi_1 \partial S} & \frac{\partial^2 \ell_\psi}{\partial \pi_1 \partial C} \\ \cdot & \cdot & \frac{\partial^2 \ell_\psi}{\partial S^2} & \frac{\partial^2 \ell_\psi}{\partial S \partial C} \\ \cdot & \cdot & \cdot & \frac{\partial^2 \ell_\psi}{\partial C^2} \end{bmatrix} \tag{3.9}$$

is the OFIM, where ℓ_ψ is the transformed log-likelihood function. We display the elements of (3.9) in Appendix B.2. We use (3.9) in the following subsection.

3.6.1 A Wald Confidence Interval for ψ

We first describe the full-likelihood Wald CI that incorporates nuisance parameters. The Wald statistic for ψ with nuisance parameter vector $\boldsymbol{\eta}$ is

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

where

$$\boldsymbol{J}(\psi, \boldsymbol{\eta}) \equiv \begin{bmatrix} J_{11} & \boldsymbol{J}_{12} \\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22} \end{bmatrix} \tag{3.10}$$

is the partitioned observed information matrix for the MLEs $\hat{\psi}$ and $\hat{\boldsymbol{\eta}}$ and $J^{11} \equiv (J_{11} - \boldsymbol{J}_{12}\boldsymbol{J}_{22}^{-1}\boldsymbol{J}_{21})^{-1}$ (see Pawitan (2001)).Thus, an approximate $100(1-\alpha)\%$ CI for the OR consists of the values of ψ that satisfy

$$(\hat{\psi} - \psi)^2[J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]^{-1} < \chi_1^2(1-\alpha), \tag{3.11}$$

where $\chi_1^2(1-\alpha)$ denotes the $(1-\alpha)$ quantile of a central chi-squared distribution with one degree of freedom. Solving (3.11) directly for ψ, we obtain

$$\hat{\psi} - \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]} < \psi < \hat{\psi} + \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\psi}, \hat{\boldsymbol{\eta}})]}, \tag{3.12}$$

an approximate $(1-\alpha)\%$ Wald confidence interval for ψ under the double-sampling procedure with non-differential misclassification. We use a bisectional algorithm to derive the endpoints of (3.12). Notice that $\boldsymbol{J}^{11}$ is evaluated at the MLEs for both ψ and $\boldsymbol{\eta}$ derived in Section 3.4.

3.6.2 A Profile Likelihood Confidence Interval for ψ

The second CI we consider in this paper is the profile likelihood interval, which is a pseudo-likelihood-based CI, generally believed to have better coverage probability properties than the full likelihood Wald CI. We eliminate the nuisance parameters in the likelihood function by replacing them with their respective RMLEs. Hence, the profile likelihood function is

$$L_P(\psi) \equiv \max_{\boldsymbol{\eta}} L(\psi, \boldsymbol{\eta}) = L(\psi, \hat{\boldsymbol{\eta}}_\psi), \tag{3.13}$$

where $\hat{\boldsymbol{\eta}}_\psi$ is the vector of profile MLEs or RMLEs for $\boldsymbol{\eta}$ in terms of ψ. We derive an EM algorithm to compute the RMLEs for $\boldsymbol{\eta}$ because we cannot obtain closed-form RMLEs (see Section 3.5). Thus, an approximate $100(1-\alpha)\%$ profile likelihood CI for the odds ratio is the set of values of ψ that satisfy

$$-2\left[\ell(\psi, \hat{\boldsymbol{\eta}}_\psi) - \ell(\hat{\psi}, \hat{\boldsymbol{\eta}})\right] < \chi_1^2(1-\alpha), \tag{3.14}$$

where $\ell(\hat{\psi}, \hat{\boldsymbol{\eta}})$ is the log-likelihood function evaluated at the MLEs $\hat{\psi}$ and $\hat{\boldsymbol{\eta}}$ (see Section 3.4), and $\ell(\psi, \hat{\boldsymbol{\eta}}_\psi)$ is the log-likelihood function evaluated at the RMLE $\hat{\boldsymbol{\eta}}$. The profile likelihood CI possesses a major flaw in that it does not account for the uncertainty in the nuisance parameter estimation. This flaw can cause optimistically short CIs for the OR parameter (see Riggs (2006)).

3.6.3 An Approximate Integrated Likelihood Confidence Interval for ψ

Finally, we describe the AIL CI, which is a pseudo-likelihood-based interval. This CI is generally considered an improvement over the PL CI, but its calculation is often computationally intensive or even intractable because one must integrate over all the nuisance parameters. To overcome this particular issue, we use the Laplace approximation method for a more computationally feasible approximation. The AIL likelihood function is

$$L_{AI}(\psi) = \int L(\psi, \boldsymbol{\eta}) d\boldsymbol{\eta} \approx \frac{c L_P(\psi)}{\left| \hat{\boldsymbol{J}}_{\boldsymbol{\eta}}(\psi, \hat{\boldsymbol{\eta}}) \right|^{1/2}}, \tag{3.15}$$

where $c = (2\pi)^{\nu/2}$, ν is the dimension of the nuisance parameter, and $L_P(\psi)$ is the profile likelihood expressed in (3.13). Recall that $\boldsymbol{J}_{\boldsymbol{\eta}}$ represents the nuisance parameter sub-matrix of the OFIM. Hence, from (3.10) we have $\hat{\boldsymbol{J}}_{\boldsymbol{\eta}}$ is $\boldsymbol{J}_{22}$ evaluated at the MLE of $\boldsymbol{\eta}$. The sub-matrix is obtained using the expressions for the OFIM given in Appendix B.2 and the MLE of $\boldsymbol{\eta}$. We use an iterative bisectional method to determine all values of ψ that satisfy

$$-2 \left[\ell_{AI}(\psi) - \ell_{AI}(\hat{\psi}_{AI}) \right] < \chi^2_1(1-\alpha), \tag{3.16}$$

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

3.7 The Monte Carlo Simulation Design

In this section, we examine the performance of three different first-order asymptotic CIs for the estimation of ψ in a 2×2 cohort study when implementing a

double-sampling procedure for a non-differential misclassification. In particular, we compare the performance of these three CIs based on coverage probability and median-interval-width properties via a Monte Carlo simulation. Specifically, we assess the effect of the sample sizes of both cohorts, M_i and N_i, respectively, for $i = 0$, 1, the effect of the disease probability π_k, for $k = 0, 1$, for the gold standard test outcome $X = 0, 1$, and the magnitude of the odds ratio ψ on the coverage probability and median interval width of the CIs (3.12), (3.14), and (3.16).

3.7.1 Simulation Parameter and Sample Size Configurations

In this Monte Carlo simulation study, we considered two values for the odds ratio parameter $\psi = 2$ and $\psi = 4$. Table 3.3 displays the joint probabilities of exposure and disease status $\pi_0 \equiv Pr(X = 1, D = 0)$ and $\pi_1 \equiv Pr(X = 1, D = 1)$. Here $X = 1$ indicates that the gold standard test has classified an individual as exposed, $D = 0$ indicates that the participant's condition is not-diseased, and $D = 1$ indicates that an individual is diseased.

Recall that we are interested only in the case of non-differential misclassification,

Table 3.3: Parameter Configurations Used in the Simulation for CONDIFF

Conf.	ψ	π_0	π_1	Misclassification	C	S
$C_{3.1}$	2	0.25	0.40	low	0.99	0.98
$C_{3.2}$	4	0.38	0.71	low	0.99	0.98
$C_{3.3}$	2	0.25	0.40	high	0.85	0.75
$C_{3.4}$	4	0.38	0.71	high	0.85	0.75

i.e, $S_0 = S_1 = S$ and $C_0 = C_1 = C$. We consider only two misclassification categories – low and high – where the corresponding the configuration parameter values are given in Table 3.3.

To obtain realistic interval comparisons, we derive CIs for eight sample sizes with low and high misclassification values with $\psi = 2$, 4. By definition we know that the validation sample sizes are much smaller than the fallible sample sizes due to

factors previously discussed. We show the sample-size values we use in our simulation in Table 3.4. The parameter configurations for this simulation are summarized

Table 3.4: Sample Sizes Used in the Simulation for CONDIFF

	A1	A2	A3	A4	A5	A6	A7	A8
M_0	100	100	150	200	250	300	350	400
M_1	60	60	74	100	124	150	174	200
N_0	500	1000	1500	2000	2500	3000	3500	4000
N_1	240	500	740	1000	1240	1500	1740	2000

by $\psi \in \{2, 4\}$, $\pi_1 \in \{0.40, 0.71\}$, $S \in \{0.98, 0.75\}$, and $C \in \{0.99, 0.85\}$. We generated 10,000 data sets under each combination of conditions and calculated the corresponding Wald, PL, and AIL 95% CIs for the OR under double-sampling and non-differential misclassification in cohort studies for each of the CI methods described in this chapter.

3.7.2 Monte Carlo Simulation Results

The parameter configuration scenario we first consider is the case of low non-differential misclassification and $\psi = 2, 4$. Refer to Table 3.3 to recall the appropriate misclassification probabilities. We compared the coverage probabilities (Figures 3.1 and 3.3) and median interval widths (Figures 3.2 and 3.4) for the three proposed CIs for the sample-size scenarios in Table 3.4.

Under parameter configuration $C_{3.1}$, we see in Figure 3.1 that the Wald CIs overestimated the nominal 95% confidence level for the sample-size scenarios A1 to A4. However, as the sample sizes increased, the Wald CI coverage probabilities converged towards the nominal 95% confidence level. Notice in Figure 3.2 that the Wald CIs had greater median interval widths and interquartile ranges for the sample-size scenarios A1 to A4 compared to the PL and AIL CIs. This fact could explain the over-coverage of the Wald CI seen in Figure 3.1. Nevertheless, as the sample sizes (A5 - A8) increased the median interval widths of the Wald CIs became similar

to the PL and AIL CIs. The PL and AIL CIs closely followed the same average coverage probability patterns and similar interval width characteristics (see Figure 3.1 and 3.2).

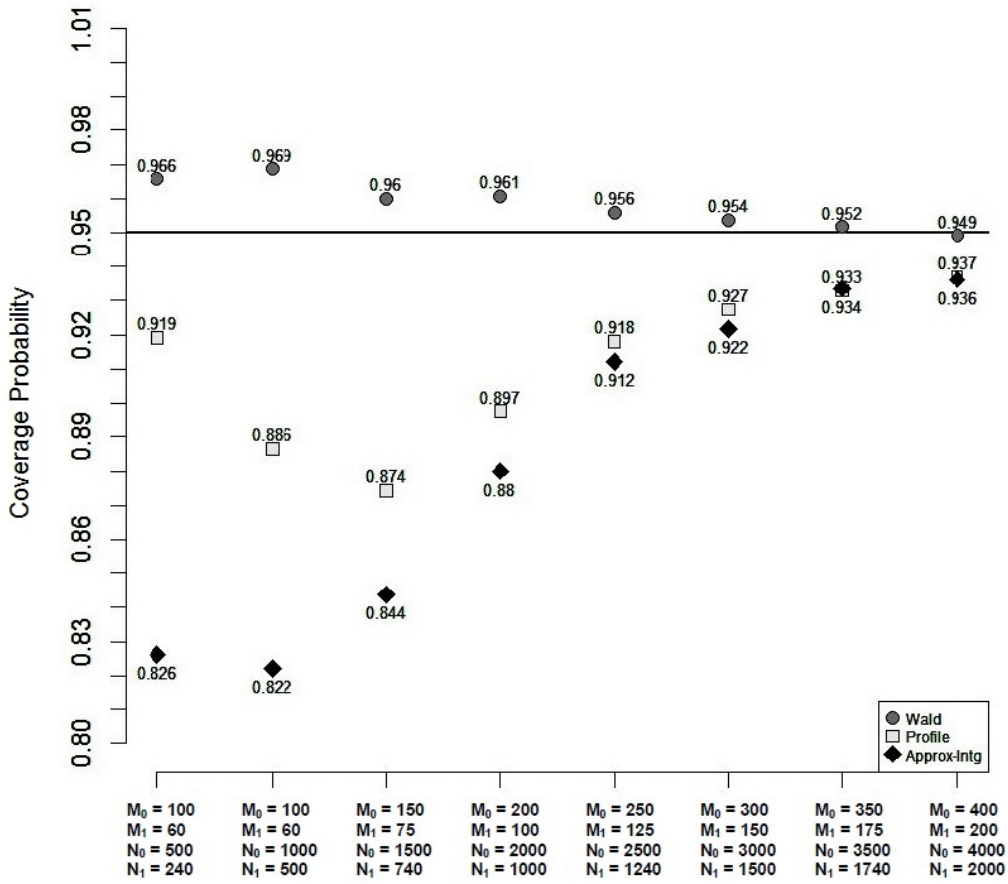

Figure 3.1: Coverage rates of W, PL, and AIL CIs for cohort study under $C_{3.1}$.

For the sample-size scenarios A1 - A4, the PL and AIL coverage probabilities were very poor in that the under-coverage was large for relatively small sample sizes. However, as the sample sizes increased, the coverage probabilities approached the nominal level but still did not achieve the desired nominal coverage. Overall, the AIL and PL CIs considerably under-covered ψ for all sample-size scenarios except the last two scenarios shown in Figure 3.2. Because the AIL and PL median interval widths were relatively short, the corresponding CIs had lower coverage probabilities. The

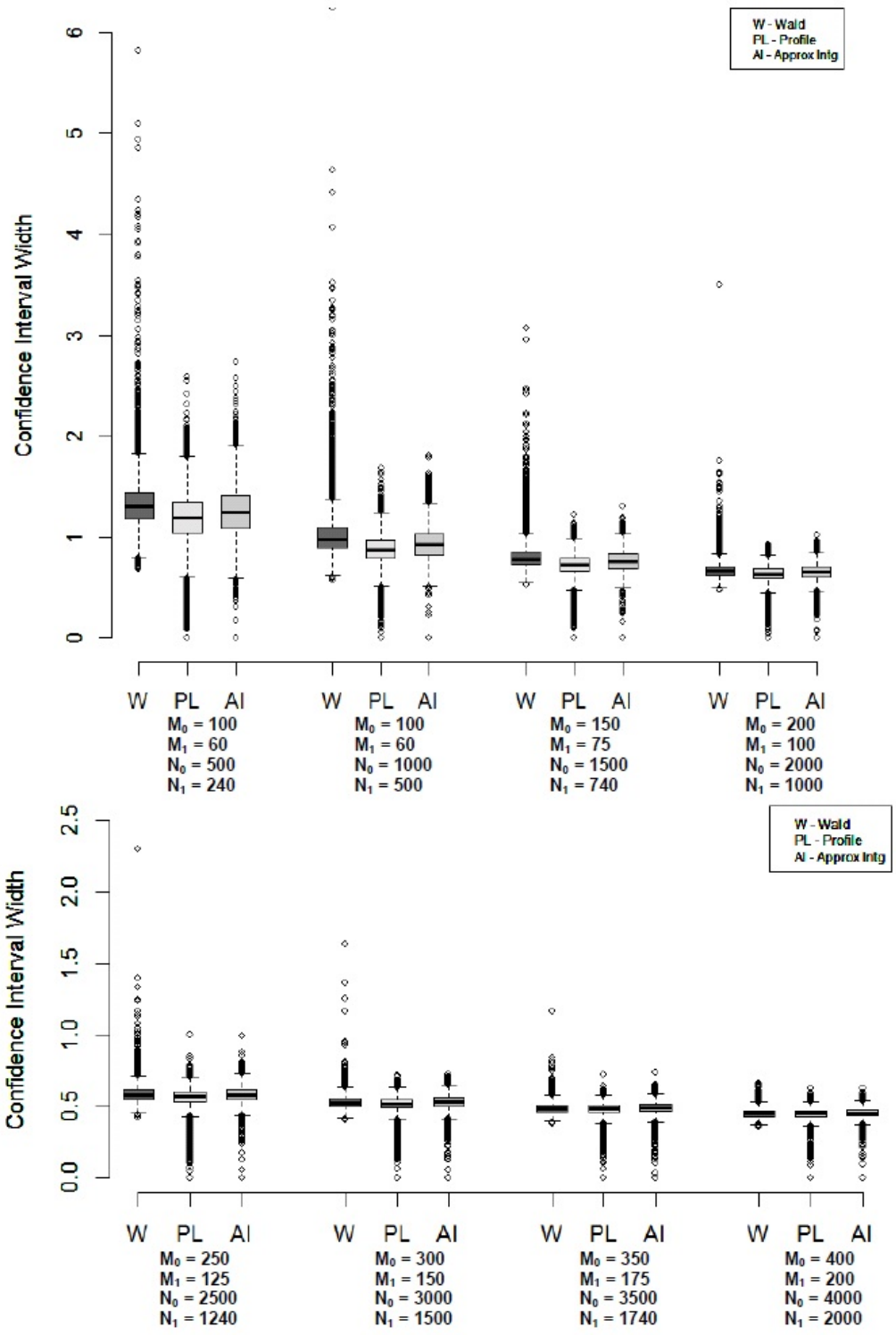

Figure 3.2: Interval widths of W, PL, and AIL CIs for cohort study under $C_{3.1}$.

results obtained for low non-differential misclassification and an odds ratio parameter $\psi = 4$ were similar to those obtained under low non-differential misclassification and an odds ratio parameter, $\psi = 2$.

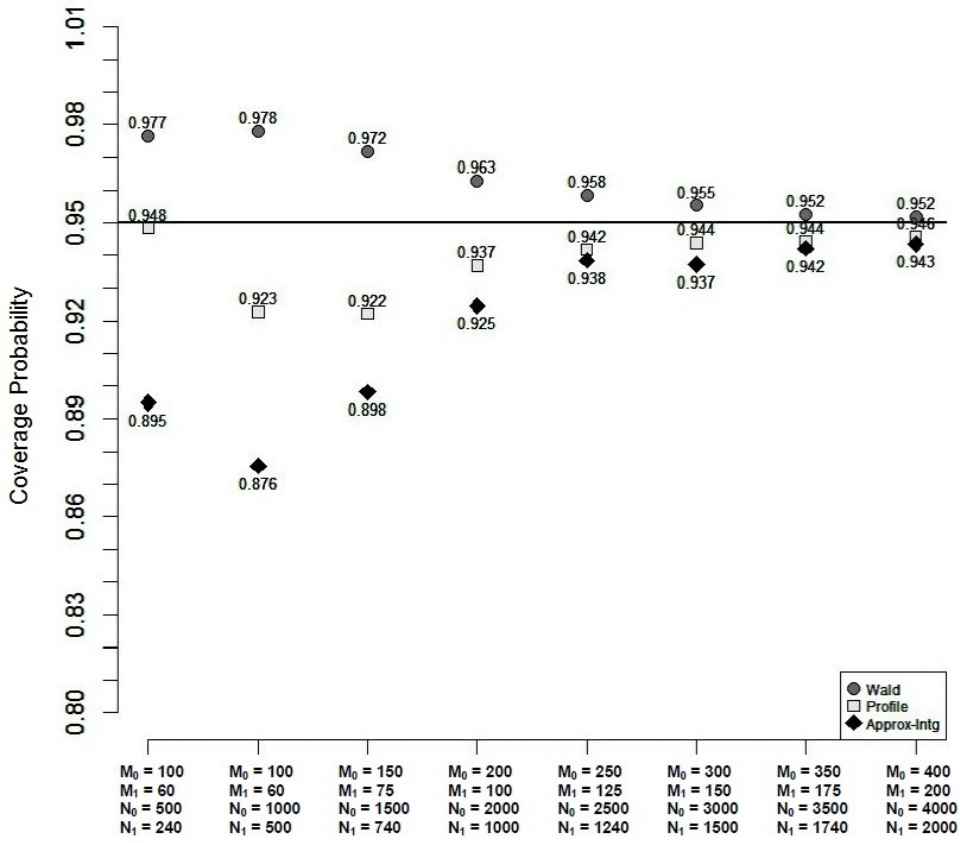

Figure 3.3: Coverage rates of W, PL, and AIL CIs for cohort study under $C_{3.2}$.

We noticed in Figure 3.3 that the Wald CI coverage probabilities for the sample-size scenarios $A1 - A4$ (Table 2.4) were conservative and overestimated the nominal 95% confidence level. However, as the sample sizes increased, the coverage probabilities improved and closely approximated the nominal confidence level. For the sample-size scenarios A1 - A4 in Figure 3.4, the Wald CIs had considerably larger median interval widths compared to the PL and AIL median interval widths. This fact could explain the over-coverage of the Wald CI seen in Figure 3.3. The PL and AIL CIs closely followed the same patterns for the case $\psi = 2$.

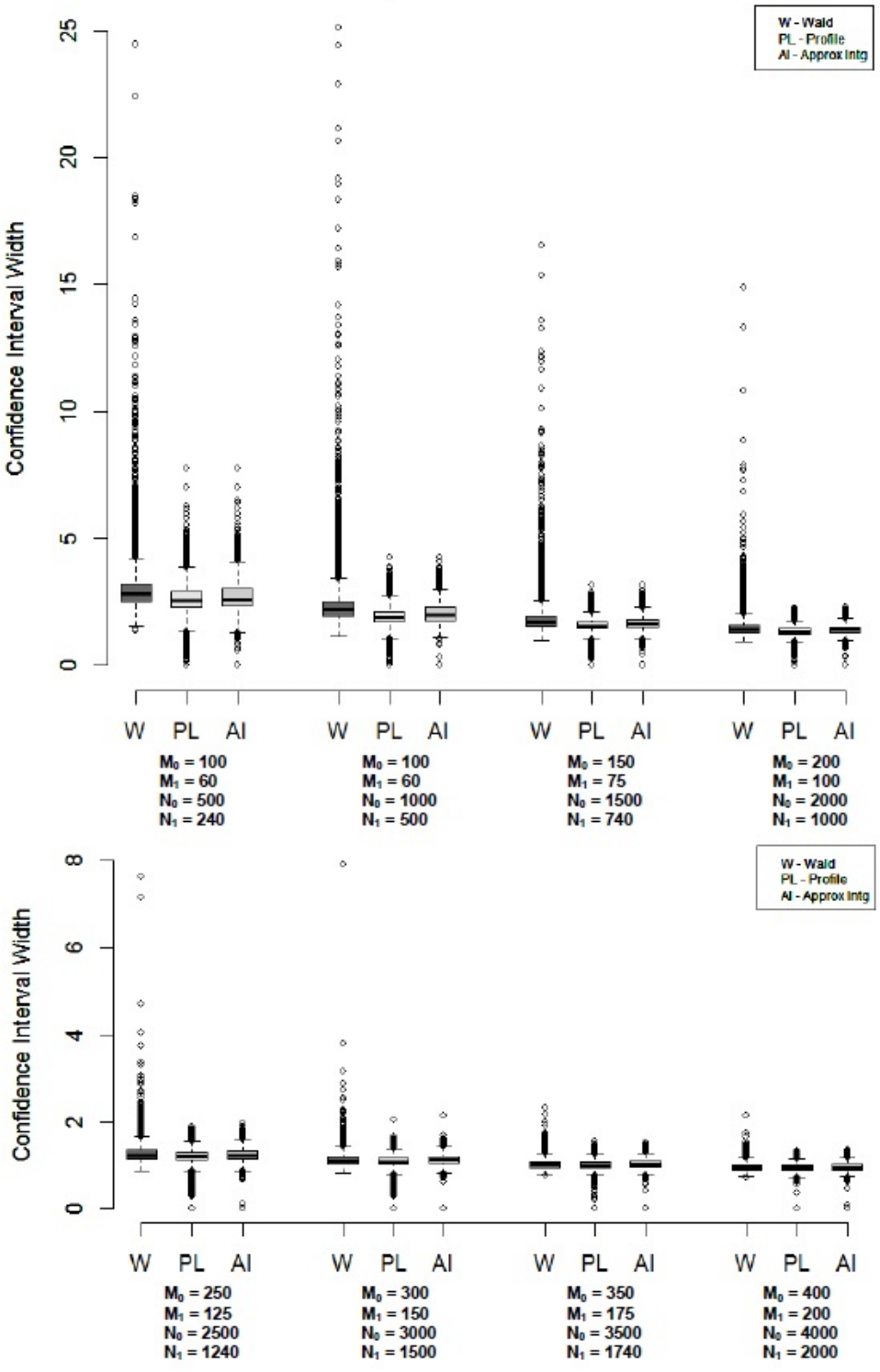

Figure 3.4: Interval widths of W, PL, and AIL CIs for cohort study under $C_{3.2}$.

Hence, we conclude that for low non-differential misclassification with an odds ratio of $\psi = 4$, the PL and AIL CIs yielded less-than-nominal coverage. This fact is possibly explained by the relatively small interquartile ranges for each sample-sizes scenario (A1 - A8) seen in Figure 3.4. Notice that although the PL and AIL CI coverage probabilities slightly underestimated the nominal confidence level of 95%, both intervals showed approximately correct coverage probability as the fallible and infallible sample sizes increased.

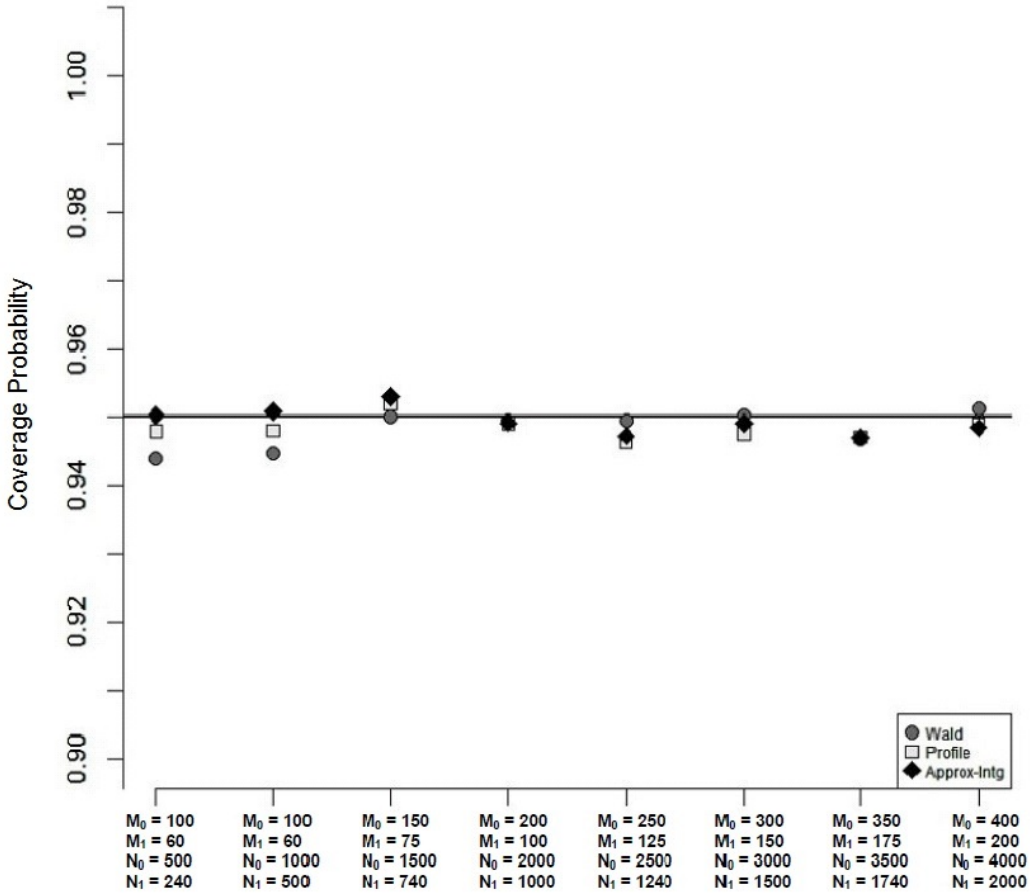

Figure 3.5: Coverage rates of W, PL, and AIL CIs for cohort study under $C_{3.3}$.

The last parameter-configuration scenarios we considered were the cases of high non-differential misclassification and odds ratios $\psi = 2, 4$. Refer to Table 3.3 for the corresponding misclassification probabilities. We compared the coverage

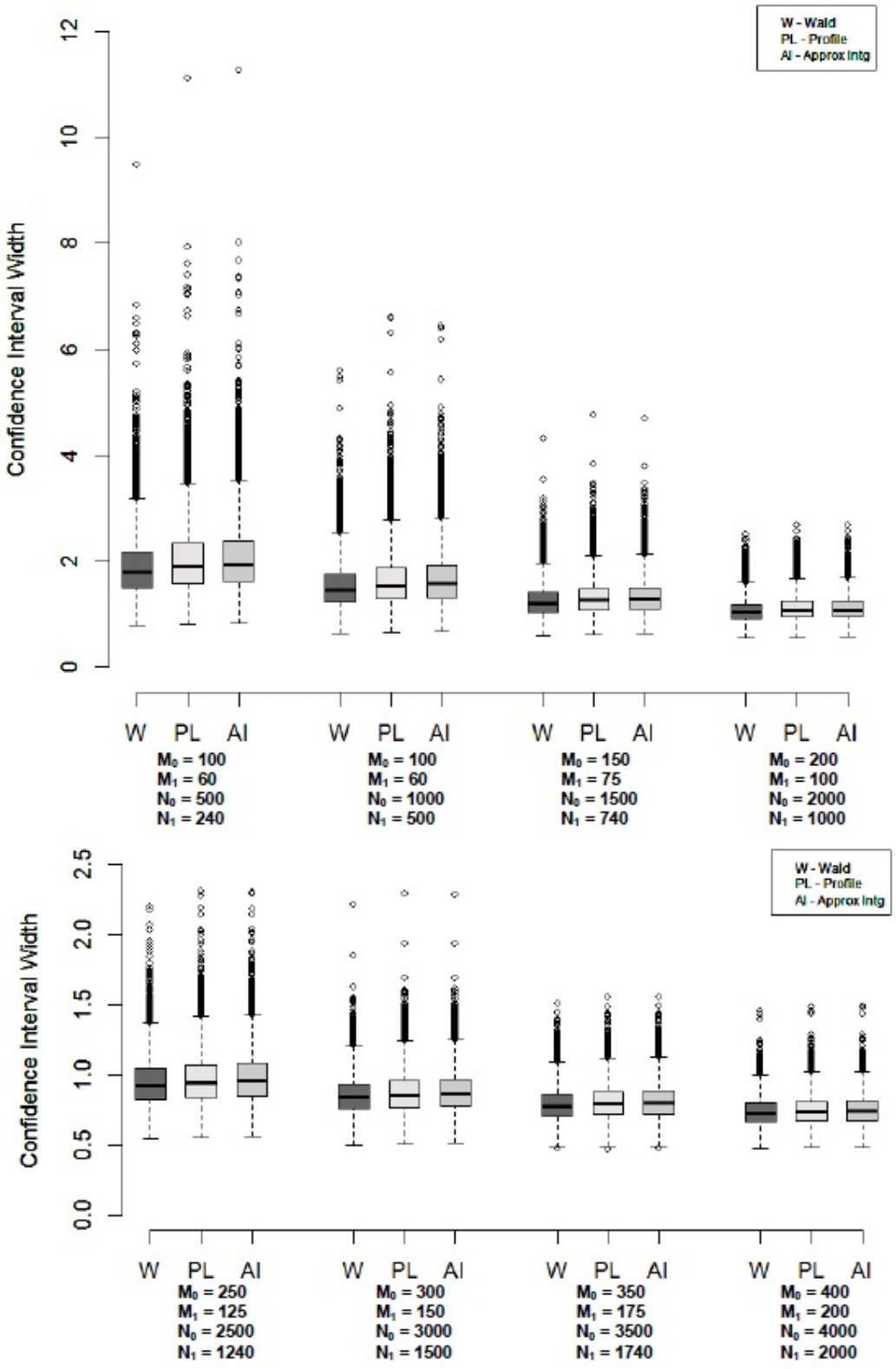

Figure 3.6: Interval widths of W, PL, and AIL CIs for cohort study under $C_{3.3}$.

probabilities (Figure 3.5 and Figure 3.7) and median interval widths (Figure 3.6 and Figure 3.8) of the Wald, PL, and AIL CIs under the different sample-size scenarios given in Table 3.4.

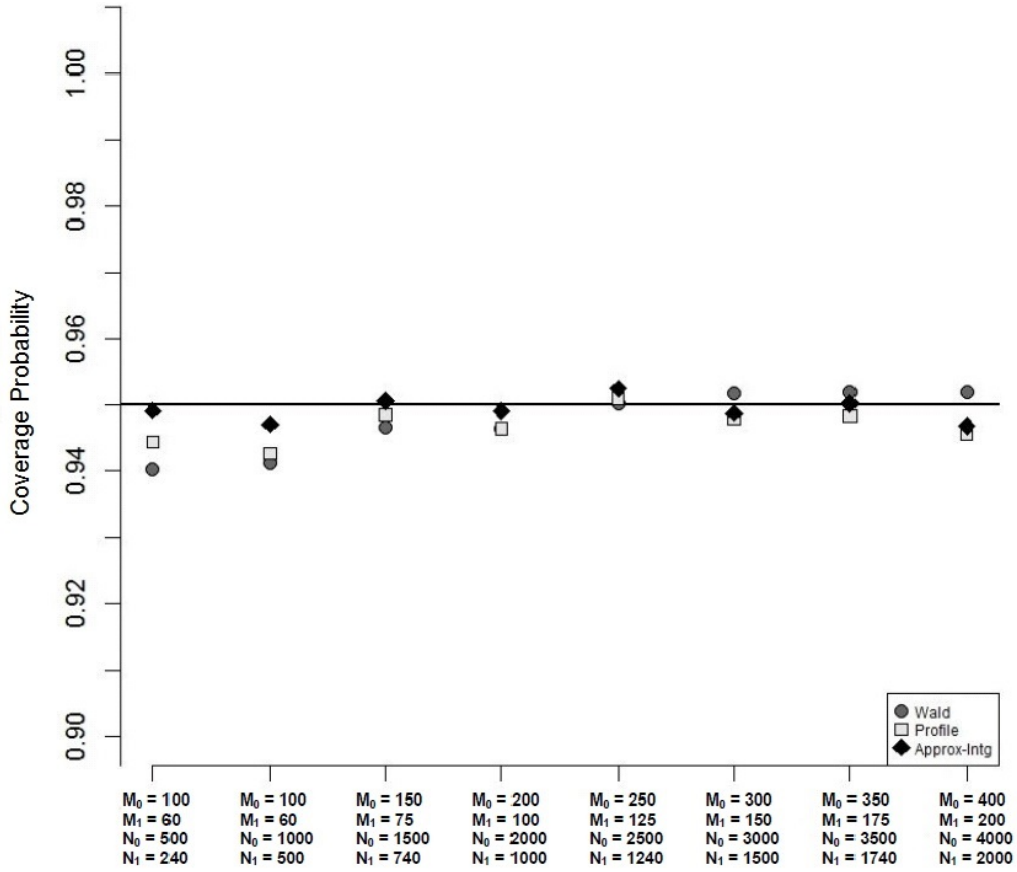

Figure 3.7: Coverage rates of W, PL, and AIL CIs for cohort study under $C_{3.4}$.

The coverage probabilities of all three CIs displayed similar behaviors (see Figure 3.5). The Wald and PL CIs slightly underestimated the nominal confidence level of 95% for the two smallest sample-size scenarios. However as the sample sizes increased (A3-A8), the Wald, PL, and AIL CIs approached the nominal 95% confidence level. Most importantly, the AIL interval consistently closely approximated the nominal 95% confidence level for all sample-size scenarios. Thus, the AIL interval yielded the best coverage probability properties for the interval estimation of ψ for the configuration $C_{3.3}$. In Figure 3.6, we see that all three CIs yielded very

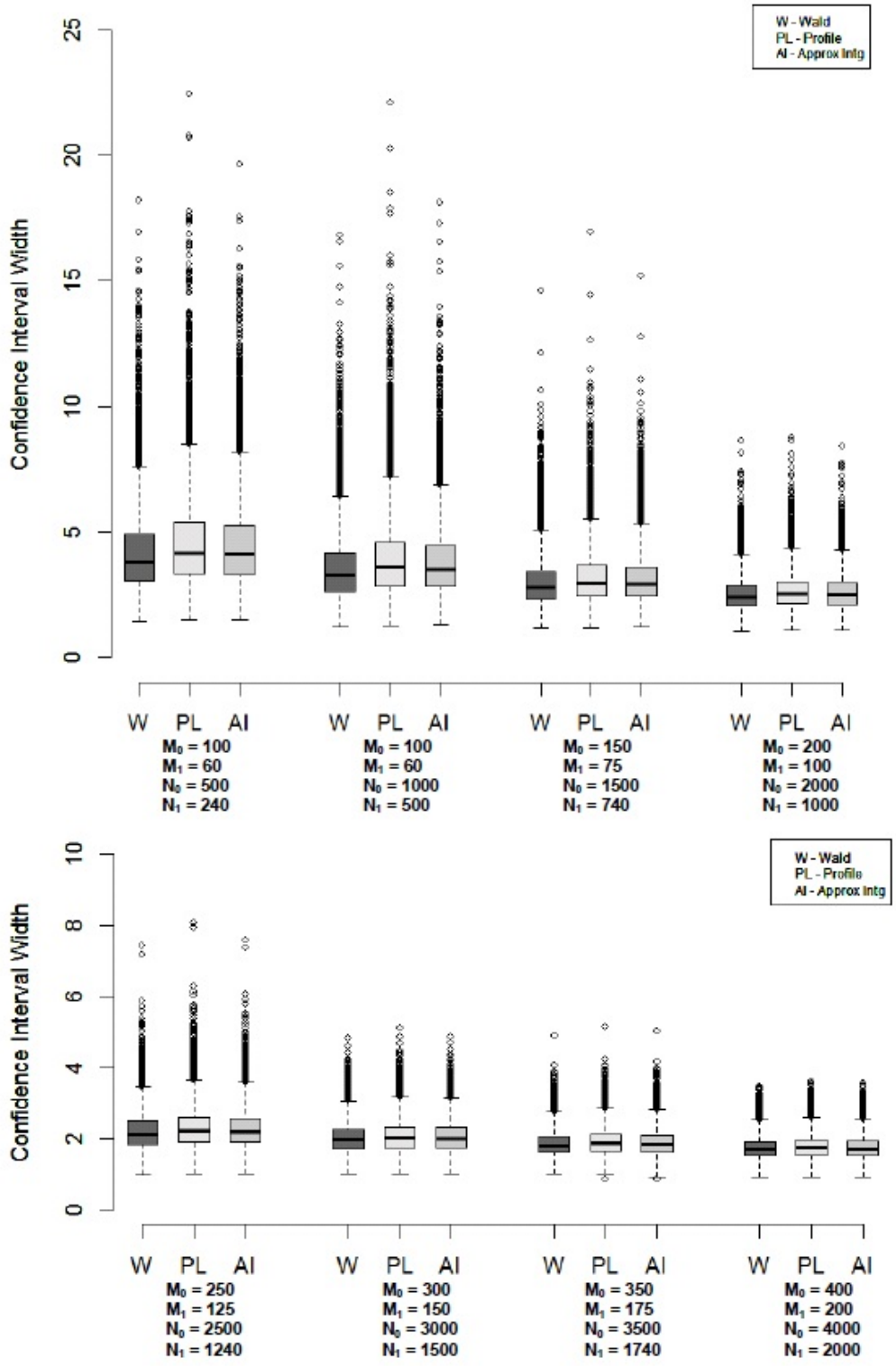

Figure 3.8: Interval widths of W, PL, and AIL CIs for cohort study under $C_{3.4}$.

similar median interval widths for configuration $C_{3.3}$. As the sample sizes increased, the median interval widths decreased as expected. Overall, for the parameter configurations considered here, the AIL CI for ψ yielded the best coverage probability properties and had reasonable interval width properties.

Under the scenario of high non-differential misclassification and an assumed OR parameter of $\psi = 4$, Figure 3.7 shows that all three CIs displayed similar coverage probability and median width behaviors as for $C_{3.3}$. Notice in Figure 3.7 that for the sample size scenarios A1 - A4, the Wald and PL CIs slightly underestimated the nominal 95% confidence level. However, as the sample sizes increased, both CIs approached the nominal 95% confidence level. The AIL CI yielded coverage probabilities that were the closest to nominal level for all sample-size scenarios. Figure 3.8 shows that all three CIs yielded very similar median-interval-widths and interquartile ranges for the corresponding interval widths. However, for sample-size scenarios A1 - A4, the Wald CI yielded slightly shorter median CI widths. As the sample sizes increased, their median interval widths consistently decreased as one would expect. For relatively high non-differential misclassification, the AIL CI was the preferred interval estimator for ψ for the parameter configuration $C_{3.4}$.

3.8 Comments

In this paper we have derived and contrasted one-likelihood-based and two pseudo-likelihood-based CIs for the OR parameter ψ in a binary data cohort study using a double-sampling scheme for binary data subject to correct for non-differential misclassification. In the case of low non-differential misclassification with an assumed OR of $\psi = 2$ or $\psi = 4$, we have concluded that the Wald CI consistently over-covered and the PL and AIL CIs consistently under-covered for relatively small sample sizes. However, as the sample sizes increased, all three approximate CIs showed excellent potential as omnibus CIs for estimating ψ. In the case of high non-

differential misclassification with an assumed odds ratio of $\psi = 2$ or $\psi = 4$, the AIL CI yielded almost exact coverage probabilities for all sample-size scenarios considered here (Table 3.4). We conjecture that the reason the AIL CI showed excellent coverage properties was because of the fact that with a high probability of misclassification, we have more information to estimate the misclassification parameters and, thus, the misclassification error rates are more stable estimates. The AIL CI coverage probabilities determined for the high non-differential misclassification case appeared to be consistent with the findings of several authors who have examined similar likelihood and pseudo-likelihood-based intervals for different proportion parameters using double-sampling (Boese (2005), Greer (2008) and Markova (2011)).

CHAPTER FOUR

Approximate Interval Estimation of a Poisson Rate Parameter Using Data Subject to Misclassification

4.1 Abstract

We derive four first-order asymptotic confidence intervals (CIs) for estimating a Poisson rate parameter using a double sample of inerrant data and data containing false-negative and false-positive observations. The four CIs are then compared in terms of average coverage probability and median interval width via a simulation experiment. We conclude that over the parameter configurations and sample sizes considered here, the Wald CI is the best omnibus CI for the Poisson rate parameter. Last, we apply all four interval estimation procedures to a real-data example.

4.2 Introduction

The biasing effects of misclassification on Poisson parameter estimation have not been documented as thoroughly as the misclassification-biasing effects of the binomial success parameter observation, especially when both false-positives and false-negatives are present. A false-positive takes place when a count occurs for any reason other than an event of interest. A false-negative occurs when an occurrence of interest is omitted (Bratcher and Stamey (2002)).

Suppose we wish to estimate the rate of an event from a population that yields data that are approximately Poisson distributed. For example, one might be interested in estimating the digestive disease mortality rate based on death certificates. Because the statistical analysis for this problem involves count data, we will very probably encounter misclassified counts. To obtain accuracy, one must adjust for this labeling error.

When misclassification is present, the double-sampling paradigm, developed by Tenenbein (1970), can be utilized to correct for the misclassification bias. This sampling method combines inerrant data from which one obtains a true count, a false-positive count, and a false-negative count, and errant data from which only a fallible count is available (Bratcher and Stamey (2002)).

In this chapter we derive two likelihood-based and one pseudo-likelihood-based confidence interval (CI) for a Poisson rate parameter using a double sample of inerrant data and data that contains both false-negative and false-positive observations. Specifically, we utilize Wald, score, profile log-likelihood (PL), and the approximated integrated likelihood (AIL) statistics to obtain CIs for a Poisson rate parameter. We assess the efficacy of these four CIs in terms of coverage probability and median interval width properties via a Monte Carlo simulation experiment. We also apply these four interval estimators that utilize information in a double-sampling scheme to a real-data example presented in Kircher, Nelson, and Burdo (1985).

Many authors have derived methods for parameter estimation for counted data with misclassification for both binomial and Poisson models. Such authors include Bross (1954); Tenenbein (1970, 1971); Hochberg (1977); Chen (1979); Whittemore and Gong (1991); Sposto, Preston, Shimizu, and Mabuchi (1992); Viana, Ramakrishnan, and Levy (1993); Joseph, Gyorkos, and Coupal (1995); Bratcher and Stamey (2002); Boese, Young, and Stamey (2006); Greer (2008); and Riggs, Young, and Stamey (2009).

In particular, Hochberg (1977) and Chen (1979) have used double sampling to correct for misclassification in categorical models to obtain maximum likelihood estimators (MLEs). Tenenbein (1970), Whittemore and Gong (1991), Viana et al. (1993), and Joseph et al. (1995) have focused on the bias of estimators due to misclassification and on methods to correct this bias. Bratcher and Stamey (2002)

have estimated Poisson rates in the presence of both false-negative and false-positive misclassification utilizing a Bayesian approach. Also, Sposto et al. (1992) have considered estimating complementary cancer and non-cancer mortality rates using data subject to misclassification. These two populations are assumed to be mutually exclusive and, thus, counts were assigned only to one group. Stamey, Young, and Stephens (2005) have found closed-form MLEs of parameters for this model in the case of no covariates using a double-sampling scheme similar to that of Tenenbein (1970). Riggs et al. (2009) have derived three interval estimators for complementary Poisson rates where the data are possibly misclassified.

We have organized this chapter as follows. In Section 4.3, we present the statistical model, basic terms, and notation. In Section 4.4, we derive the full-data likelihood along with the parameter MLEs. In Section 4.5, we provide an EM algorithm for estimating the restricted maximum-likelihood estimators (RMLEs)for two nuisance parameters. In Section 4.6, we present the observed Fisher information matrix (OFIM) that is used to derive the four CIs for a Poisson rate parameter using a double sample of infallible and fallible data. We examine the performance of the proposed four CIs in Section 4.7 using a Monte Carlo simulation, and in Section 4.8 we apply the four MLE-motivated CIs for a Poisson rate parameter model under a double-sampling paradigm to a real data set. We conclude with some brief comments in Section 4.9.

4.3 The Model

To derive likelihood- and pseudo-likelihood-based CIs for the Poisson rate parameter, we consider a statistical model. For our current problem, we use a double-sampling procedure with the assumption of possible misclassification. Hence, we collect fallible and infallible training data from a population that can be approximately modeled using a Poisson distribution. On the first observation-opportunity

size A_0, we apply both a fallible and an infallible test classifier. For the second observation-opportunity size A, which is generally much larger than A_0, the data collected is less costly and consists solely of counts using only a fallible classification method. Let t be the actual number of occurrences, y the number of false-positives, and x the number of false-negatives. Then, we assume $z = t + y - x$ is the number of observed occurrences actually reported from A. The unobservable variables t, y, and x are conditionally independent and have the distributions

$$t|\lambda \sim Poisson(A\lambda),$$

$$y|\phi, t \sim Poisson(A\phi),$$

and (4.1)

$$x|t, \theta \sim Binomial(t, \theta).$$

The occurrence false-positive and false-negative rates, are represented by λ, ϕ, and θ, respectively. From (4.1) and using the transformation $t = z + x - y$, we have that

$$f(z, y, x|\lambda, \phi, \theta) = \frac{e^{-A\lambda}(A\lambda)^{z+x-y}}{(z+x-y)!} \frac{e^{-A\phi}(A\phi)^{y}}{y!} \begin{pmatrix} z+x-y \\ x \end{pmatrix} \theta^{x} (1-\theta)^{z-y} \quad (4.2)$$

is the joint density function of the unobservable variables, z, y, and x. Thus, from (4.2), $g(z|\lambda, \theta, \phi) \sim Poisson(A[\lambda(1-\theta) + \phi])$.

If the misclassification parameters ϕ and θ are unknown, we can use a training sample to determine the MLEs for λ, ϕ, and θ analogous to the approach taken by Tenenbein (1970) for the binomial model with false-negative and false-positive misclassification. We use two different search techniques on the training observation-opportunity size A_0. One search method is an expensive error-free method that results in a true count of size t_0. A less-expensive, error-prone search technique is also utilized that yields a fallible count of $z = t_0 + y_0 - x_0$, where x_0 and y_0 are false-negative and false-positive occurrences, respectively. We remark that because

the actual count is distributed as a Poisson random variable, the maximum number of false-negatives is t_0 and, thus, x_0 is distributed as a binomial random variable. Also, we assume that we are sampling from a countably infinite population, and, therefore, the actual count has no effect on the number of false-positives. Hence, we obtain

$$\begin{aligned} t_0 &\sim Poisson(A_0\lambda), \\ y_0 &\sim Poisson(A_0\phi), \\ x_0 &\sim Binomial(t_0, \theta), \end{aligned} \tag{4.3}$$

and

$$z \sim Poisson(A[\lambda(1-\theta)+\phi]).$$

From (4.3), the joint density function of the observable variables, t_0, y_0, x_0, and z is

$$\begin{aligned} f\left(t_0, x_0, y_0, z|\lambda, \phi, \theta\right) = \frac{e^{-A_0\lambda}(A_0\lambda)^{t_0}}{t_0!}\,\frac{e^{-A_0\phi}(A_0\phi)^{y_0}}{y_0!}\begin{pmatrix} t_0 \\ x_0 \end{pmatrix}\theta^{x_0}\left(1-\theta\right)^{t_0-x_0} \\ \times \frac{[\lambda(1-\theta)+\phi]^z e^{-A[\lambda(1-\theta)+\phi]}}{z!}. \end{aligned} \tag{4.4}$$

4.4 The Full Data Likelihood and Maximum-Likelihood Estimators

Let $\boldsymbol{\rho} \equiv (\lambda, \boldsymbol{\eta}')'$ represent the parameter vector where λ is the occurrence rate parameter of interest and $\boldsymbol{\eta}$ is the vector of nuisance parameter. Also, let $\mathbf{d} \equiv (z, t_0, y_0, x_0)'$ denote the observed data counts. From (4.4), if we let $L_o \equiv L_o(\lambda, \phi, \theta|\mathbf{d})$ then the concentrated observed data likelihood function is

$$L_o \propto \lambda^{t_0} e^{-\lambda A_0} \phi^{y_0} e^{-\phi A_0} \theta^{x_0} (1-\theta)^{t_0-x_0} [\lambda(1-\theta)+\phi]^z e^{-A[\lambda(1-\theta)+\phi]}, \tag{4.5}$$

so that the observed data log-likelihood function ℓ_o is

$$\begin{aligned} \ell_o \propto\, & t_0 \ln(\lambda) + y_0 \ln(\phi) + x_0 \ln(\theta) + (t_0 - x_0)\ln(1-\theta) + z\ln[\lambda(1-\theta)+\phi] \\ & + (-A_0\lambda - A_0\phi - A[\lambda(1-\theta)+\phi]). \end{aligned} \tag{4.6}$$

Taking partial derivatives of (4.6) with respect to λ, ϕ, and θ yields the three estimating equations

$$\frac{t_0}{\lambda} - A_0 + \frac{z(1-\theta)}{\lambda(1-\theta)+\phi} - A(1-\theta) = 0,$$
$$\frac{y_0}{\phi} - A_0 + \frac{z}{\lambda(1-\theta)+\phi} - A = 0,$$

and (4.7)

$$\frac{x_0}{\theta} - \frac{t_0 - x_0}{1-\theta} - \frac{z\lambda}{\lambda(1-\theta)+\phi} + A\lambda = 0,$$

respectively. Solving the estimating equations (4.7) for the respective parameters of interest yields the MLEs

$$\hat{\lambda} = \alpha_1 \left(\frac{t_0 + z(1 - y_0/z_0)}{A_0} \right) + \alpha_2 \left(\frac{x_0}{A_0} \right), \tag{4.8}$$

$$\hat{\phi} = \frac{(z+z_0)y_0}{(A+A_0)z_0}, \tag{4.9}$$

and

$$\hat{\theta} = \frac{x_0(A+A_0)}{A_0\left(z + t_0 + \frac{x_0}{A_0}A - \frac{y_0}{z_0}z\right)}, \tag{4.10}$$

where $z_0 = t_0 + y_0 - x_0$, $\alpha_1 = A_0/(A+A_0)$, and $\alpha_1 + \alpha_2 = 1$. Hence, z_0 is the number of occurrences observed using the fallible search technique in the training sample, y_0/z_0 is the proportion of false positives in the observed training-sample data, and x_0/A_0 is the estimated rate of false negatives. We remark that (4.8) can be expressed as

$$\hat{\lambda} = \frac{z + t_0 + (x_0/A_0)A - (y_0/z_0)z}{A_0 + A}. \tag{4.11}$$

The numerator of (4.11) is composed of three terms. The first term is the sum of the observed occurrences using the infallible method on the training sample and using the fallible method on the larger sample. The second term is the weighted rate of false-negatives in the fallible sample and the third term is the weighted proportion of false-positives from the training sample. The sum of those three terms is then averaged over the total sample size $A_0 + A$.

The misclassification estimators $\hat{\phi}$ and $\hat{\theta}$ have straightforward interpretations. The false-negative proportion estimator (4.10) is the proportion of false-negatives from the training sample per the estimated rate of occurrence. The false-positive rate estimator (4.9) is the average of the observed fallible data from both samples multiplied by the proportion of false-positives from the training sample.

4.5 Restricted Maximum-Likelihood Estimators

To compute the score, the PL and AIL CIs, we need the unrestricted MLEs (4.8), (4.9), and (4.10), as well as the RMLEs of the nuisance parameters. Unlike the unrestricted MLEs, we can derive no closed form for the RMLEs. Therefore, we detail an EM algorithm to determine the RMLEs for a fixed value of λ in Appendix C.1.

The likelihood used to obtain the EM algorithm results from the two independent samples. First, a training observation-opportunity size A_0 where fallible and infallible classifiers are both applied, and the fallible observation-opportunity size A, where only the fallible classifier is applied. Multiplying the fallible data associated likelihood (4.2) with the likelihood for the training sample (4.5), we obtain the complete-data likelihood $L^c(\lambda, \boldsymbol{\eta}|\boldsymbol{d}^c)$ expressed in (C.1.1) of Appendix C.1, where $\boldsymbol{d}^c \equiv (z, y, x, t_0, y_0, x_0)'$ represents the complete data with x and y unobservable.

4.6 Four Asymptotic Confidence Intervals for λ

Here, we define and develop four first-order asymptotic CIs for a Poisson rate parameter λ where the data is subject to false-negative and false-positive misclassification. First, we derive the OFIM that we use to estimate the MLE variances for λ and $\boldsymbol{\eta}$. We then construct one likelihood and two pseudo-likelihood-based CIs for

λ. The OFIM is

$$\mathbf{J}(\lambda, \boldsymbol{\eta}) = - \begin{bmatrix} \frac{\partial^2 \ell_o}{\partial \lambda^2} & \frac{\partial^2 \ell_o}{\partial \lambda \partial \phi} & \frac{\partial^2 \ell_o}{\partial \lambda \partial \theta} \\ . & \frac{\partial^2 \ell_o}{\partial \phi^2} & \frac{\partial^2 \ell_o}{\partial \phi \partial \theta} \\ . & . & \frac{\partial^2 \ell_o}{\partial \theta^2} \end{bmatrix}, \tag{4.12}$$

where ℓ_o is the log-likelihood function in (4.5). In Appendix C.2 we give expressions for each of the terms in (4.12). Also, we partition (4.12) so that

$$\mathbf{J}\left(\lambda, \boldsymbol{\eta}\right) \equiv \begin{bmatrix} J_{11} & \boldsymbol{J}_{12} \\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22} \end{bmatrix}, \tag{4.13}$$

where $J_{11} = J_\lambda$ is a scalar and we emphasize that (4.13) is used in the construction of the subsequent CIs for λ.

4.6.1 A Wald Confidence Interval for λ

The Wald CI is based on the large-sample properties of the Wald statistic with nuisance parameters. Here, the corresponding Wald statistic for λ with nuisance parameters $\boldsymbol{\eta}$ is

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

where $J^{11} = (J_{11} - \boldsymbol{J}_{12}\boldsymbol{J}_{22}^{-1}\boldsymbol{J}_{21})^{-1}$ (refer to Pawitan (2001)). An approximate $100(1-\alpha)\%$ Wald CI for the occurrence rate parameter consists of the values of λ that satisfy

$$(\hat{\lambda} - \lambda)^2 [J^{11}(\hat{\lambda}, \hat{\boldsymbol{\eta}})]^{-1} < \chi_1^2(1-\alpha), \tag{4.14}$$

where $\chi_1^2(1-\alpha)$ denotes the $(1-\alpha)$ quantile of a central chi-squared distribution with one degree of freedom. Thus, solving (4.14) directly for λ, we get

$$\hat{\lambda} - \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\lambda}, \hat{\boldsymbol{\eta}})]} < \lambda < \hat{\lambda} + \sqrt{\chi_1^2(1-\alpha)[J^{11}(\hat{\lambda}, \hat{\boldsymbol{\eta}})]} \tag{4.15}$$

as an approximate $(1-\alpha)\%$ Wald CI for λ using the double-sampling paradigm with misclassified counts. The OFIM (4.13) is derived in Appendix C.2.

4.6.2 A Score Confidence Interval for λ

Second, we derive a score CI, which is a likelihood related interval. The score-based CI for λ involves inverting the score statistic

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

where λ is the parameter of interest and $\boldsymbol{\eta}$ is the vector of nuisance parameters. An approximate $100(1-\alpha)\%$ score CI is composed of the values of λ that satisfy

$$[S(\lambda, \hat{\boldsymbol{\eta}}_\lambda)]^2 \left[J^{11}(\lambda, \hat{\boldsymbol{\eta}}_\lambda)\right] < \chi^2_1(1-\alpha), \tag{4.16}$$

where $\hat{\boldsymbol{\eta}}_\lambda$ is the vector of RMLEs of the nuisance parameters evaluated at a fixed value of λ. The interval of solutions to (4.16) must be found numerically.

4.6.3 A Profile Likelihood Confidence Interval for λ

The profile log-likelihood CI for λ involves inverting the profile log-likelihood statistic. For sufficiently large n,

$$-2\left[\ell(\lambda, \hat{\boldsymbol{\eta}}_\lambda) - \ell(\hat{\lambda}, \hat{\boldsymbol{\eta}})\right] \dot{\sim} \chi^2_1.$$

Therefore, an approximate $100(1-\alpha)\%$ profile likelihood CI for the occurrence rate is the set of values of λ that satisfy

$$-2\left[\ell(\lambda, \hat{\boldsymbol{\eta}}_\lambda) - \ell(\hat{\lambda}, \hat{\boldsymbol{\eta}})\right] < \chi^2_1(1-\alpha), \tag{4.17}$$

where $\hat{\lambda}$ is given in (4.8) and $\hat{\boldsymbol{\eta}}_\lambda$ is defined in Subsection 4.6.2. As in (4.16), the values of λ that satisfy (4.17) must be determined numerically.

4.6.4 An Approximate Integrated Likelihood Confidence Interval for λ

Last, we obtain a CI for ψ based on the approximate integrated likelihood (AIL). We use the Laplace approximation to derive a more computationally feasible likelihood and express the AIL function as

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

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

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

where $\boldsymbol{J}_{\boldsymbol{\eta}}$ is the nuisance parameter sub-matrix of the observed information matrix. Thus, from (4.13) we have $\hat{\boldsymbol{J}}_{\boldsymbol{\eta}}$ is $\boldsymbol{J}_{22}$ evaluated at the MLEs of the nuisance parameters. We use an iterative bisectional method to determine the interval of λ values that satisfy the inequality

$$-2\left[\ell_{AI}(\lambda) - \ell_{AI}(\hat{\lambda}_{AI})\right] < \chi^2_1(1-\alpha), \tag{4.18}$$

where $\ell_{AI}(\lambda)$ is the log AIL evaluated at a fixed value of λ.

4.7 A Monte Carlo Simulation

Here, we examine coverage probability and interval width properties of the four competing interval estimators of λ described in (4.15) - (4.18) by using a Monte Carlo simulation. We studied these intervals under parameter configurations $C_1 - C_4$ displayed in Table 4.1. Each parameter configuration is examined with a fixed value

Table 4.1: Parameter Configurations for Study of CIs with $\lambda = 20$

Configuration	θ	ϕ	A	A_0
C_1	0.15	0.6	$10, 20, 40, 10, 20, 40$	$3, 3, 3, 8, 8, 8$
C_2	0.25	0.6	$10, 20, 40, 10, 20, 40$	$3, 3, 3, 8, 8, 8$
C_3	0.15	0.9	$10, 20, 40, 10, 20, 40$	$3, 3, 3, 8, 8, 8$
C_4	0.25	0.9	$10, 20, 40, 10, 20, 40$	$3, 3, 3, 8, 8, 8$

of $\lambda = 20$, three different values of the fallible sample size $A \in \{10, 20, 40\}$, two different values of the training sample $A_0 \in \{3, 8\}$ and one unique combination of the misclassification parameters $\theta \in \{0.15, 0.25\}$ and $\phi \in \{0.6, 09\}$. For each parameter configuration, we generated 10,000 data sets and calculated the 95% CIs for the Poisson rate parameter using double sampling with misclassified counts for each of the four CI methods presented in Subsections 4.6.1 - 4.6.4.

4.7.1 Simulation Results

We first considered the simulation scenarios for the parameter configurations C_1 and C_2 shown in Table 4.1. We compared the coverage probabilities in Figure 4.1 and Figure 4.2 and the median interval widths in Figure 4.3 and Figure 4.4 for our four proposed CIs.

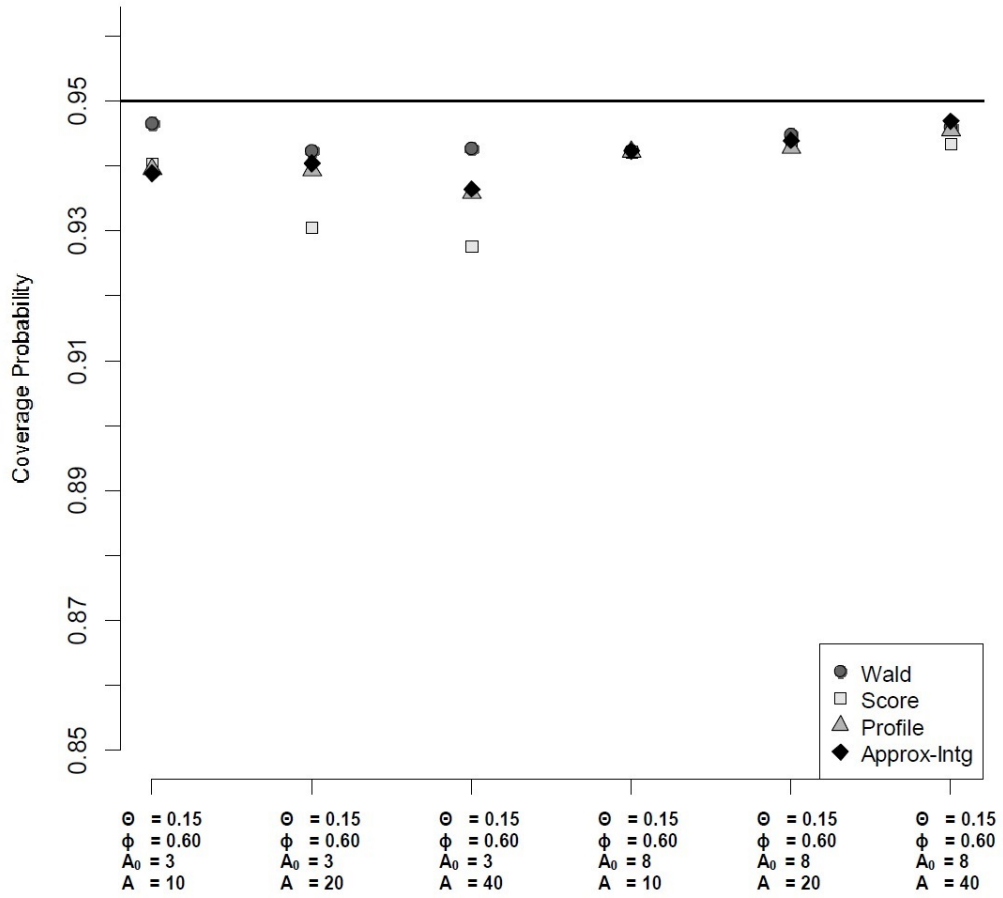

Figure 4.1: Coverage Rates for λ for parameter configuration C_1.

Note in Figure 4.1 and Figure 4.2 that as the fallible data observation-opportunity sizes ($A = 10, 20, 40$) increased, the training observation opportunity-size ($A_0 = 3$) remained relatively small and the rate of false positive observations was moderate ($\phi = 0.60$), the score, PL and AIL CIs slightly underestimated the nominal 95% confidence level. Moreover, as the training observation-opportunity size ($A_0 = 8$)

increased by more than 50%, we noticed that the coverage probabilities of the Wald, score, PL and AIL CIs all improved and approached the nominal level. A possible explanation for the under-coverage of the score, PL and AIL CIs is that the error-free observation-opportunity size A_0 is small relative to to the tainted observation-opportunity size A, and, thus, the misclassification rates ϕ and θ are not well estimated.

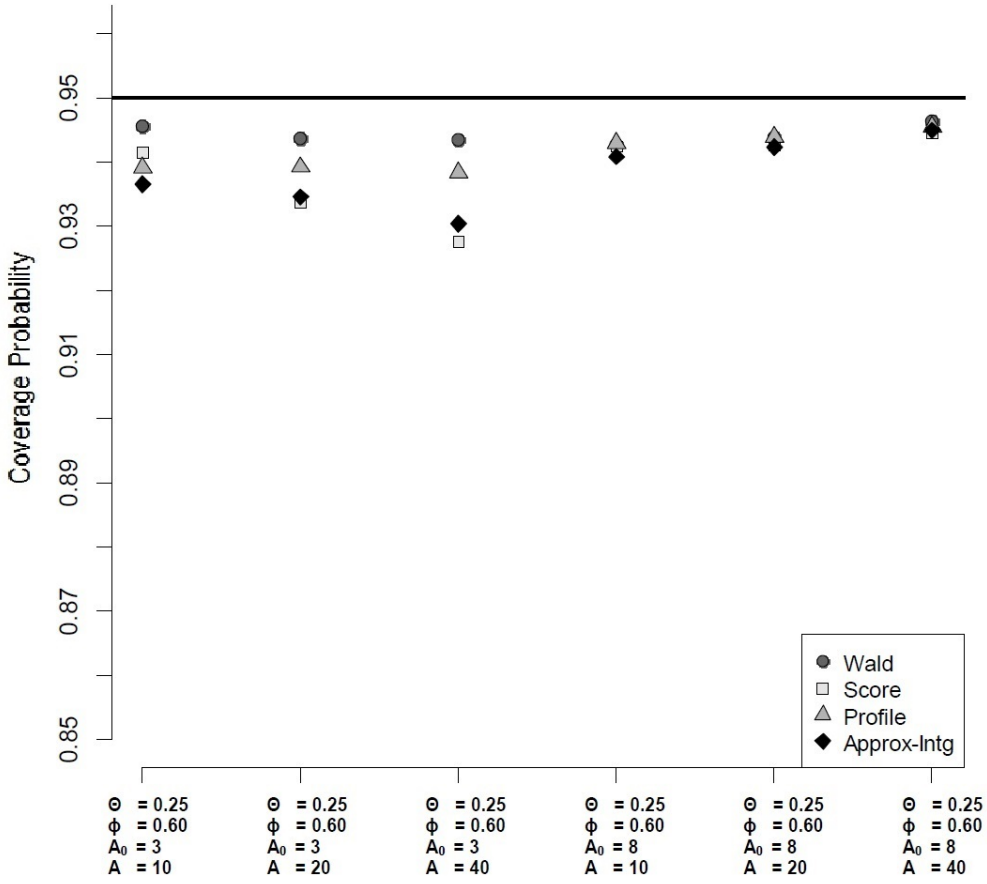

Figure 4.2: Coverage Rates for λ for parameter configuration C_2.

For parameter configurations C_2, we observed that for the misclassification rates $\theta = 0.25$ and $\phi = 0.60$, the PL CI yielded on average better coverage rates that the score and AIL CIs (see Figure 4.2). However, for parameter configurations C_1 and C_2, the Wald CI yielded the best coverage properties for the occurrence-rate parameter λ.

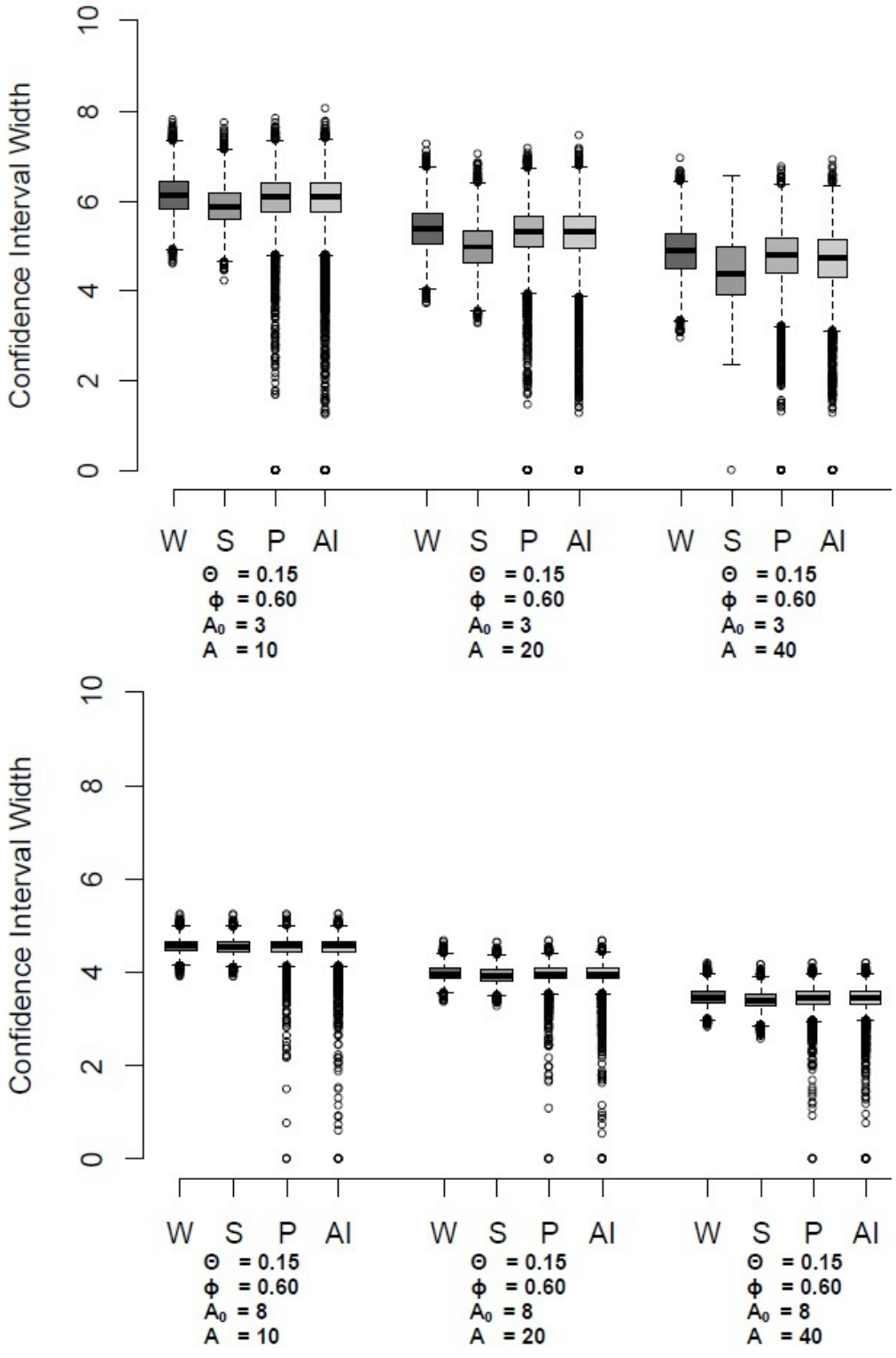

Figure 4.3: Confidence Interval Widths for λ for parameter configuration C_1.

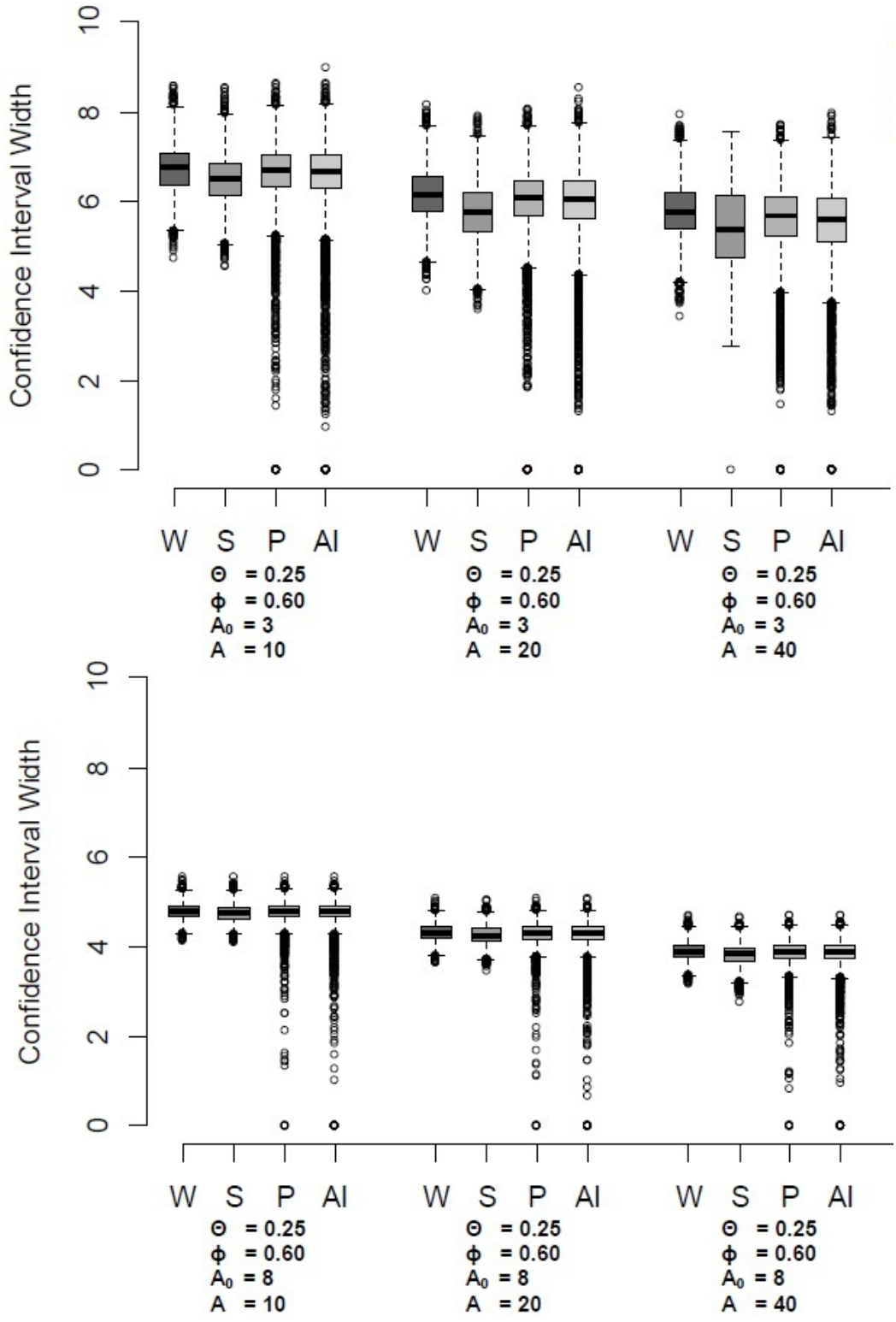

Figure 4.4: Confidence Interval Widths for λ for parameter configuration C_2.

From Table 4.2 and Figures 4.3 - 4.4, we observed that as the fallible data observation-opportunity size A and training data observation-opportunity size A_0 increased, the Wald CIs exhibited larger median widths compared to the other three CIs. This fact could explain why the Wald CI coverage probabilities outperformed the PL, AIL, and score CIs coverage behavior and that the Wald CI is more likely to contain λ. Also, from Figures 4.3 - 4.4, we observed that as A or A_0 increased, the widths of the CIs decreased, as expected.

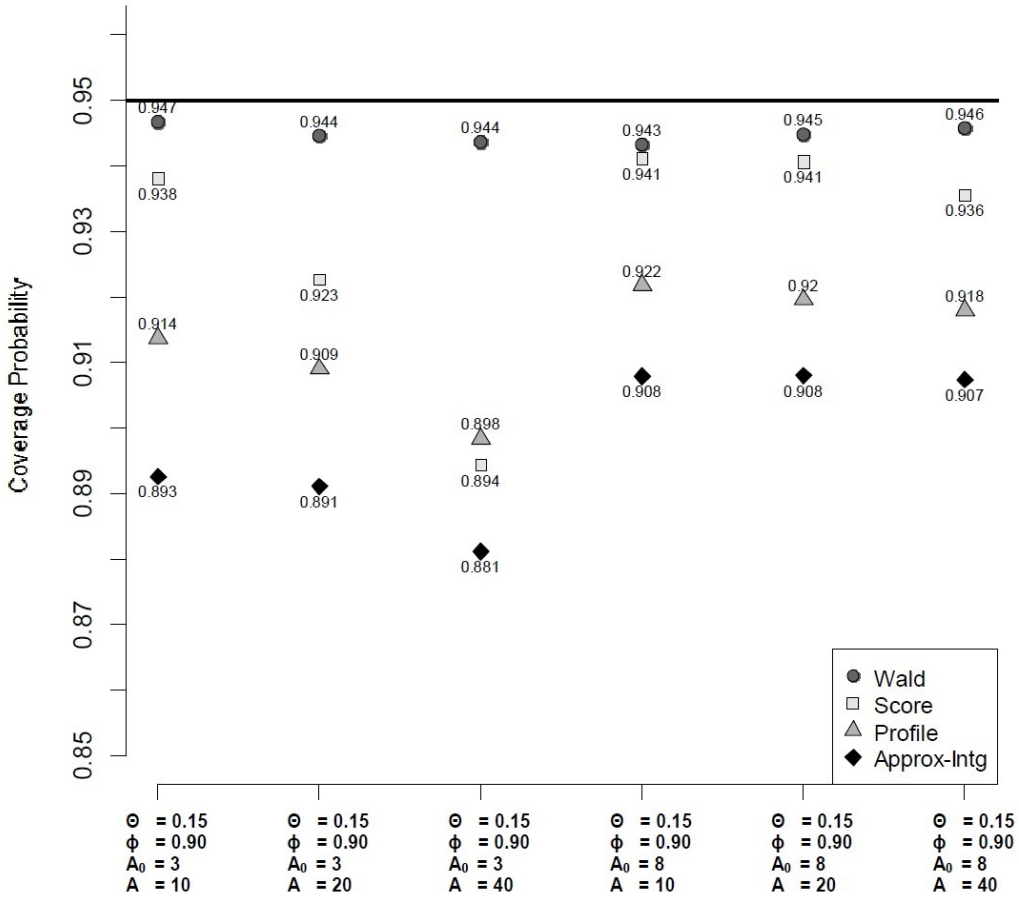

Figure 4.5: Coverage Rates for λ for parameter configuration C_3.

Moreover, the median interval widths for all four CIs intervals decreased as the training observation-opportunity size ($A_0 = 8$) increased, confirming the improved coverage probabilities for the PL, AIL, and score CIs shown in Figure 4.1 and Figure

4.2. However, the decrease in available training data ($A_0 = 3$) comes at the cost of noticeable under-coverage as shown in Figures 4.1 - 4.2.

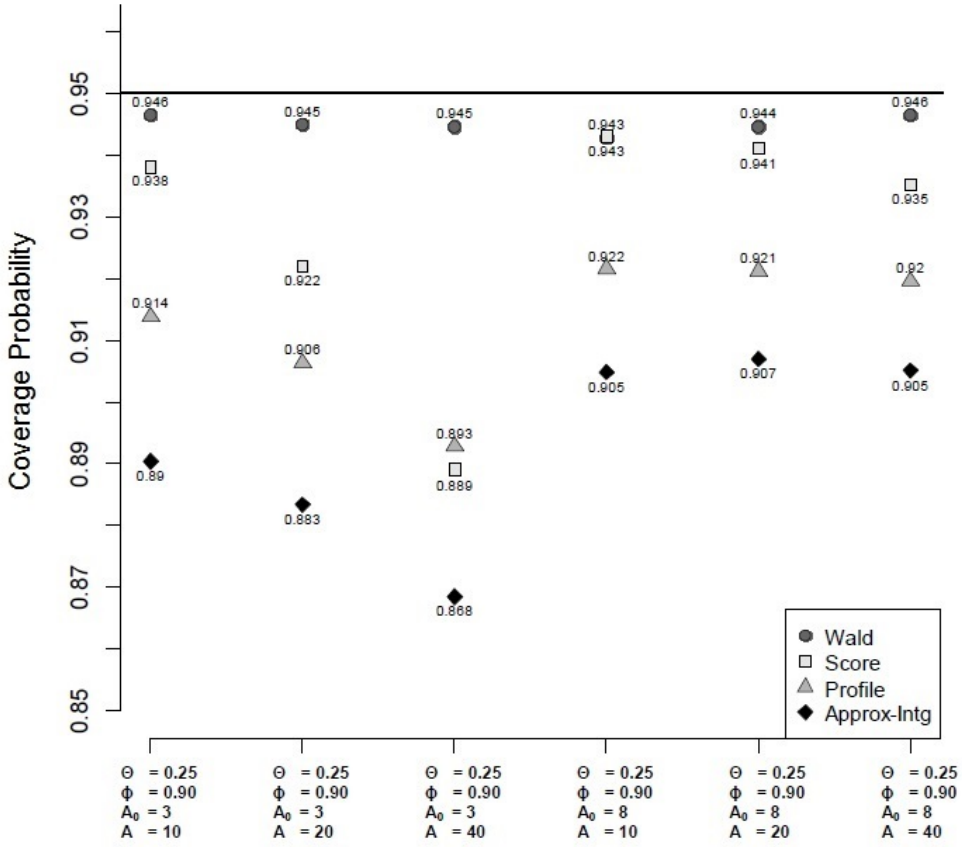

Figure 4.6: Coverage Rates for λ for parameter configuration C_4.

Next, we examined the simulation scenarios C_3 and C_4. Refer to Table 4.1 for the corresponding misclassification probabilities. Figure 4.5 and Figure 4.6 yielded insight into the effect of the error prone and training observation-opportunity sizes on the CIs' coverage probabilities when $\lambda = 20$ and the rate of false-positive observations was high ($\phi = 0.90$), while the rates of false-negative observations fluctuated from very small ($\theta = 0.15$) to relatively small ($\theta = 0.25$). As the fallible observation-opportunity size A increased, the amount of potential misinformation increased, causing a substantial decrease in coverage probabilities. However, this effect was offset by an increase in the training observation-opportunity size A_0.

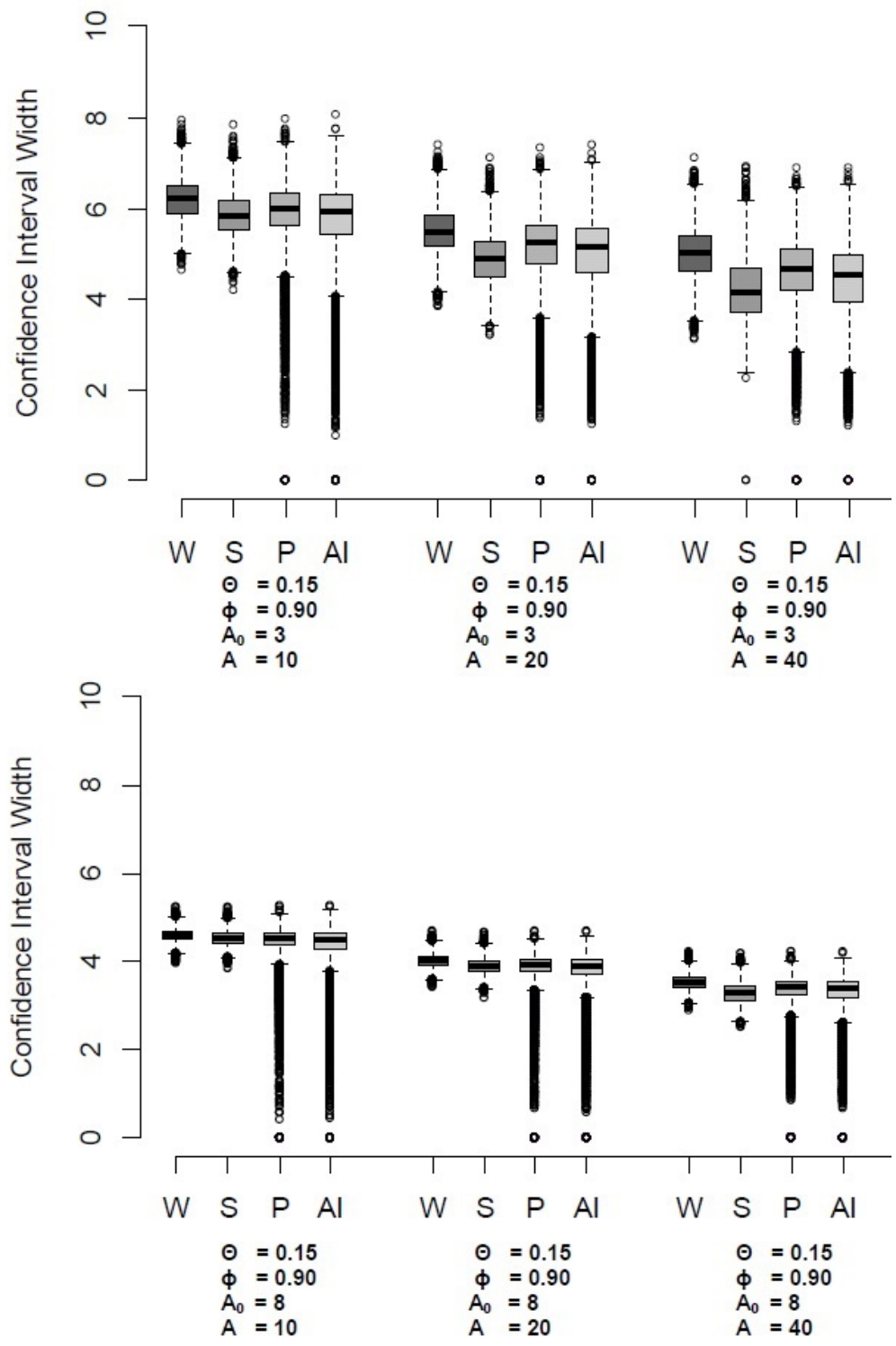

Figure 4.7: Confidence Interval Widths for λ for parameter configuration C_3.

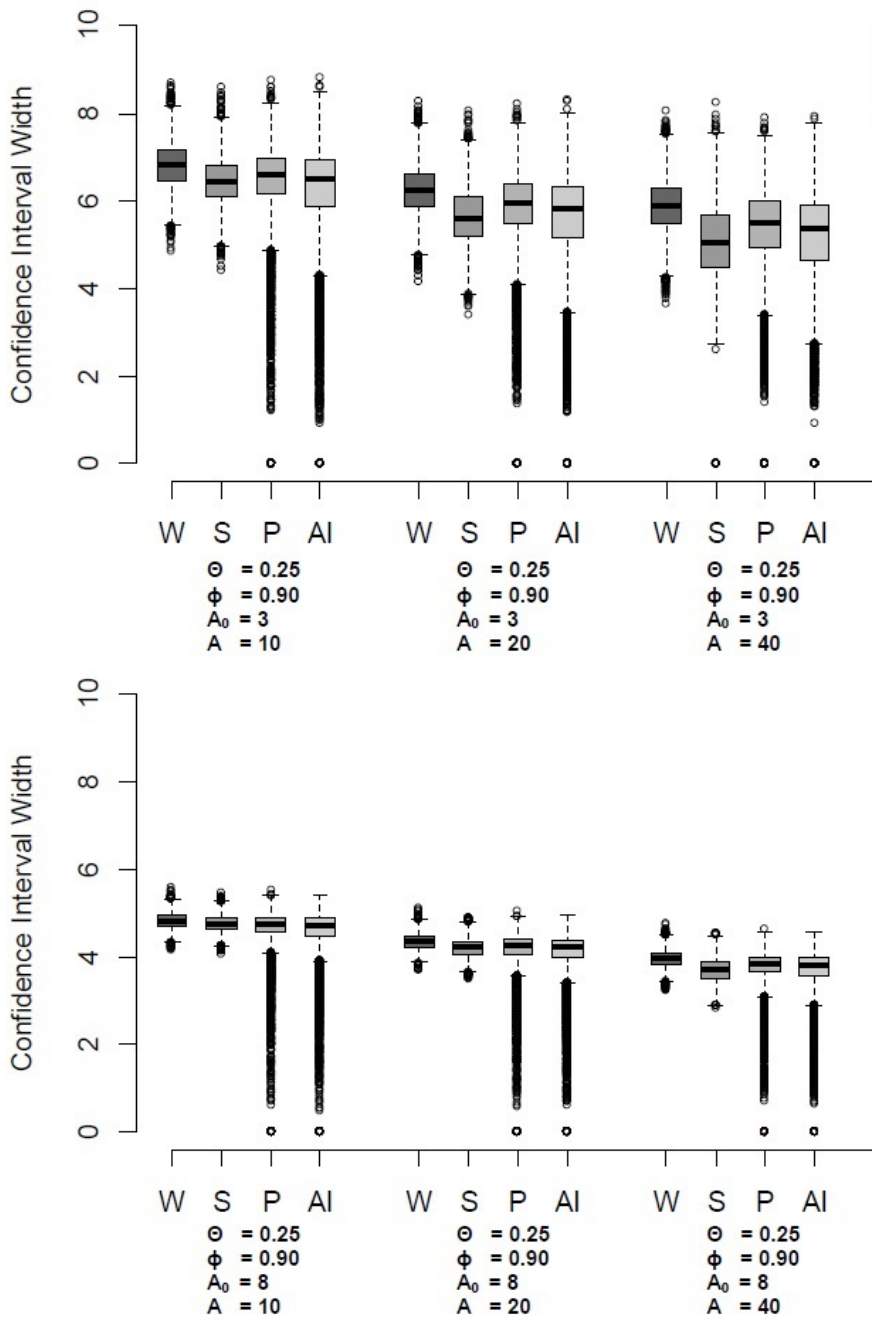

Figure 4.8: Confidence Interval Widths for λ for parameter configuration C_4.

From Figure 4.5 and Figure 4.6, we observed that as the observation opportunity sizes of the fallible data increased, and the training observation opportunity sizes ($A_0 = 3$) remained small while the rate of false positive observations ($\phi = 0.90$)

was high, the score, the PL and AIL CIs considerably undercovered λ. However the Wald CI maintained good coverage probability properties. We also noticed that for a larger rate of false positive observations ($\phi = 0.90$), the score CIs yielded larger coverage probabilities than the PL and AIL CIs.

Under a small to high misclassification rates ($\theta = 0.25, 0.15, \phi = 0.9$) and a small training observation opportunity size ($A_0 = 3$), we observed from Figures 4.5-4.6 that the coverage probabilities of the score, PL, AIL CIs considerably decreased as compared to Figures 4.1-4.2.

We next considered the median interval width properties of the Wald, score, PL and AIL CIs for configurations C_3 and C_4. From Table 4.2 and Figures 4.7 and 4.8, the ordering of the median interval widths from largest to smallest was the Wald, score, PL and AIL CIs. Hence, we conjecture that it is highly likely that the actual parameter value is more frequently contained in the wider Wald CIs, therefore resulting in the best overall coverage probability properties. Evidence for this conjecture appears in Figure 4.5 and Figure 4.6. Because the score CI median interval widths were larger than the corresponding AIL and PL median interval widths, its average coverage properties were superior to the AIL and PL CIs (see Figures 4.5 and 4.6). Also, as anticipated, an increase in the training observation-opportunity size A_0 and the fallible observation-opportunity size A produced shorter interval widths.

4.8 An Application

We now apply the four new approximate CIs for the Poisson rate parameter λ using both fallible and infallible data to a data set from Kircher et al. (1985). The data involve the death certificates and corresponding autopsy reports of all individuals who died in Connecticut in 1979. Suppose that our interest lies in the rate of death due to digestive disease per 10,000 person-years. Death certificates have

long been criticized as fallible, but follow-ups with autopsy reports are assumed to be infallible. An infallible count of 272 deaths were thoroughly autopsied and used as training data to gain information concerning the misclassification parameters. A larger sample of 3604 death certificates was used as the fallible sample, subject to both false-negative and false-positive misclassification. Here, $A = 39.4$, in terms of 10,000 person-years.

The infallible data contained $t_0 = 32$ deaths due to digestive disease. For the observation-opportunity size $A_0 = 3.0$ in terms of 10,000 person-years, the search of inaccurate death certificate information resulted in $z_0 = 18$ deaths. Comparing the errant and inerrant counts, we concluded that $y_0 = 2$ false positives and $x_0 = 16$ false negatives were contained in this sample. For the fallible observation-opportunity size A, $z = 219$ death certificates were recorded with digestive disease as the cause of death. Here, the rate parameter λ is the death rate due to digestive disease per 10,000 person-years.

Table 4.2: MLEs, Estimated Standard Errors and Approximate 95% CIs for λ

	Est.	S.E	Wald	Score	PL	AIL
λ	10.30	1.44	$(7.49, 13.11)$	$(7.77, 13.09)$	$(7.79, 13.50)$	$(7.57, 13.12)$
θ	0.52	0.07	n/a	n/a	n/a	n/a
ϕ	0.62	0.43	n/a	n/a	n/a	n/a

Table 4.2 gives the double-sample-based MLE estimates for λ, θ, and ϕ, the corresponding estimated standard errors estimated from both the fallible and infallible counts, and approximate 95% Wald, score, PL and AIL CIs for λ. From Table 4.2, we see that the score CI yielded the shortest interval width, and the PL CI yielded the widest interval width. Hence, the actual value of λ maybe more likely to be contained in the PL CI. However, because the discrepancy in interval widths among the considered CIs was essentially negligible, all four CIs could be reasonable interval estimates for the rate of death of individuals who died in Connecticut in

1979 due to digestive disease per 10,000 person-years. Nevertheless, we suggest that one use the Wald CI to estimate λ when the Poisson count size and misclassification rates are moderate to high as is the case in this example. Therefore, using the Wald 95% CI, we are highly confident that the rate of death of all individuals who died in Connecticut in 1979 due to digestive disease per 10,000 person-years is contained in $(7.49, 13.11)$.

4.9 Discussion

In this paper, we noticed that in the case of small to moderate misclassification rates ($\theta = 0.15, 0.25, \phi = 0.60$) and a small ratio of infallible observation-opportunity size to fallible observation-opportunity size (A_0/A), the Wald CI displayed good overall coverage properties compared to the score CIs. However, in the scenario of moderate to high misclassification rates ($\theta = 0.15, 0.25, \phi = 0.90$) and a small ratio of infallible to fallible observation-opportunity sizes, the score CI yielded better coverage probabilities than the PL and AIL CIs. From Table 4.2 we also noticed that for a fixed ratio A_0/A, the median interval widths of all four intervals increased as misclassification increased. Summarizing, for all parameter configurations considered here, we observed that the Wald interval maintained the best overall coverage properties for the estimation of the Poisson rate parameter λ for the case where the data counts are subject to misclassification. This conclusion was primarily attributable to the fact that the Wald CIs maintained slightly wider interval widths and, therefore, larger coverage probabilities. The simulation results determined here appear to be consistent with the findings of Riggs (2006) and Riggs et al. (2009) who have examined the Wald, score and PL CIs using various parameter and observation-opportunity size and parameter configurations in some Poisson related double-sampling paradigms.

CHAPTER FIVE

Approximate Interval Estimation of a Poisson Rate Parameter Using Data Subject to Visibility Bias

5.1 Abstract

We derive three first-order asymptotic confidence intervals (CIs) for estimating a Poisson rate parameter using a double sample of both an inerrant count and a count containing false-negative observations. In particular, we obtain a Wald-based, a score-based, and a profile log-likelihood-based CI. We then compare these three CIs in terms of coverage probability and median interval width properties via a Monte Carlo simulation experiment. For the parameter configurations considered here, we conclude that the CI based on the profile log-likelihood statistic is superior. Finally, we apply our three new CIs to real data.

5.2 Introduction

Suppose one is interested in estimating the rate of gallinule nests along a certain waterway. One may find a thorough search difficult for such a large area, and if only a cursory search is undertaken, nests obstructed from view could be uncounted. Thus, a cursory search would very likely result in unregistered counts that cause an underestimate of the nest rate. Such observations are known as false-negative observations, and this type of count misclassification is known as visibility bias (Stamey, Young, and Cecchini, 2003a).

To account for visibility bias in counted data, we can implement a double-sampling method analogous to the double-sampling scheme of Tenenbein (1970), which consists of an infallible training count and a count that is under-reported. The double-sampling method combines data from a large, usually inexpensive, fallible sample with data from a smaller, more expensive, infallible sample that is used to

estimate the parameter of interest (Greer, 2008). Here, we derive three first-order asymptotic CIs for a Poisson rate parameter using the double-sampling paradigm when data is subject to visibility bias. In particular, we derive Wald, score, and profile-likelihood (PL) CIs and compare the efficacy of these three CIs for ψ in terms of coverage probability and median interval width properties via a Monte Carlo simulation. We also apply the three proposed interval estimators to a real data set that has a double sample and is subject to under-reporting bias.

Various methods and models have been proposed to estimate parameters for counted data subject to visibility bias. Anderson, Bratcher, and Kutran (1994) have considered a Bayesian approach to estimate of a Poisson rate parameter λ when the data is subject to visibility bias. Their estimation of λ and the "visibility" parameter θ were based on independent cursory searches, exhaustive searches, and prior information about λ and θ. Fader and Hardie (2000) have presented a parsimonious Bayesian model for the analysis of under-reported Poisson count data and were able to derive an analytic expression for the key marginal posterior distributions of interest. Stamey and Young (2005) have derived maximum likelihood estimators (MLEs) for a Poisson model that allows for both false-positives and false-negatives and have also provided rough guidelines for sample-size determination. However, the papers listed above have given little information regarding interval estimation for λ.

The remainder of this chapter is organized in the following manner. In Section 5.3, we present the statistical model and an under-reporting misclassification parameter in our visibility bias model. In Section 5.4, we derive the likelihood function along with the Poisson rate parameter and the misclassification parameter MLEs. In Section 5.5, we give an EM algorithm used to determine the restricted maximum likelihood estimators (RMLEs) or maximum profile-likelihood estimators needed in two of the new CIs. In Section 5.6, we derive the three large-sample CIs for the Poisson rate parameter using a double sample. In Section 5.7, we provide a simulation

study in which we compare the efficacy of the three proposed interval estimation methods in terms of coverage probability and median interval width properties. In Section 5.8, we apply the three MLE-motivated CIs assuming a visibility-bias model to a real data set and conclude with some brief comments in Section 5.9.

5.3 The Model

First, we introduce the statistical model that we utilize to construct the likelihood-based CIs for a Poisson rate parameter. Under the double-sampling paradigm and the assumption of visibility bias, we utilize two counts: a fallible count obtained from a relatively large observation-opportunity size A using a cursory search method and a training count obtained from a small observation-opportunity size A_0. In the fallible count obtained from A, there are t true occurrences, while only $z = t - x$ occurrences are actually spotted, where x represents the missed counts in the cursory search. Hence, x represents the number of false-negatives. We assume the following distributions for the unobservable variables t and x:

$$t|\lambda \sim Poisson(A\lambda)$$

and (5.1)

$$x|t, \theta \sim Binomial(t, \theta).$$

The occurrence-rate and the false-negative misclassification rate parameters are λ and θ, respectively. Using (5.1) and the fact that $t = z + x$, we determine the joint density function of the unobservable variables z and x to be

$$f\left(z, x|\lambda, \theta\right) = \frac{e^{-A\lambda}(A\lambda)^{z+x}}{(z+x)!} \begin{pmatrix} z+x \\ x \end{pmatrix} \theta^x \left(1-\theta\right)^z. \tag{5.2}$$

From (5.2), we have that $z|\lambda, \theta \sim Poisson(A[\lambda(1-\theta)])$. One can find a proof of this result in Stamey (2000).

If the rate of false-negatives θ is unknown, then we use sample data from an observation-opportunity time or size to determine the MLEs similar to the method developed by Tenenbein (1970) for the binomial model with misclassification. A cursory search and an exhaustive search are both performed on a training observation-opportunity size A_0. The exhaustive search produces a true count t_0, and the cursory search yields a count of $z = t_0 - x_0$, where x_0 is the number of false-negative occurrences. Because the actual count t_0 is distributed as a Poisson random variable, the maximum number of false-negatives is t_0 and, thus, x_0 is distributed as a binomial random variable. We then obtain the observable data random variables

$$t_0 \sim Poisson(A_0\lambda),$$

$$x_0 \sim Binomial(t_0, \theta),$$

and (5.3)

$$z \sim Poisson(A[\lambda(1-\theta)]).$$

From (5.3), we define the joint density function of the observable variables t_0, x_0, and z as

$$f\left(t_0, x_0, z|\lambda, \theta\right) = \frac{e^{-A_0\lambda}(A_0\lambda)^{t_0}}{t_0!} \begin{pmatrix} t_0 \\ x_0 \end{pmatrix} \theta^{x_0}\left(1-\theta\right)^{t_0-x_0} \times \frac{[\lambda(1-\theta)]^z e^{-A[\lambda(1-\theta)]}}{z!}. \tag{5.4}$$

5.4 The Full-Data Likelihood and Maximum-Likelihood Estimators

Let $\mathbf{\Psi} \equiv (\lambda, \theta)'$ represent the parameter vector, where λ is the Poisson rate parameter of interest and θ is the misclassification parameter, and let $\mathbf{d} \equiv (z, t_0, x_0)'$ denote the observed data. Also, let $L_o(\lambda, \theta|\mathbf{d})$ denote the observed-data likelihood function, and let ℓ_o represent the observed data log-likelihood function. From (5.4), we have

$$L_o(\lambda, \theta|\mathbf{d}) \propto \lambda^{t_0} e^{-\lambda A_0} \theta^{x_0} (1-\theta)^{t_0-x_0} [\lambda(1-\theta)]^z e^{-A[\lambda(1-\theta)]}. \tag{5.5}$$

Therefore,

$$\ell_o \propto t_0 \ln(\lambda) + x_0 \ln(\theta) + (t_0 - x_0)\ln(1-\theta) + z\ln[\lambda(1-\theta)] + (-A_0\lambda - A[\lambda(1-\theta)]). \tag{5.6}$$

Let $\hat{\boldsymbol{\Psi}} = (\hat{\lambda}, \hat{\theta})'$ represent the unrestricted MLEs. Stamey et al. (2003a) have shown that the MLEs for λ and θ are

$$\hat{\lambda} = \frac{A_0(t_0 + z) + Ax_0}{A_0(A + A_0)}$$

and (5.7)

$$\hat{\theta} = \frac{(A + A_0)x_0}{A_0(t_0 + z) + Ax_0}.$$

Note that $\hat{\lambda}$ can be rewritten as

$$\hat{\lambda} = \alpha_1 x_0/A_0 + \alpha_2(t_0 + z)/A_0, \tag{5.8}$$

where $\alpha_1 = A/(A_0 + A)$, and $\alpha_1 + \alpha_2 = 1$. One can see from (5.8) that $\hat{\lambda}$ is a weighted average of the number of observations missed in the small area x_0 and the total number observed in both samples, $(t_0 + z)$. Thus, the first component of (5.8) simply "adds back" the number of observations missed proportionally to the size of A (Stamey et al. (2003a)).

5.5 Restricted Maximum-Likelihood Estimation

We utilize the unrestricted MLEs in (5.7) along with the restricted maximum-likelihood estimators (RMLEs) of the nuisance parameters to derive the score and the PL CIs. We implement an EM algorithm to determine the RMLEs given λ because there appears to be no closed form for the RMLEs. If we let $\boldsymbol{d}^c \equiv (z, x, t_0, x_0)'$ denote the complete data, the EM algorithm steps, detailed in Appendix 4.1.1, use the complete-data likelihood function $L^c(\lambda, \theta|\boldsymbol{d}^c)$, which is expressed in Appendix D.1.1.

5.6 Three First-Order Asymptotic Confidence Intervals for Estimating λ

In this section, we obtain three approximate $100(1-\alpha)\%$ CIs for λ by inverting the appropriate Wald, score, and profile log-likelihood statistics. Note that all three interval estimators are based on first-order asymptotic approximations.

We first obtain the observed Fisher information matrix for $\boldsymbol{\Psi} = (\lambda, \theta)'$ which is used to construct Wald and score statistics that yield CIs for the Poisson rate parameter λ. The observed Fisher information matrix is

$$\begin{aligned} \mathbf{I}(\Psi) &\equiv -E\left[\frac{\partial^2 \ell_o}{\partial \Psi \partial \Psi'}\right] \\ &= -\begin{bmatrix} -\frac{t_0+z}{\lambda^2} & A \\ A & -\frac{x_0+\theta(-2x_0+\theta(t_0+z))}{(\theta-1)^2\theta^2} \end{bmatrix}. \end{aligned} \tag{5.9}$$

We partition (5.9) so that

$$\mathbf{I}(\Psi) \equiv \begin{bmatrix} I_{11} & \boldsymbol{I}_{12} \\ \boldsymbol{I}_{21} & \boldsymbol{I}_{22} \end{bmatrix} \tag{5.10}$$

and remark that we use (5.10) in the construction of the score and PL CIs.

5.6.1 A Wald Confidence Interval for λ

We first derive the Wald CI for λ by inverting the appropriate Wald statistic. Here, the Wald statistic for λ with nuisance parameter θ is

$$W(\lambda) \ = (\hat{\lambda} - \lambda)^2 [I^{11}(\hat{\lambda}, \hat{\theta})]^{-1},$$

where $I^{11} = (I_{11} - \boldsymbol{I}_{12}\boldsymbol{I}_{22}^{-1}\boldsymbol{I}_{21})^{-1}$ (refer to Pawitan (2001)). As $A, A_0 \to \infty$ then $W(\lambda) \xrightarrow{d} \chi_1^2$. Hence, an approximate $100(1-\alpha)\%$ Wald CI for the Poisson rate parameter consists of all the values of λ satisfying

$$(\hat{\lambda} - \lambda)^2 [I^{11}(\hat{\lambda}, \hat{\theta})]^{-1} < \chi_1^2(1-\alpha), \tag{5.11}$$

where $\chi^2_1(1-\alpha)$ denotes the $(1-\alpha)$ quantile of a central chi-squared distribution with one degree of freedom. Thus, solving (5.11) for λ, we have that

$$\hat{\lambda} - \sqrt{\chi^2_1(1-\alpha)[I^{11}(\hat{\lambda},\hat{\theta})]} < \lambda < \hat{\lambda} + \sqrt{\chi^2_1(1-\alpha)[I^{11}(\hat{\lambda},\hat{\theta})]} \tag{5.12}$$

is an approximate $(1-\alpha)\%$ Wald CI for λ.

5.6.2 A Score Confidence Interval for λ

Next, we describe a score-based CI for λ that involves inverting the score statistic

$$Sc = \left[S(\hat{\mathbf{\Psi}}_\lambda)\right]^2 \left[I^{11}(\hat{\mathbf{\Psi}}_\lambda)\right] \dot{\sim} \chi^2_p,$$

where $\hat{\mathbf{\Psi}} \equiv (\hat{\lambda},\hat{\theta})'$ and $S(\hat{\mathbf{\Psi}}_\lambda) \equiv \frac{\partial \ell}{\partial \lambda}$, where ℓ is given in (5.6). An approximate $100(1-\alpha)\%$ score CI is composed of the values of λ that satisfy

$$\left[S(\hat{\mathbf{\Psi}}_\lambda)\right]^2 \left[I^{11}(\hat{\mathbf{\Psi}}_\lambda)\right] < \chi^2_1(1-\alpha). \tag{5.13}$$

We determine the interval (5.13) numerically.

5.6.3 A Profile Likelihood Confidence Interval for λ

One can determine a profile log-likelihood CI for λ by inverting the profile log-likelihood statistic. For sufficiently large n,

$$-2\left[\ell(\hat{\mathbf{\Psi}}) - \ell(\hat{\mathbf{\Psi}}_\lambda)\right] \dot{\sim} \chi^2_1.$$

Therefore, an approximate $100(1-\alpha)\%$ CI consists of all values of λ that satisfy

$$-2\left[\ell(\hat{\mathbf{\Psi}}) - \ell(\hat{\mathbf{\Psi}}_\lambda)\right] < \chi^2_1(1-\alpha). \tag{5.14}$$

The CI given in (5.14) must be determined numerically.

5.7 A Monte Carlo Simulation

Here, using a Monte Carlo simulation, we examine coverage and interval-width properties of the three interval estimators for λ described in (5.12), (5.13), and (5.14).

We fix $\lambda = 20$ and examine the effects of varying observation-opportunity sizes A_0, A, and false negative rates ($\theta = 0.4$ - moderate, $\theta = 0.7$ - moderate to high). We studied these intervals for the two different parameter configurations shown in Table 5.1. For each parameter configuration, we generated 10,000 data sets and calculated

Table 5.1: Parameter Configurations for CIs with $\lambda = 20$

Config	θ	A	A_0
Co_1	0.4	$2, 10, 20, 2, 10, 20, 2, 10, 20$	$.25, .25, .25, .5, .5, .5, 1, 1, 1$
Co_2	0.7	$2, 10, 20, 2, 10, 20, 2, 10, 20$	$.25, .25, .25, .5, .5, .5, 1, 1, 1$

approximate 95% Wald, score, and PL CIs for the Poisson rate parameter λ using a double sample of an inerrant count and an under-counted count.

5.7.1 Simulation Results

For all parameter and sample-size configurations considered in Co_1 and Co_2, we see from Figures 5.2 and 5.3 that the median interval widths of the score CIs were the shortest while the median interval widths of the PL CIs were the widest. However, as the observation-opportunity sizes A_0 and A increased, the discrepancy in median interval widths among the Wald, score, and PL CIs significantly decreased as did the Wald, score, and PL CI coverage probabilities (see Figures 5.1 and 5.4).

For a moderate false-negative rate ($\theta = 0.4$), we remark from Figure 5.1 that the Wald and score CIs underestimated the nominal 95% confidence level for all observation-opportunity sizes considered in Co_1. For observation-opportunity size $A_0 = 0.25$, we noticed that the PL CIs also somewhat underestimated the nominal 95% confidence level. However, for $A_0 = 0.5, 1$, the PL CI maintained almost nominal coverage. This coverage property behavior could be attributable to the fact that as the ratio of error-prone data to error-free data (A/A_0) decreases, the better each CI will cover the parameter λ.

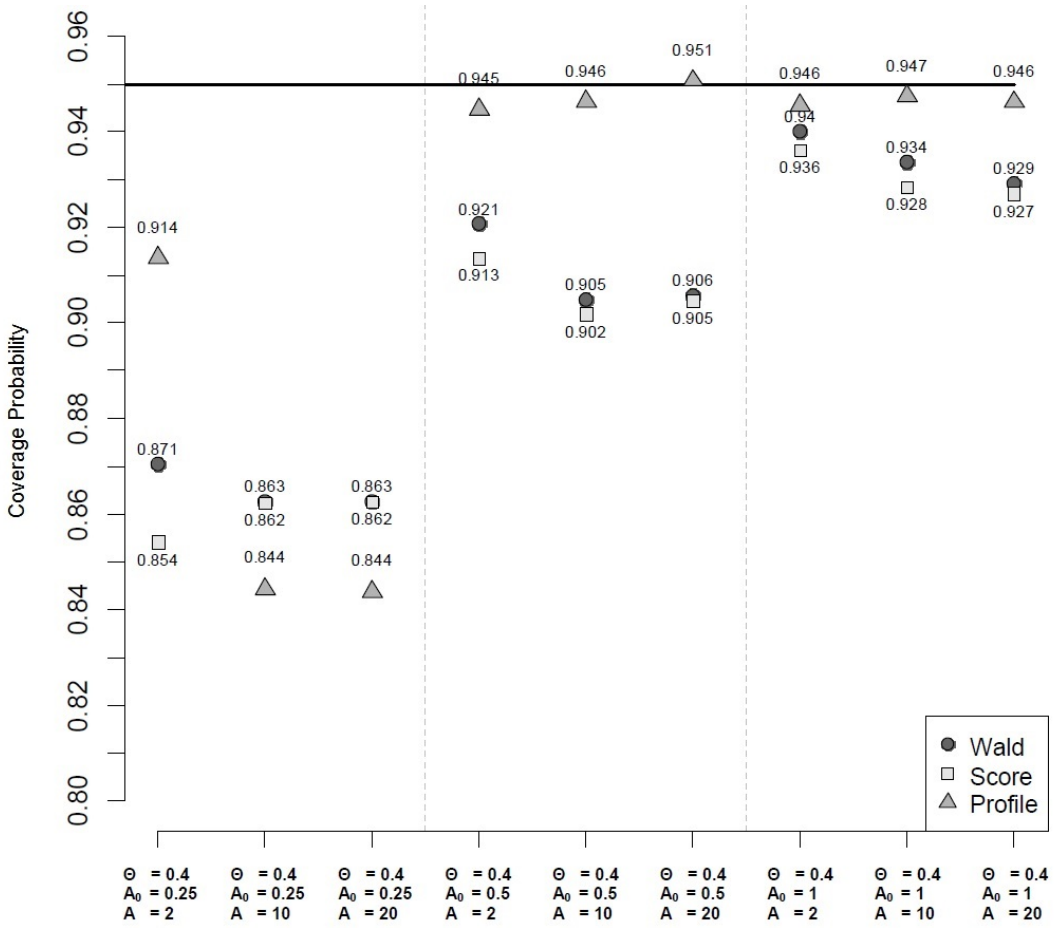

Figure 5.1: Coverage rates for λ under parameter configuration Co_1.

We observed from Figures 5.2 and 5.3 that the differences among the median interval widths of all three CIs when $A_0 = 1$ are negligible. Hence, the coverage probabilities of the Wald and score intervals for parameter configurations Co_1 and Co_2 when $A_0 = 1$ were very similar (see Figures 5.1 and 5.4). Comparing Figures 5.2 and 5.3, we noticed that as the rate of false-negative observations increased, the median interval widths decreased and the coverage probabilities improved toward the nominal 95% confidence level. Also, we noticed from Figures 5.1 and 5.4 that as the observation-opportunity size ($A_0 = 0.25$) increased by more than 50%, the Wald and score CI coverage probabilities greatly improved.

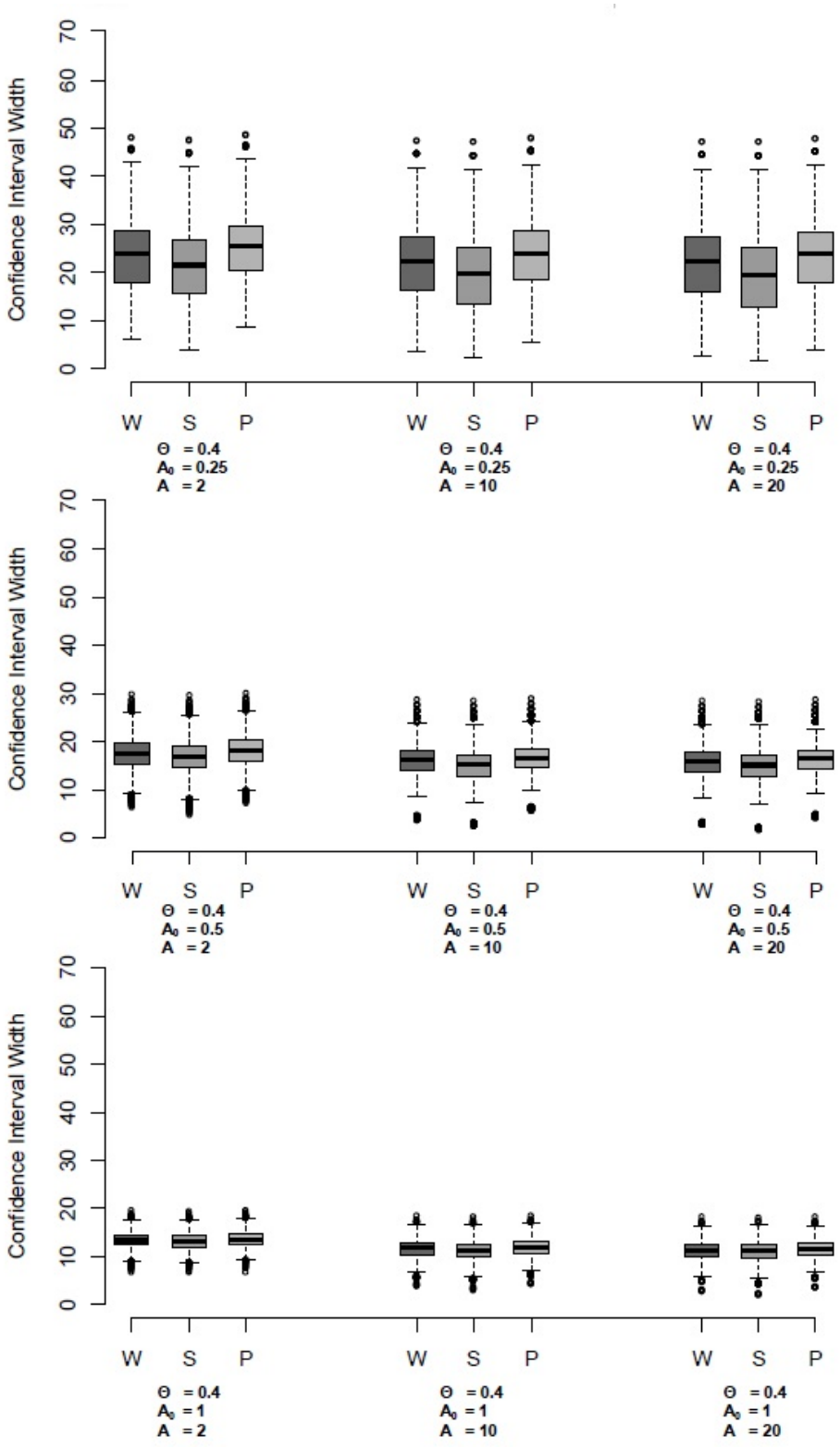

Figure 5.2: Confidence interval widths for λ under parameter configuration Co_1.

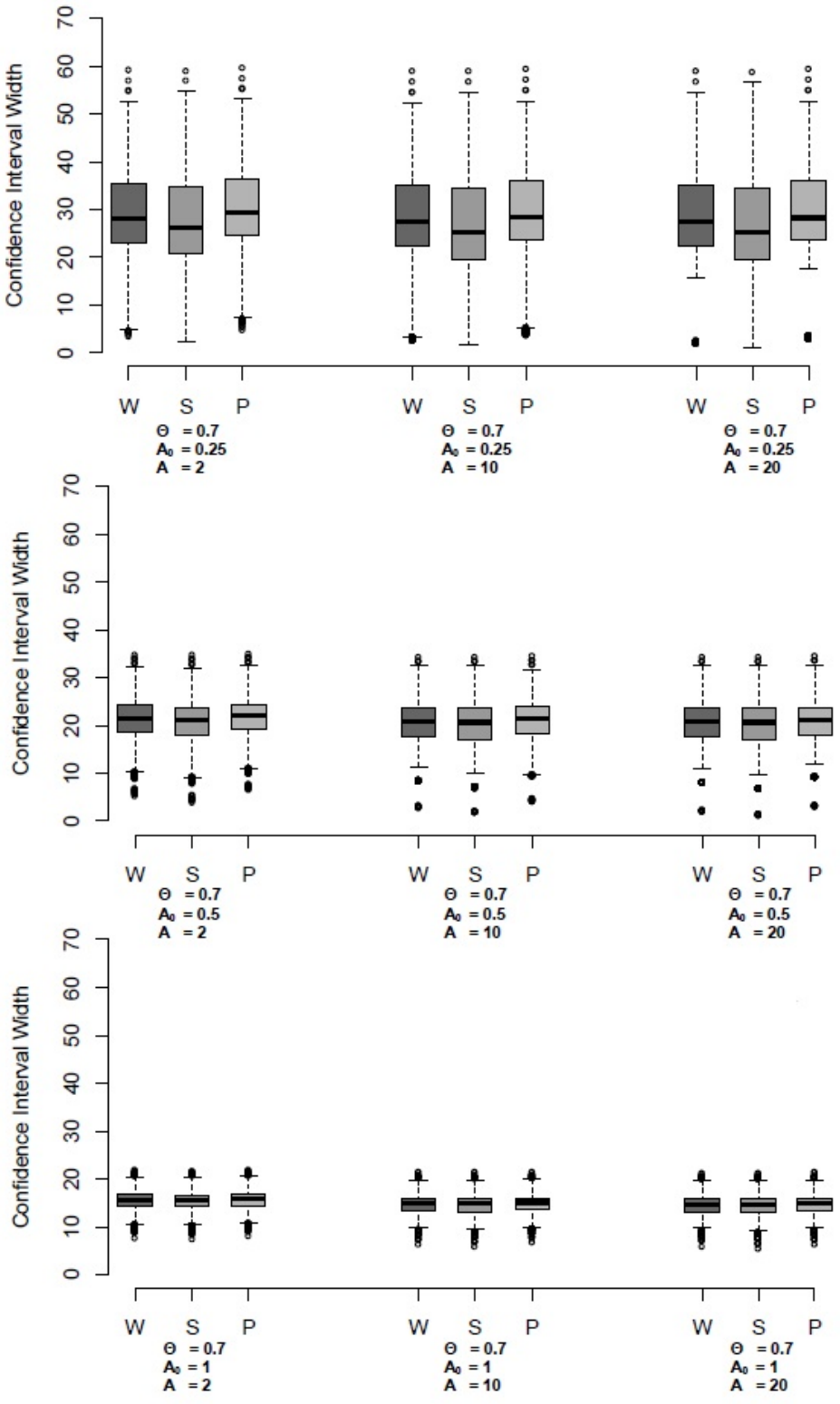

Figure 5.3: Confidence interval widths for λ under parameter configuration Co_2.

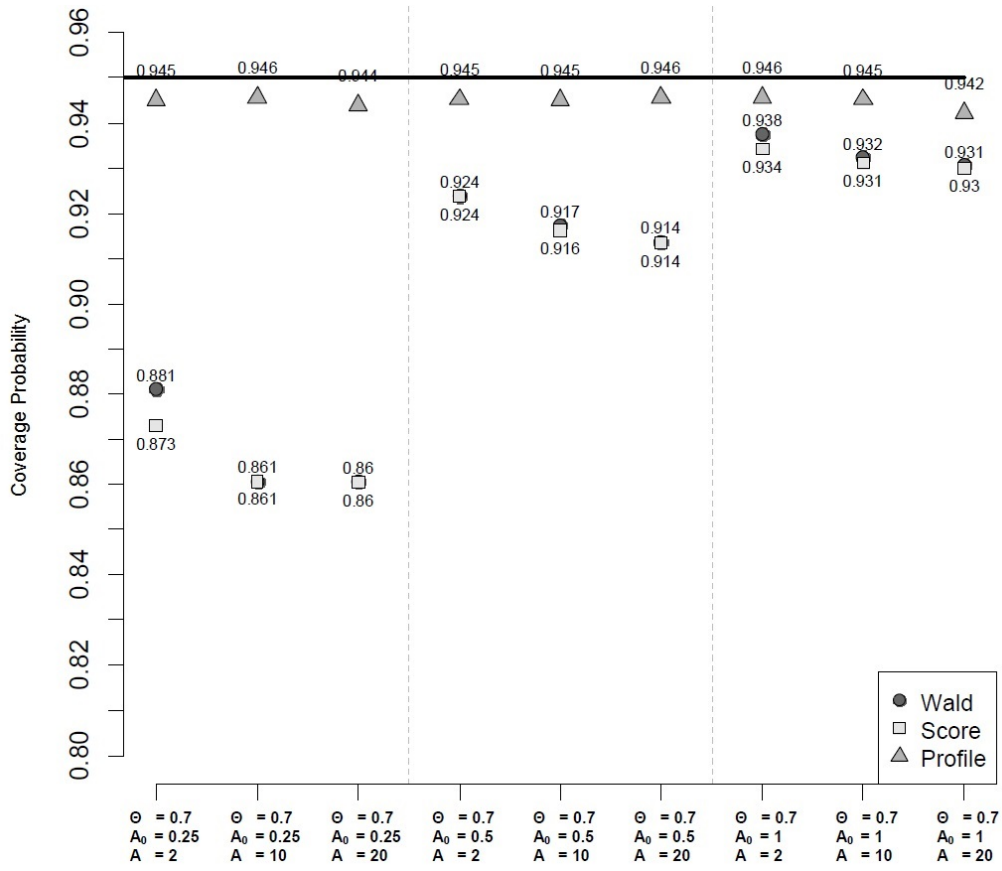

Figure 5.4: Coverage rates for λ under parameter configuration Co_2.

Under the parameter configuration Co_2, for the case of a moderate to high false-negative rate ($\theta = 0.7$), we noted from Figure 5.4 that while the Wald and score CIs considerably under-covered λ, the PL CI yielded the best overall coverage properties for the estimation of Poisson rate λ. Overall, for a fixed ratio of A_0/A, the median interval widths of all three CIs increased as the rate of false-negative observations increased, which is the behavior one would expect.

5.8 A Real-Data Example

Here, we apply the proposed double-sample Poisson rate-estimation CI with data subject to visibility bias to a real-data problem. We analyze data from Anderson et al. (1994) for which the parameter of interest is the rate of gallinule nests along

the water of Lacassine National Wildlife Refuge in southern Louisiana. A cursory (fallible) search along the waterway is conducted along with a thorough (infallible) search by air-boat over a smaller area. The fallible search is over the area $A = 4300$ linear feet, and the infallible search is applied over the smaller area $A_0 = 500$ linear feet. Using the thorough search, researchers spotted nine nests, five of which were missed in using a cursory search over the same area.

For the larger area over which only a cursory search was applied, 21 nests were spotted. Using (5.7) and (5.4), one obtains the resulting Poisson rate estimate $\hat{\lambda}$ and the false-negative rate $\hat{\theta}$ displayed in Table 5.2.

Table 5.2: MLEs, Estimated Standard Errors, and 95% Confidence Intervals for λ

	Est.	S.E.	Wald	Score	PL
$\hat{\lambda}$	15.21	4.60	$(6.21, 24.21)$	$(6.79, 24.17)$	$(8.39, 26.85)$
$\hat{\theta}$	0.34	0.11	n/a	n/a	n/a

Table 5.2 provides the double-sample-based MLEs, estimated standard errors for the two MLEs (5.7) using both the fallible and infallible data, and approximate 95% Wald, score, and PL CIs for λ. From Table 5.2, for $\theta < 0.4$ and a relatively small ratio of infallible to fallible observation-opportunity sizes ($A_0/A = 0.116$), we observed that the score CI yielded the shortest interval width and the PL CI yielded the widest interval width. Hence, using the PL 95% CI in Table 5.2, we are highly confident that the rate of gallinule nests along the water of Lacassine National Wildlife Refuge per 1000 linear feet is in the interval (8.39, 26.85).

5.9 Comments

In this chapter, we have derived three CIs for a Poisson rate parameter using under-reported data and utilizing likelihood and pseudo-likelihood methods to account for a nuisance parameter in the model. The Wald, score, and PL Poisson rate CI estimators are based on a first-order asymptotic approximation. To gain in-

formation concerning the under-reporting rate, we have utilized a double-sampling method in which we collect training counts from a small observation-opportunity size A_0 in addition to a fallible count from a large observation-opportunity size A. Through a simulation analysis, we have found that the profile likelihood approach yielded better coverage properties than the Wald and score approaches for the parameter configurations considered here. Hence, we suggest that one use the PL CI to estimate λ when the ratio A_0/A is relatively small, as in our example application.

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 for ψ, π_i, S_i, and C_i, we first re-parameterize using the transformations

$$\begin{aligned}
\alpha_i &= \pi_i S_i + (1-\pi_i)(1-C_i), \\
\beta_i &= \frac{\pi_i S_i}{\alpha_i}, \\
\gamma_i &= \frac{\pi_i(1-S_i)}{1-\alpha_i}, \\
1-\beta_i &= \frac{(1-\pi_i)(1-C_i)}{\alpha_i},
\end{aligned}$$

and

$$1-\gamma_i = \frac{(1-\pi_i)C_i}{1-\alpha_i},$$

where $i = 0, 1$, indicates the disease group of the participant.

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

The concentrated observed-data likelihood function is

$$\begin{aligned}
L_o \propto\ & [\pi_1 S_1 + (1-\pi_1)(1-C_1)]^{W_{11}} [1-\pi_1 S_1 - (1-\pi_1)(1-C_1)]^{N_1 - W_{11}} \\
& \times [\pi_0 S_0 + (1-\pi_0)(1-C_0)]^{W_{01}} [1-\pi_0 S_0 - (1-\pi_0)(1-C_0)]^{N_0 - W_{01}} \\
& \times (\pi_1 S_1)^{V_{111}} [\pi_1(1-S_1)]^{V_{110}} [(1-\pi_1)(1-C_1)]^{V_{011}} \\
& \times (\pi_0 S_0)^{V_{101}} [\pi_0(1-S_0)]^{V_{100}} [(1-\pi_0)(1-C_0)]^{V_{001}} \\
& \times [(1-\pi_1)C_1]^{[M_1 - (V_{111}+V_{110}+V_{011})]} \times [(1-\pi_0)C_0]^{[M_0 - (V_{101}+V_{100}+V_{001})]}.
\end{aligned} \tag{A.1.1}$$

For ease of derivation, we consider each part of the likelihood that comes from each disease status group ($D = 1$ or $D = 0$) separately. Rewriting the observed-data likelihood in (A.1.1) in terms of α_i, β_i, and $\gamma_i, i = 0, 1$, we have the transformed concentrated likelihood

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

so that the concentrated observed-data log-likelihood is

$$\begin{aligned} \ell_\psi \propto\ & (W_{11}+V_{111}+V_{011})\ln(\alpha_1) + (W_{10}+V_{110}+V_{010})\ln(1-\alpha_1) \\ & + V_{111}\ln(\beta_1) + V_{011}\ln(1-\beta_1) + V_{110}\ln(\gamma_1) + V_{010}\ln(1-\gamma_1) \\ & + (W_{01}+V_{101}+V_{001})\ln(\alpha_0) + (W_{00}+V_{100}+V_{000})\ln(1-\alpha_0) \\ & + V_{101}\ln(\beta_0) + V_{001}\ln(1-\beta_0) + V_{100}\ln(\gamma_0) + V_{000}\ln(1-\gamma_0). \end{aligned} \tag{A.1.2}$$

Thus, the estimating equations are

$$\begin{aligned} \frac{\partial \ln L_o}{\partial \alpha_i} &= \frac{W_{i1}+V_{1i1}+V_{0i1}}{\alpha_i} - \frac{W_{i0}+V_{1i0}+V_{0i0}}{1-\alpha_i} = 0, \\ \frac{\partial \ln L_o}{\partial \beta_i} &= \frac{V_{1i1}}{\beta_i} - \frac{V_{0i1}}{1-\beta_i} = 0 \end{aligned}$$

and

$$\frac{\partial \ln L_o}{\partial \gamma_i} = \frac{V_{1i0}}{\gamma_i} - \frac{V_{0i0}}{1-\gamma_i} = 0,$$

where $i = 0, 1$. Expression (2.7) is obtained by using the invariance property of MLEs and the observed-data MLEs for α_i, β_i, and $\gamma_i, i = 0, 1$. Thus, we see that

$$\begin{aligned} \hat{\alpha}_i &= \frac{W_{i1}+V_{1i1}+V_{0i1}}{W_{i1}+V_{1i1}+V_{0i1}+W_{i0}+V_{1i0}+V_{0i0}}, \\ \hat{\beta}_i &= \frac{V_{1i1}}{V_{1i1}+V_{0i1}}, \end{aligned}$$

and

$$\hat{\gamma}_i = \frac{V_{1i0}}{V_{1i0}+V_{0i0}}.$$

Let $L_U \equiv L_U(\pi_1, \pi_0, C_1, C_0, S_1, S_0 | \boldsymbol{d}^{full})$ be the full likelihood function in terms of the complete observed and unobserved data. Then,

$$L_U = \prod_{i=0}^{1} \begin{pmatrix} U_{1i1} + U_{1i0} \\ U_{1i1} \end{pmatrix} (\pi_i S_i)^{U_{1i1}} [\pi_i (1 - S_i)]^{U_{1i0}}$$

$$\times \begin{pmatrix} N_i \\ U_{1i1} + U_{1i0} \end{pmatrix} (\pi_i)^{U_{1i1}+U_{1i0}} (1 - \pi_i)^{N_i - (U_{1i1}+U_{1i0})}$$

$$\times \begin{pmatrix} N_i - (U_{1i1} + U_{1i0}) \\ W_{i0} - U_{1i0} \end{pmatrix} [(1 - \pi_i) C_i]^{W_{i0} - U_{1i0}} [(1 - \pi_i)(1 - C_i)]^{W_{i1} - U_{1i1}}$$

$$\times \left(\frac{4!}{V_{1i1}! V_{1i0}! V_{0i1}! V_{0i0}!} \right) (\pi_i S_i)^{V_{1i1}} [\pi_i (1 - S_i)]^{V_{1i0}} [(1 - \pi_i)(1 - C_i)]^{V_{0i1}} [(1 - \pi_i) C_i]^{V_{0i0}},$$

where $i = 0, 1$, indicates the true disease status of an individual in the study (0 indicates diseased, 1 indicates non-diseased). We then re-express L_U as

$$L_U \propto \prod_{i=0}^{1} \begin{pmatrix} N_i \\ U_{1i1} + U_{1i0} \end{pmatrix} \begin{pmatrix} U_{1i1} + U_{1i0} \\ U_{1i1} \end{pmatrix} \begin{pmatrix} N_i - (U_{1i1} + U_{1i0}) \\ W_{i0} - U_{1i0} \end{pmatrix}$$

$$\times (\pi_i S_i)^{U_{1i1}} [\pi_i (1 - S_i)]^{U_{1i0}} \times [(1 - \pi_i)(1 - C_i)]^{W_{i1} - U_{1i1}} [(1 - \pi_i)(C_i)]^{W_{i0} - U_{1i0}}$$

$$\times (\pi_i S_i)^{V_{1i1}} [\pi_i (1 - S_i)]^{V_{1i0}} [(1 - \pi_i)(1 - C_i)]^{V_{0i1}} \times [(1 - \pi_i) C_i]^{[M_i - (V_{1i1} + V_{1i0} + V_{0i1})]}. \tag{A.2.1}$$

Let

$$\pi_0 = \frac{\pi_1}{\pi_1 - \pi_1 \psi + \psi}, \tag{A.2.2}$$

and let $f_i(V_{1i1}, V_{1i0}, V_{0i1}, V_{0i0})$, i = 0, 1, be the multinomial probability functions for the complete data. Then substituting the right-hand side of (A.2.2) for π_0 into

(A.2.1), we get

$$
\begin{aligned}
L_U \propto & \prod_{i=0}^{1} \frac{W_{i1}!}{U_{1i1}!(W_{i1}-U_{1i1})!} \frac{W_{i0}!}{U_{1i0}!(W_{i0}-U_{1i0})!} \\
& \times \left[\frac{\pi_1 S_1}{\pi_1 S_1 + (1-\pi_1)(1-C_1)}\right]^{U_{111}} \left[\frac{(1-\pi_1)(1-C_1)}{\pi_1 S_1 + (1-\pi_1)(1-C_1)}\right]^{W_{11}-U_{111}} \\
& \times \left[\frac{\pi_1 S_0}{\pi_1 S_0 + \psi(1-\pi_1)(1-C_0)}\right]^{U_{101}} \left[\frac{\psi(1-\pi_1)(1-C_0)}{\pi_1 S_0 + \psi(1-\pi_1)(1-C_0)}\right]^{W_{01}-U_{101}} \\
& \times \left[\frac{\pi_1 (1-S_1)}{\pi_1 (1-S_1) + (1-\pi_1) C_1}\right]^{U_{110}} \left[\frac{(1-\pi_1) C_1}{\pi_1 (1-S_1) + (1-\pi_1) C_1}\right]^{W_{10}-U_{110}} \\
& \times \left[\frac{\pi_1 (1-S_0)}{\pi_1 (1-S_0) + \psi(1-\pi_1) C_0}\right]^{U_{100}} \left[\frac{\psi(1-\pi_1) C_0}{\pi_1 (1-S_0) + \psi(1-\pi_1) C_0}\right]^{W_{00}-U_{100}} \\
& \times \prod_{i=0,1} f_i(V_{1i1}, V_{1i0}, V_{0i1}, V_{0i0}).
\end{aligned}
\tag{A.2.3}
$$

Thus, the latent variables $U_{111}, U_{110}, U_{101}$, and U_{100} are conditionally binomially distributed.

1.2.1 An EM Algorithm

E-step: Let $\mathbf{\Phi}^{(r)} \equiv (\psi, \pi_1^{(r)}, S_0^{(r)}, S_1^{(r)}, C_0^{(r)}, C_1^{(r)})'$ be the current parameter vector at the r^{th} iteration. We first determine that the conditional expectations of the four unobserved variables are

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

$$
U_{110}^* \equiv E[U_{110}|\boldsymbol{d}, \mathbf{\Phi}^{(r)}] = \frac{W_{10}\pi_1^{(r)} (1-S_1^{(r)})}{\pi_1^{(r)} (1-S_1^{(r)}) + (1-\pi_1^{(r)}) C_1^{(r)}},
$$

$$
U_{101}^* \equiv E[U_{101}|\boldsymbol{d}, \mathbf{\Phi}^{(r)}] = \frac{W_{01}\pi_1^{(r)} S_0^{(r)}}{\pi_1^{(r)} S_0^{(r)} + \psi(1-\pi_1^{(r)})(1-C_0^{(r)})},
$$

and

$$
U_{100}^* \equiv E[U_{100}|\boldsymbol{d}, \mathbf{\Phi}^{(r)}] = \frac{W_{00}\pi_1^{(r)} (S_0^{(r)}-1)}{\pi_1^{(r)} (1-S_0^{(r)}) + \psi(1-\pi_1^{(r)}) C_0^{(r)}}.
$$

M-step: Recall that the conditional likelihood function L_U is given in (A.2.3). The full-data estimating equations are

$$\begin{aligned}
\frac{\partial \ell_U}{\partial \pi_1} &= -\frac{M_0 + M_1 - U^*_{100} - U^*_{101} - U^*_{110} - U^*_{111} + W_{00} + W_{01} + W_{10} + W_{11}}{1-\pi_1} \\
&\quad - \frac{V_{100} - V_{101} - V_{110} - V_{111}}{1-\pi_1} + \frac{(M_0 + W_{00} + W_{01})(\psi - 1)}{\pi_1 - \pi_1\psi + \psi} \\
&\quad + \frac{U^*_{100} + U^*_{101} + U^*_{110} + U^*_{111} + V_{100} + V_{101} + V_{110} + V_{111}}{\pi_1} = 0, \\
\frac{\partial \ell_U}{\partial S_0} &= \frac{V_{101} + U^*_{101}}{S_0} - \frac{V_{100} + U^*_{100}}{1 - S_0} = 0, \\
\frac{\partial \ell_U}{\partial S_1} &= \frac{V_{111} + U^*_{111}}{S_1} - \frac{V_{110} + U^*_{110}}{1 - S_1} = 0, \\
\frac{\partial \ell_U}{\partial C_0} &= \frac{W_{00} + V_{000} - U^*_{100}}{C_0} - \frac{W_{01} - U^*_{101} + M_0 - V_{101} - V_{100} - V_{000}}{1 - C_0} = 0,
\end{aligned}$$

and

$$\frac{\partial \ell_U}{\partial C_1} = \frac{W_{10} + V_{010} - U^*_{110}}{C_1} - \frac{W_{11} - U^*_{111} + M_1 - V_{111} - V_{110} - V_{010}}{1 - C_1} = 0.$$

Solving these estimating equations for the respective nuisance parameters π_1, S_i, and $C_i, i = 0, 1$, we obtain their complete-data MLEs in terms of ψ. We then update the current parameter estimates $\pi_1^{(r)}$, $S_i^{(r)}$, and $C_i^{(r)}$, $i = 0, 1$, using

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

where

$$\begin{aligned}
A &= (\psi - 1)(M_1 + W_{10} + W_{11}), \\
B &= M_0 + W_{00} + W_{01} + \psi(M_1 + W_{10} + W_{11}) \\
&\quad + (\psi - 1)(V_{100} + V_{101} + V_{110} + V_{111} + U^*_{100} + U^*_{101} + U^*_{110} + U^*_{111}),
\end{aligned}$$

and

$$C = \psi(V_{100} + V_{101} + V_{110} + V_{111} + U^*_{100} + U^*_{101} + U^*_{110} + U^*_{111}).$$

Also,

$$C_i^{(r+1)} = \frac{W_{i0} - U^*_{1i0} + V_{0i0}}{W_{i0} + W_{i1} - U^*_{1i0} - U^*_{1i1} + M_i - V_{1i1} - V_{1i0}},$$

and

$$S_i^{(r+1)} = \frac{V_{1i1} + U^*_{1i1}}{V_{1i0} + V_{1i1} + U^*_{1i0} + U^*_{1i1}}.$$

A.3 The Observed Fisher Information Matrix

To calculate the Wald, score, and AIL CIs, we derive each of the elements on and above the diagonal of the OFIM $\mathbf{J}(\psi, \boldsymbol{\eta})$ defined in (2.8). We have

$$\begin{aligned}\frac{\partial^2 \ell_\psi}{\partial \psi^2} = \Bigg[\; & (\pi_1 - 1)^2 (V_{100} + V_{101}) - \frac{\pi_1^2 (V_{000} + V_{001})}{\psi^2} + \frac{2(\pi_1 - 1)\pi_1 (V_{001} + V_{000})}{\psi} \\ & + \frac{2\pi_1 (\pi_1 - 1)^2 (S_0 + C_0 - 1)(N_0 - W_{01})}{-C_0\psi + \pi_1 (S_0 + C_0\psi - 1)} - \frac{\pi_1^2 (\pi_1 - 1)^2 (S_0 + C_0 - 1)^2 (N_0 - W_{01})}{[C_0\psi - \pi_1 (S_0 + C_0\psi - 1)]^2} \\ & + \frac{2\pi_1 (\pi_1 - 1)^2 (S_0 + C_0 - 1) W_{01}}{(C_0 - 1)(\pi_1 - 1)\psi + \pi_1 S_0} - \frac{\pi_1^2 (\pi_1 - 1)^2 (S_0 + C_0 - 1) W_{01}}{[(C_0 - 1)(\pi_1 - 1)\psi + \pi_1 S_0]^2} \; \Bigg] \\ & \times \frac{1}{(\pi_1 + \psi - \pi_1\psi)^2},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell_\psi}{\partial \psi \partial \pi_1} = & \frac{-\pi_1 (V_{001} + V_{000})}{\psi(\pi_1 - 1)(-\psi - \pi_1 + \psi\pi_1)} - \frac{\pi_1 (V_{001} + V_{000})}{\psi(\pi_1 - 1)(\psi + \pi_1 - \psi\pi_1)^2} \\ & + \frac{(V_{101} + V_{100})(\pi_1 - 1)(\psi - 1)}{(\psi + \pi_1 - \psi\pi_1)^2} + \frac{V_{101} + V_{100}}{\psi + \pi_1 - \psi\pi_1} \\ & + \frac{2\pi_1 (V_{001} + V_{000})(\psi - 1)}{\psi(\psi + \pi_1 - \psi\pi_1)^2} + \frac{(V_{001} + V_{000})}{\psi(\psi + \pi_1 - \psi\pi_1)} \\ & + \frac{(-\psi + \pi_1 + \psi\pi_1)(S_0 + C_0 - 1)(N_0 - W_{01})}{(\psi + \pi_1 - \psi\pi_1)^2 (C_0\psi(\pi_1 - 1) + \pi_1 (S_0 - 1))} \\ & - \frac{\psi\pi_1 (\pi_1 - 1)(S_0 + C_0 - 1)^2 (N_0 - W_{01})}{(\psi + \pi_1 - \psi\pi_1)^2 (C_0\psi(1 - \pi_1) + \pi_1 (1 - S_0))^2} \\ & - \frac{\psi\pi_1 (\pi_1 - 1)(S_0 + C_0 - 1)^2 W_{01}}{(\psi + \pi_1 - \psi\pi_1)^2 ((C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0)^2} \\ & + \frac{(\psi(\pi_1 - 1) + \pi_1)(S_0 + C_0 - 1) W_{01}}{(\psi + \pi_1 - \psi\pi_1)^2 ((C_0 - 1)\psi(\pi_1 - 1) + \pi_1 S_0)}\end{aligned}$$

$$\frac{\partial^2 \ell_\psi}{\partial \psi \partial S_1} = \frac{\partial^2 \ell_\psi}{\partial \psi \partial C_1} = \frac{\partial^2 \ell_\psi}{\partial S_0 \partial S_1} = \frac{\partial^2 \ell_\psi}{\partial S_0 \partial C_1} = \frac{\partial^2 \ell_\psi}{\partial S_1 \partial C_0} = \frac{\partial^2 \ell_\psi}{\partial C_0 \partial C_1} = 0,$$

$$\frac{\partial^2 \ell_\psi}{\partial \psi \partial S_0} = \Big[\; C_0 N_0[(C_0-1)(\pi_1-1)\psi + \pi_1 S_0]^2$$
$$+ W_{01}(C_0-1)[\pi_1^2 - 2C_0(\pi_1-1)\pi_1\psi + C_0(\pi_1-1)^2\psi^2]$$
$$+ 2W_{01}(C_0-1)\pi_1^2 S_0 - W_{01}\pi_1^2 S_0^2 \quad \Big]$$
$$\times \frac{-\pi_1(\pi_1-1)}{[C_0\psi(1-\pi_1)+\pi_1(1-S_0)]^2[(C_0-1)\psi(\pi_1-1)+\pi_1 S_0]^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \psi \partial C_0} = \Big[\; 2S_0\psi\pi_1(\pi_1-1)(S_0-1) - S_0\pi_1^2(S_0-1)$$
$$+ W_{01}\psi^2(\pi_1-1)^2\left[1 + C_0^2 + 2C_0(S_0-1) - S_0\right]$$
$$+ N_0(S_0-1)[(C_0-1)\psi(\pi_1-1)+\pi_1 S_0]^2 \quad \Big]$$
$$\times \frac{\pi_1(\pi_1-1)}{[(C_0-1)\psi(\pi_1-1)+\pi_1 S_0]^2[C_0\psi - \pi_1(S_0+C_0\psi-1)]^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \pi_1^2} = \frac{2(V_{001}+V_{000})(\psi-1)}{(\pi_1-1)(\psi+\pi_1-\psi\pi_1)^2} + \frac{\psi(V_{101}+V_{100})}{(-\psi-\pi_1+\psi\pi_1)\pi_1^2} + \frac{\psi(\psi-1)(V_{101}+V_{100})}{\pi_1(\psi+\pi_1-\psi\pi_1)^2}$$
$$- \frac{(V_{110}+V_{111})}{\pi_1^2} - \frac{(V_{010}+V_{011})}{(\pi_1-1)^2} + \frac{2\psi(\psi-1)(S_0+C_0-1)(N_0-W_{01})}{(\psi+\pi_1-\psi\pi_1)^2(C_0\psi(\pi_1-1)+\pi_1(S_0-1))}$$
$$- \frac{(V_{001}+V_{000})}{(\pi_1-1)^2(\psi+\pi_1-\psi\pi_1)^2} - \frac{(S_1+C_1-1)^2 W_{11}}{(1+C_1(\pi_1-1)+\pi_1(S_1-1))^2}$$
$$- \frac{\psi^2(S_0+C_0-1)^2(N_0-W_{01})}{(\psi+\pi_1-\psi\pi_1)^2(C_0\psi(1-\pi_1)+\pi_1(1-S_0))^2}$$
$$- \frac{\psi^2(S_0+C_0-1)^2 W_{01}}{(\psi+\pi_1-\psi\pi_1)^2((C_0-1)\psi(\pi_1-1)+\pi_1 S_0)^2}$$
$$+ \frac{2\psi(\psi-1)(S_0+C_0-1)W_{01}}{(\psi+\pi_1-\psi\pi_1)^2((C_0-1)\psi(\pi_1-1)+\pi_1 S_0)}$$
$$- \frac{(S_1+C_1-1)^2(N_1-W_{11})}{(C_1(\pi_1-1)+\pi_1(S_1-1))^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \pi_1 \partial S_0} = \Big[\quad - C_0 N_0 [(C_0-1)(\pi_1-1)\psi + \pi_1 S_0]^2$$
$$+ W_{01}(C_0-1)[-\pi_1^2 + 2C_0(\pi_1-1)\pi_1\psi - C_0(\pi_1-1)^2\psi^2]$$
$$+ 2W_{01}(C_0-1)\pi_1^2 S_0 + W_{01}\pi_1^2 S_0^2 \quad \Big]$$
$$\times \frac{\psi}{[C_0\psi(1-\pi_1) + \pi_1(1-S_0)]^2[(C_0-1)\psi(\pi_1-1) + \pi_1 S_0]^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \pi_1 \partial S_1} = \Big[\quad - C_1 N_1 (C_1 + \pi_1 - \pi_1 C_1 - \pi_1 S_1 - 1)^2 + W_{11} C_1^2 (\pi_1^2 - 1)$$
$$+ W_{11} C_1 [2\pi_1^2 (S_1 - 1) + 1] + W_{11}\pi_1^2 (S_1 - 1)^2 \quad \Big]$$
$$\times \frac{1}{\left[C_1 (\pi_1 - 1) + \pi_1 (S_1 - 1)\right]^2 \left[(C_1 - 1)(\pi_1 - 1) + \pi_1 S_1\right]^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \pi_1 \partial C_0} = \Big[\quad 2S_0\psi\pi_1(\pi_1-1)(S_0-1) - S_0\pi_1^2(S_0-1)$$
$$+ W_{01}\psi^2(\pi_1-1)^2\left[1 + C_0^2 + 2C_0(S_0-1) - S_0\right]$$
$$+ N_0(S_0-1)[(C_0-1)\psi(\pi_1-1) + \pi_1 S_0]^2 \quad \Big]$$
$$\times \frac{\psi}{[(C_0-1)\psi(\pi_1-1) + \pi_1 S_0]^2 [C_0\psi - \pi_1(S_0 + C_0\psi - 1)]^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \pi_1 \partial C_1} = \Big[\quad N_1(S_1-1)[(C_1-1)(\pi_1-1) + \pi_1 S_1]^2$$
$$+ W_{11} C_1^2 (\pi_1-1)^2 + 2W_{11} C_1 (\pi_1-1)^2 (S_1-1)$$
$$+ W_{11}(S_1-1)(\pi_1(S_1-1)(\pi_1-2) - 1) \quad \Big]$$
$$\times \frac{1}{\left[C_1 (\pi_1 - 1) + \pi_1 (S_1 - 1)\right]^2 \left[(C_1 - 1)(\pi_1 - 1) + \pi_1 S_1\right]^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial S_0^2} = -\frac{V_{101}}{S_0^2} - \frac{V_{100}}{(S_0-1)^2} + \frac{\pi_1^2(W_{01}-N_0)}{(C_0\psi - \pi_1(S_0+C_0\psi-1))^2} - \frac{\pi_1^2 W_{01}}{((C_0-1)(\pi_1-1)\psi+\pi_1 S_0)^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial S_0 \partial C_0} = \left[\frac{-(\pi_1-1)\pi_1\psi}{((C_0-1)(\pi_1-1)\psi+\pi_1 S_0)^2 (C_0\psi-\pi_1(C_0\psi+S_0-1))^2}\right] \times \left[\begin{array}{l} N_0((C_0-1)(\pi_1-1)\psi+\pi_1 S_0)^2 \\ +W_{01}(\pi_1(\psi-1)-\psi)(\psi-2C_0\psi+\pi_1(2S_0+2C_0\psi-\psi-1)) \end{array}\right],$$

$$\frac{\partial^2 \ell_\psi}{\partial S_1^2} = -\frac{V_{111}}{S_1^2} - \frac{V_{110}}{(S_1-1)^2} - \frac{\pi_1^2(N_1-W_{11})}{(C_1(\pi_1-1)+\pi_1(S_1-1))^2} - \frac{\pi_1^2 W_{11}}{(1+C_1(\pi_1-1)+\pi_1(S_1-1))^2},$$

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

$$\frac{\partial^2 \ell_\psi}{\partial C_0^2} = -\frac{V_{001}}{(C_0-1)^2} - \frac{V_{000}}{C_0^2} - \frac{(\pi_1-1)^2\psi^2(N_0-W_{01})}{(C_0\psi-\pi_1(S_0+C_0\psi-1))^2} - \frac{(\pi_1-1)^2\psi^2 W_{01}}{((C_0-1)(\pi_1-1)\psi+\pi_1 S_0)^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial C_1^2} = -\frac{V_{011}}{(C_1-1)^2} - \frac{V_{010}}{C_1^2} - \frac{(\pi_1-1)^2(N_1-W_{11})}{(C_1(\pi_1-1)+\pi_1(S_1-1))^2} - \frac{(\pi_1-1)^2 W_{11}}{((C_1-1)(\pi_1-1)+\pi_1 S_1)^2}.$$

The terms beneath the OFIM diagonal are provided by the symmetry of the OFIM.

APPENDIX B

Derivations for Chapter Three

B.1 Restricted Maximum-Likelihood Estimators

The complete-data likelihood function is

$$
\begin{aligned}
L_U = \prod_{i=0}^{1} & \begin{pmatrix} U_{1i1} + U_{1i0} \\ U_{1i1} \end{pmatrix} (\pi_i S)^{U_{1i1}} [\pi_i (1-S)]^{U_{1i0}} \\
& \times \begin{pmatrix} N_j \\ U_{1i1} + U_{1i0} \end{pmatrix} (\pi_i)^{U_{1i1}+U_{1i0}} (1-\pi_i)^{N_i - (U_{1i1}+U_{1i0})} \\
& \times \begin{pmatrix} N_i - (U_{1i1} + U_{1i0}) \\ W_{i0} - U_{1i0} \end{pmatrix} [(1-\pi_i)C]^{W_{i0}-U_{1i0}} [(1-\pi_i)(1-C)]^{W_{i1}-U_{1i1}} \\
& \times \left(\frac{4!}{V_{1i1}! V_{1i0}! V_{0i1}! V_{0i0}!} \right) (\pi_i S)^{V_{1i1}} [\pi_i (1-S)]^{V_{1i0}} [(1-\pi_i)(1-C)]^{V_{0i1}} [(1-\pi_i)C]^{V_{0i0}},
\end{aligned}
\tag{B.1.1}
$$

We next regroup the latent variables $U_{1ij}, i, j = 0, 1$, and rewrite (B.1.1) as the concentrated complete-data likelihood function

$$
\begin{aligned}
L_U \propto \prod_{i=0}^{1} & \begin{pmatrix} N_i \\ U_{1i1} + U_{1i0} \end{pmatrix} \begin{pmatrix} U_{1i1} + U_{1i0} \\ U_{1i1} \end{pmatrix} \begin{pmatrix} N_i - (U_{1i1} + U_{1i0}) \\ W_{i0} - U_{1i0} \end{pmatrix} \\
& \times (\pi_i S)^{U_{1i1}} [\pi_i (1-S)]^{U_{1i0}} \\
& \times [(1-\pi_i)(1-C)]^{W_{i1}-U_{1i1}} [(1-\pi_i)(C)]^{W_{i0}-U_{1i0}} \\
& \times (\pi_i S)^{V_{1i1}} [\pi_i (1-S)]^{V_{1i0}} [(1-\pi_i)(1-C)]^{V_{0i1}} \times [(1-\pi_i)C]^{[M_i - (V_{1i1}+V_{1i0}+V_{0i1})]}.
\end{aligned}
\tag{B.1.2}
$$

Let

$$
\pi_0 = \frac{\pi_1}{\pi_1 - \pi_1 \psi + \psi}. \tag{B.1.3}
$$

After substituting the right-hand side of (B.1.3) into (B.1.2), we get

$$
\begin{aligned}
L_U \propto & \frac{W_{11}!}{U_{111}!(W_{11}-U_{111})!}\frac{W_{01}!}{U_{101}!(W_{01}-U_{101})!}\frac{W_{00}!}{U_{100}!(W_{00}-U_{100})!}\frac{W_{00}!}{U_{100}!(W_{00}-U_{100})!} \\
& \times \left[\frac{\pi_1 S}{\pi_1 S+(1-\pi_1)(1-C)}\right]^{U_{111}}\left[\frac{(1-\pi_1)(1-C)}{\pi_1 S+(1-\pi_1)(1-C)}\right]^{W_{11}-U_{111}} \\
& \times \left[\frac{\pi_1 S}{\pi_1 S+\psi(1-\pi_1)(1-C)}\right]^{U_{101}}\left[\frac{\psi(1-\pi_1)(1-C)}{\pi_1 S+\psi(1-\pi_1)(1-C)}\right]^{W_{01}-U_{101}} \\
& \times \left[\frac{\pi_1(1-S)}{\pi_1(1-S)+(1-\pi_1)C}\right]^{U_{110}}\left[\frac{(1-\pi_1)C}{\pi_1(1-S)+(1-\pi_1)C}\right]^{W_{10}-U_{110}} \\
& \times \left[\frac{\pi_1(1-S)}{\pi_1(1-S)+\psi(1-\pi_1)C}\right]^{U_{100}}\left[\frac{\psi(1-\pi_1)C}{\pi_1(1-S)+\psi(1-\pi_1)C}\right]^{W_{00}-U_{100}} \\
& \times f_0(V_{101},V_{100},V_{001},V_{000})f_1(V_{111},V_{110},V_{011},V_{010}),
\end{aligned}
$$

where the notation $f_i(V_{1i1},V_{1i0},V_{0i1},V_{0i0})$ for $i=0,1$, represents the multinomial distributions for the two groups that compose the observed data. Thus, the full conditional distributions of the latent variables $U_{1ij}, i,j=0,1$, are distributed as binomial distributions.

2.1.1 An EM Algorithm for Estimating the RMLEs

E-step: Let $\mathbf{\Phi}^{(r)} \equiv (\psi, \pi_1^{(r)}, S^{(r)}, C^{(r)})'$ be the current parameter vector at the r^{th} iteration. We next derive the conditional expectations for the latent variables $\mathbf{U} \equiv (U_{111}, U_{110}, U_{101}, U_{100})'$, given the observable counts and current parameter values. The conditional expectations of the unobserved variables are

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

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

$$
U^*_{101} \equiv E[U_{101}|\boldsymbol{d},\mathbf{\Phi}^{(r)}] = \frac{W_{01}\pi_1^{(r)}S^{(r)}}{\pi_1^{(r)}S^{(r)}+\psi(1-\pi_1^{(r)})(1-C^{(r)})},
$$

and

$$
U^*_{100} \equiv E[U_{100}|\boldsymbol{d},\mathbf{\Phi}^{(r)}] = \frac{W_{00}\pi_1^{(r)}(S^{(r)}-1)}{\pi_1^{(r)}(1-S^{(r)})+\psi(1-\pi_1^{(r)})C^{(r)}}.
$$

M-step: We update the parameter of interest in each iteration by using the solutions to the full-data log-likelihood estimating equations

$$\begin{aligned}
\frac{\partial \ell_U}{\partial \pi_1} &= -\frac{M_0 + M_1 - U^*_{100} - U^*_{101} - U^*_{110} - U^*_{111} + W_{00} + W_{01} + W_{10} + W_{11}}{1-\pi_1} \\
&\quad - \frac{V_{100} - V_{101} - V_{110} - V_{111}}{1-\pi_1} + \frac{(M_0 + W_{00} + W_{01})(\psi - 1)}{\pi_1 - \pi_1\psi + \psi} \\
&\quad + \frac{U^*_{100} + U^*_{101} + U^*_{110} + U^*_{111} + V_{100} + V_{101} + V_{110} + V_{111}}{\pi_1} = 0, \\
\frac{\partial \ell_U}{\partial C} &= \frac{W_{00} + V_{000} - U^*_{100} + W_{10} + V_{010} - U^*_{110}}{C} \\
&\quad - \frac{W_{01} + V_{001} - U^*_{101} + W_{11} + V_{011} - U^*_{111}}{1-C} = 0,
\end{aligned}$$

and

$$\frac{\partial \ell_U}{\partial S} = \frac{V_{101} + U^*_{101} + V_{111} + U^*_{111}}{S} - \frac{V_{100} + V_{110} + U^*_{100} + U^*_{110}}{1-S} = 0. \tag{B.1.4}$$

Solving the three estimating equations in (B.1.4) for the respective elements of nuisance parameter vector $\boldsymbol{\eta}$, we have for the r^{th} iteration,

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

where

$$\begin{aligned}
A &= (\psi - 1)(M_1 + W_{10} + W_{11}), \\
B &= M_0 + W_{00} + W_{01} + \psi(M_1 + W_{10} + W_{11}) \\
&\quad + (\psi - 1)(V_{100} + V_{101} + V_{110} + V_{111} + U^*_{100} + U^*_{101} + U^*_{110} + U^*_{111}),
\end{aligned}$$

and

$$C = \psi(V_{100} + V_{101} + V_{110} + V_{111} + U^*_{100} + U^*_{101} + U^*_{110} + U^*_{111}).$$

Also, we have

$$C^{(r+1)} = \frac{W_{00} + W_{10} - U^*_{100} - U^*_{110} + V_{010} + V_{000}}{W_{00} + W_{01} + W_{10} + W_{11} - U^*_{100} - U^*_{101} - U^*_{110} - U^*_{111} + M_0 + M_1 - T_{10}},$$

where

$$T_{10} = (V_{111} + V_{100} + V_{101} + V_{110}),$$

and

$$S^{(r+1)} = \frac{V_{101} + V_{111} + U^*_{101} + U^*_{111}}{V_{100} + V_{101} + V_{110} + V_{111} + U^*_{100} + U^*_{101} + U^*_{110} + U^*_{111}}.$$

B.2 The Observed Fisher Information Matrix

Below, we give the OFIM elements used to calculate the Wald, score, and AIL CIs. The elements on and above the diagonal of $\mathbf{J}(\psi, \boldsymbol{\eta})$, given in (3.9), are

$$\begin{aligned}
\frac{\partial^2 \ell_\psi}{\partial \psi^2} = \Bigg[& (\pi_1 - 1)^2 (V_{100} + V_{101}) - \frac{\pi_1^2 (V_{000} + V_{001})}{\psi^2} + \frac{2(\pi_1 - 1)\pi_1 (V_{001} + V_{000})}{\psi} \\
& + \frac{2\pi_1 (\pi_1 - 1)^2 (S + C - 1)(N_0 - W_{01})}{-C\psi + \pi_1 (S + C\psi - 1)} - \frac{\pi_1^2 (\pi_1 - 1)^2 (S + C - 1)^2 (N_0 - W_{01})}{[C\psi - \pi_1 (S + C\psi - 1)]^2} \\
& + \frac{2\pi_1 (\pi_1 - 1)^2 (S + C - 1) W_{01}}{(C - 1)(\pi_1 - 1)\psi + \pi_1 S} - \frac{\pi_1^2 (\pi_1 - 1)^2 (S + C - 1) W_{01}}{[(C - 1)(\pi_1 - 1)\psi + \pi_1 S]^2} \Bigg] \\
& \times \frac{1}{(\pi_1 + \psi - \pi_1 \psi)^2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell_\psi}{\partial \psi \partial \pi_1} = & \frac{-\pi_1 (V_{001} + V_{000})}{\psi (\pi_1 - 1)(-\psi - \pi_1 + \psi \pi_1)} - \frac{\pi_1 (V_{001} + V_{000})}{\psi (\pi_1 - 1)(\psi + \pi_1 - \psi \pi_1)^2} \\
& + \frac{(V_{101} + V_{100})(\pi_1 - 1)(\psi - 1)}{(\psi + \pi_1 - \psi \pi_1)^2} + \frac{V_{101} + V_{100}}{\psi + \pi_1 - \psi \pi_1} \\
& + \frac{2\pi_1 (V_{001} + V_{000})(\psi - 1)}{\psi (\psi + \pi_1 - \psi \pi_1)^2} + \frac{(V_{001} + V_{000})}{\psi (\psi + \pi_1 - \psi \pi_1)} \\
& + \frac{(-\psi + \pi_1 + \psi \pi_1)(S + C - 1)(N_0 - W_{01})}{(\psi + \pi_1 - \psi \pi_1)^2 (C\psi(\pi_1 - 1) + \pi_1 (S - 1))} \\
& - \frac{\psi \pi_1 (\pi_1 - 1)(S + C - 1)^2 (N_0 - W_{01})}{(\psi + \pi_1 - \psi \pi_1)^2 (C\psi(1 - \pi_1) + \pi_1 (1 - S))^2} \\
& - \frac{\psi \pi_1 (\pi_1 - 1)(S + C - 1)^2 W_{01}}{(\psi + \pi_1 - \psi \pi_1)^2 ((C - 1)\psi(\pi_1 - 1) + \pi_1 S)^2} \\
& + \frac{(\psi(\pi_1 - 1) + \pi_1)(S + C - 1) W_{01}}{(\psi + \pi_1 - \psi \pi_1)^2 ((C - 1)\psi(\pi_1 - 1) + \pi_1 S)},
\end{aligned}$$

$$\frac{\partial^2 \ell_\psi}{\partial \psi \partial S} = \Big[\; CN_0[(C-1)(\pi_1-1)\psi + \pi_1 S]^2$$
$$+ W_{01}(C-1)[\pi_1^2 - 2C(\pi_1-1)\pi_1\psi + C(\pi_1-1)^2\psi^2]$$
$$+ 2W_{01}(C-1)\pi_1^2 S - W_{01}\pi_1^2 S^2 \quad \Big]$$
$$\times \frac{-\pi_1(\pi_1-1)}{[C\psi(1-\pi_1)+\pi_1(1-S)]^2[(C-1)\psi(\pi_1-1)+\pi_1 S]^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \psi \partial C} = \Big[\; 2S\psi\pi_1(\pi_1-1)(S-1) - S\pi_1^2(S-1)$$
$$+ W_{01}\psi^2(\pi_1-1)^2\left[1 + C^2 + 2C(S-1) - S\right]$$
$$+ N_0(S-1)[(C-1)\psi(\pi_1-1)+\pi_1 S]^2 \quad \Big]$$
$$\times \frac{\pi_1(\pi_1-1)}{[(C-1)\psi(\pi_1-1)+\pi_1 S]^2[C\psi - \pi_1(S+C\psi-1)]^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \pi_1^2} = \frac{2(V_{001}+V_{000})(\psi-1)}{(\pi_1-1)(\psi+\pi_1-\psi\pi_1)^2} + \frac{\psi(V_{101}+V_{100})}{(-\psi-\pi_1+\psi\pi_1)\pi_1^2} + \frac{\psi(\psi-1)(V_{101}+V_{100})}{\pi_1(\psi+\pi_1-\psi\pi_1)^2}$$
$$- \frac{(V_{110}+V_{111})}{\pi_1^2} - \frac{(V_{010}+V_{011})}{(\pi_1-1)^2} + \frac{2\psi(\psi-1)(S+C-1)(N_0-W_{01})}{(\psi+\pi_1-\psi\pi_1)^2(C\psi(\pi_1-1)+\pi_1(S-1))}$$
$$- \frac{(V_{001}+V_{000})}{(\pi_1-1)^2(\psi+\pi_1-\psi\pi_1)^2} - \frac{(S+C-1)^2 W_{11}}{(1+C(\pi_1-1)+\pi_1(S-1))^2}$$
$$- \frac{\psi^2(S+C-1)^2(N_0-W_{01})}{(\psi+\pi_1-\psi\pi_1)^2(C\psi(1-\pi_1)+\pi_1(1-S))^2}$$
$$- \frac{\psi^2(S+C-1)^2 W_{01}}{(\psi+\pi_1-\psi\pi_1)^2((C-1)\psi(\pi_1-1)+\pi_1 S)^2}$$
$$+ \frac{2\psi(\psi-1)(S+C-1)W_{01}}{(\psi+\pi_1-\psi\pi_1)^2((C-1)\psi(\pi_1-1)+\pi_1 S)}$$
$$- \frac{(S+C-1)^2(N_1-W_{11})}{(C(\pi_1-1)+\pi_1(S-1))^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial \pi_1 \partial S} = \frac{2\psi V_{101}}{\pi_1 S(\pi_1\psi - \pi_1 - \psi)(\pi_1 + \psi - \psi\pi_1)} + \frac{N_1 - W_{11}}{C(\pi_1 - 1) + \pi_1(S-1)}$$
$$+ \frac{W_{11}}{(C-1)(\pi_1 - 1) + \pi_1 S} + \frac{\pi_1\psi(S + C - 1)(N_0 - W_{01})}{(\pi_1\psi - \pi_1 - \psi)(C\psi - \pi_1(S + C\psi - 1))^2}$$
$$- \frac{\pi_1 W_{11}(S + C - 1)}{((C-1)(\pi_1 - 1) + \pi_1 S)^2} + \frac{\pi_1\psi(S + C - 1)W_{01}}{(\pi_1\psi - \pi_1 - \psi)(\psi(C-1)(\pi_1 - 1) + \pi_1 S)^2}$$
$$- \frac{\psi W_{01}}{(\pi_1\psi - \pi_1 - \psi)(\psi(C-1)(\pi_1 - 1) + \pi_1 S)} - \frac{\pi_1(S + C - 1)(N_1 - W_{11})}{(C(\pi_1 - 1) + \pi_1(S-1))^2}$$
$$+ \frac{\psi(-N_0 + W_{01})}{(\pi_1\psi - \pi_1 - \psi)(-C\psi + \pi_1(S + C\psi - 1))},$$

$$\frac{\partial^2 \ell_\psi}{\partial \pi_1 \partial C} = \frac{(\pi_1 - 1)\psi^2(S + C - 1)(N_0 - W_{01})}{(\pi_1\psi - \pi_1 - \psi)(C\psi - \pi_1(S + C\psi - 1))^2} + \frac{W_{11}}{(C-1)(\pi_1 - 1) + \pi_1 S}$$
$$+ \frac{(\pi_1 - 1)\psi^2(S + C - 1)W_{01}}{(\pi_1\psi - \pi_1 - \psi)(\psi(C-1)(\pi_1 - 1) + \pi_1 S)^2} + \frac{N_1 - W_{11}}{C(\pi_1 - 1) + \pi_1(S-1)}$$
$$- \frac{(\pi_1 - 1)(S + C - 1)(N_1 - W_{11})}{(C(\pi_1 - 1) + \pi_1(S-1))^2} - \frac{(\pi_1 - 1)(S + C - 1)W_{11}}{((C-1)(\pi_1 - 1) + \pi_1 S)^2}$$
$$- \frac{\psi W_{01}}{(\pi_1\psi - \pi_1 - \psi)(\psi(C-1)(\pi_1 - 1) + \pi_1 S)}$$
$$+ \frac{\psi(-N_0 + W_{01})}{(\pi_1\psi - \pi_1 - \psi)(-C\psi + \pi_1(S + C\psi - 1))},$$

$$\frac{\partial^2 \ell_\psi}{\partial S^2} = -\frac{(V_{101} + V_{111})}{S^2} - \frac{(V_{100} + V_{110})}{(S-1)^2}$$
$$- \frac{\pi_1^2 W_{01}}{(\psi(C-1)(\pi_1 - 1) + \pi_1 S)^2} - \frac{\pi_1^2 W_{11}}{((C-1)(\pi_1 - 1) + \pi_1 S)^2}$$
$$+ \frac{\pi_1^2(W_{01} - N_0)}{(C\psi - \pi_1(S + \psi C - 1))^2} + \frac{\pi_1^2(W_{11} - N_1)}{(C(\pi_1 - 1) + \pi_1(S-1))^2},$$

$$\frac{\partial^2 \ell_\psi}{\partial S \partial C} = \pi_1 \times \Bigg[- \frac{\psi(\pi_1 - 1)(N_0 - W_{01})}{(C\psi - \pi_1(S + \psi C - 1))^2} - \frac{(\pi_1 - 1)\psi W_{01}}{(\psi(C-1)(\pi_1 - 1) + \pi_1 S)^2}$$
$$- \frac{(\pi_1 - 1)(N_1 - W_{11})}{(C(\pi_1 - 1) + \pi_1(S-1))^2} + \frac{(\pi_1 - 1)W_{11}}{((C-1)(\pi_1 - 1) + \pi_1 S)^2} \Bigg]$$

$$\frac{\partial^2 \ell_\psi}{\partial C^2} = -\frac{(V_{000}+V_{010})}{C^2} - \frac{(V_{001}+V_{011})}{(C-1)^2} \\ - \frac{(\pi_1-1)^2\psi^2 W_{01}}{(\psi(C-1)(\pi_1-1)+\pi_1 S)^2} - \frac{(\pi_1-1)^2 W_{11}}{((C-1)(\pi_1-1)+\pi_1 S)^2} \\ - \frac{\psi^2(\pi_1-1)^2(N_0-W_{01})}{(C\psi-\pi_1(S+\psi C-1))^2} - \frac{(\pi_1-1)^2(N_1-W_{11})}{(C(\pi_1-1)+\pi_1(S-1))^2}.$$

We obtain the remaining terms beneath the diagonal of $\mathbf{J}(\psi, \boldsymbol{\eta})$ from its symmetry.

APPENDIX C

Derivations for Chapter Four

C.1 Restricted Maximum-Likelihood Estimators

Recall that λ is the occurrence rate parameter of interest, and $\boldsymbol{\eta}$ is the nuisance parameter vector. Also, let $\boldsymbol{d}^c \equiv (z, y, x, t_0, y_0, x_0)'$ be the complete-data. Then the concentrated complete-data likelihood is

$$L^c(\lambda, \boldsymbol{\eta}|\boldsymbol{d}^c) \propto \lambda^{z+x+t_0-y} e^{-\lambda(A+A_0)} \phi^{y+y_0} e^{-\phi(A+A_0)} \theta^{x+x_0} (1-\theta)^{z+t_0-y-x_0}. \quad \text{(C.1.1)}$$

Furthermore, the complete data log likelihood is

$$\ell^c = constant + (z + x + t_0 - y) \ln \lambda - \lambda(A + A_0) + (y + y_0) \ln \phi$$
$$- \phi(A + A_0) + (x + x_0) \ln \theta + (z + t_0 - y - x_0) \ln(1 - \theta).$$

To implement the EM algorithm, one must know the full conditional distributions of the unobserved variables x and y. The conditional distribution of x is

$$f(x|z, y, \lambda, \phi, \theta) = \frac{f(z, y, x, |\lambda, \phi, \theta)}{\sum_{x=0}^{\infty} f(z, y, x|\lambda, \phi, \theta)}. \quad \text{(C.1.2)}$$

After some algebraic manipulations, we have that

$$\sum_{x=0}^{\infty} f(z, y, x|\lambda, \phi, \theta) = e^{-A\lambda} (A\lambda)^{z-y} \frac{e^{-A\phi}(A\phi)^y}{y!(z-y)!} (1-\theta)^{z-y} \sum_{x=0}^{\infty} \frac{(A\lambda\theta)^x}{x!}.$$

Hence,

$$f(x|z, y, \lambda, \phi, \theta) = \frac{(A\lambda\theta)^x e^{-A\lambda\theta}}{x!} \quad \text{(C.1.3)}$$

and

$$f(y|z, x, \lambda, \phi, \theta) = \begin{pmatrix} z \\ y \end{pmatrix} \left(\frac{\phi}{\phi + \lambda(1-\theta)} \right)^y \left(\frac{\lambda(1-\theta)}{\phi + \lambda(1-\theta)} \right)^{z-y}. \quad \text{(C.1.4)}$$

3.1.1 EM Algorithm

E-step: Let $\boldsymbol{\rho}^{(r)} = (\lambda^{(r)}, \phi^{(r)}, \theta^{(r)})'$ be the current parameter vector at the r^{th} iteration. From (C.1.3) and (C.1.4), we determine that the conditional expectations of the unobserved variables are

$$x^* \equiv E[x|\boldsymbol{d}, \boldsymbol{\rho}^{(r)}] = A\lambda^{(r)}\theta^{(r)}$$

and

$$y^* \equiv E[y|\boldsymbol{d}, \boldsymbol{\rho}^{(r)}] = z\frac{\phi^{(r)}}{\phi^{(r)} + \lambda^{(r)}(1-\theta^{(r)})}.$$

M-step: In this step we update the parameter values λ, ϕ, and θ at each iteration by using the solutions to

$$\frac{\partial \ell^c}{\partial \lambda} = -(A + A_0) + \frac{t_0 + x - y + z}{\lambda} = 0,$$

$$\frac{\partial \ell^c}{\partial \phi} = -(A + A_0) + \frac{y + y_0}{\phi} = 0,$$

and

$$\frac{\partial \ell^c}{\partial \theta} = \frac{x + x_0}{\theta}\frac{t_0 - x_0 - y + z}{\theta - 1} = 0. \tag{C.1.5}$$

We solve (C.1.5) for the respective nuisance parameters λ, ϕ, and θ. Then, for the r^{th} iteration, we have

$$\lambda^{(r+1)} = \frac{t_0 + x - y + z}{A + A_0},$$
$$\phi^{(r+1)} = \frac{y + y_0}{A + A_0},$$

and

$$\theta^{(r+1)} = \frac{x + x_0}{t_0 + x - y + z}.$$

Below, we obtain the terms inside the OFIM in order to calculate the Wald, score, PL, and AIL CIs. The terms on the upper diagonal of the matrix $\mathbf{J}(\lambda, \boldsymbol{\eta})$ are

$$\begin{aligned}
\frac{\partial^2 \ell_o}{\partial \lambda^2} &= -\frac{t_0}{\lambda^2} - \frac{(1-\theta)^2 z}{(\lambda + \phi - \lambda\theta)^2}, \\
\frac{\partial^2 \ell_o}{\partial \lambda \partial \phi} &= -\frac{(1-\theta) z}{(\lambda + \phi - \lambda\theta)^2}, \\
\frac{\partial^2 \ell_o}{\partial \lambda \partial \theta} &= A - \frac{\phi z}{(\lambda + \phi - \lambda\theta)^2}, \\
\frac{\partial^2 \ell_o}{\partial \phi^2} &= -\frac{y_0}{\phi^2} - \frac{z}{(\lambda + \phi - \lambda\theta)^2}, \\
\frac{\partial^2 \ell_o}{\partial \phi \partial \theta} &= \frac{\lambda z}{(\lambda + \phi - \lambda\theta)^2},
\end{aligned}$$

and

$$\frac{\partial^2 \ell_o}{\partial \theta^2} = -\frac{x_0}{\theta^2} + \frac{x_0 - t_0}{(\theta - 1)^2} - \frac{\lambda^2 z}{(\lambda + \phi - \lambda\theta)^2}.$$

We obtain the remaining terms via the symmetry property of $\mathbf{J}(\lambda, \boldsymbol{\eta})$.

APPENDIX D

Derivations for Chapter Five

D.1 Restricted Maximum-Likelihood Estimators

Let $\boldsymbol{\Psi} \equiv (\lambda, \theta)'$ represent the vector of parameters, where λ is the occurrence rate parameter of interest, and θ is the false-negative misclassification parameter. Let $\boldsymbol{d}^c \equiv (z, x, t_0, x_0)'$ be the complete data. Then, the complete-data likelihood is

$$L^c(\lambda, \boldsymbol{\eta}|\boldsymbol{d}^c) \propto \lambda^{z+x+t_0} e^{-\lambda(A+A_0)} \theta^{x+x_0} (1-\theta)^{z+t_0-x_0}. \tag{D.1.1}$$

and the complete-data log-likelihood is

$$\ell^c = constant + (z + x + t_0) \ln \lambda - \lambda(A + A_0) +$$
$$(x + x_0) \ln \theta + (z + t_0 - x_0) \ln(1 - \theta).$$

Using

$$\sum_{x=0}^{\infty} f(z, x|\lambda, \theta) = e^{-A\lambda} \frac{(A\lambda)^z}{z!} (1-\theta)^z \sum_{x=0}^{\infty} \frac{(A\lambda\theta)^x}{x!},$$

we have that the full conditional distribution of x is

$$f(x|z, \lambda, \theta) = \frac{(A\lambda\theta)^x e^{-A\lambda\theta}}{x!}. \tag{D.1.2}$$

4.1.1 An EM Algorithm

E-step: Let $\boldsymbol{\Psi}^{(r)} \equiv (\lambda^{(r)}, \theta^{(r)})'$ be the current parameter vector at the r^{th} iteration. From (D.1.2), the conditional expectation of the unobserved undercount variable x is

$$x^* \equiv E[x|\boldsymbol{d}, \boldsymbol{\Psi}^{(r)}] = A\lambda^{(r)}\theta^{(r)}.$$

M-step: We update the parameter estimates by using the solutions to

$$\frac{\partial \ell^c}{\partial \lambda} = -A - A_0 + \frac{(t_0 + x + z)}{\lambda} = 0$$

and (D.1.3)

$$\frac{\partial \ell^c}{\partial \theta} = \frac{x + x_0}{\theta} \; \frac{t_0 - x_0 + z}{\theta - 1} = 0.$$

Solving equations (D.1.3) at each iteration for the two parameters, we have

$$\lambda^{(r+1)} = \frac{t_0 + x + z}{A + A_0}$$

and

$$\theta^{(r+1)} = \frac{x + x_0}{t_0 + x + z}$$

for the r^{th} iteration.

BIBLIOGRAPHY

Adams, T. D., Gress, R. E., Smith, S. C., Halverson, R. C., Simper, S. C., Rosamond, W. D., LaMonte, M. J., Stroup, A. M., and Hunt, S. C. (2007), "Long-Term Mortality after Gastric Bypass Surgery," *New England Journal of Medicine*, 357, 753 - 761, pMID: 17715409.

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

Anderson, C., Bratcher, T., and Kutran, K. (1994), "Bayesian Estimation of Population Density and Visibility," *Texas Journal of Science*, 46, 7 - 12.

Aschengrau, A. and Seage, G. R. (2013), *Essentials of Epidemiology in Public Health*, Jones and Bartlett Learning, 3rd ed.

Barker, L. (2002), "A Comparison of Nine Confidence Intervals for a Poisson Parameter When the Expected Number of Events is 5," *The American Statistician*, 56, 85 - 89.

Barndorff-Nielsen, O. E. (1994), "Adjusted Versions of Profile Likelihood and Directed Likelihood, and Extended Likelihood," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56, 125 - 140.

Barnett, V., Haworth, J., and Smith, T. M. F. (2001), "A Two-Phase Sampling Scheme with Application to Auditing," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 164, 407 - 422.

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.

Blumenthal, U. J., Fleisher, J. M., Esrey, S. A., and Peasey, A. (2001), "Epidemiology: a Tool for the Assessment of Risk," *Water Quality: Guidelines and, Standards and Health. IWA Publishing, London*, 135 - 160.

Blyth, C. R. and Still, H. A. (1983), "Binomial Confidence Intervals," *Journal of the American Statistical Association*, 78, 108 - 116.

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.

Bonett, D. G. and Price, R. M. (2012), "Adjusted Wald Confidence Interval for a Difference of Binomial Proportions Based on Paired Data," *Journal of Educational and Behavioral Statistics*, 37, 479 - 488.

Bratcher, T. L. and Stamey, J. D. (2002), "Estimation of Poisson Rates with Misclassified Counts," *Biometrical Journal*, 44, 946 - 956.

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, T. T. (1979), "Log-Linear Models for Categorical Data With Misclassification and Double Sampling," *Journal of the American Statistical Association*, 74, 481 - 488.

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-Villaseñor, 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.

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.

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

Fader, P. and Hardie, B. (2000), "A Note on Modelling Underreported Poisson Counts," *Journal of Applied Statistics*, 27 (8), 953 - 964.

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.

Gaba, A. and Winkler, R. L. (1992), "Implications of Errors in Survey Data: A Bayesian Model," *Management Science*, 38, 913 - 925.

Geng, Z. and Asano, C. (1989), "Bayesian Estimnation and Methods for Categorical Data with Misclassifications," *Communications in Statistics: Theory and Methods*, 18, 2935 - 2954.

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.

Heider, D., Kitze, K., Zieger, M., Riedel-Heller, S. G., and Angermeyer, M. C. (2007), "Health-Related Quality of Life in Patients after Lumbar Disc Surgery: A Longitudinal Observational Study," *Quality of Life Research*, 16, 1453 - 1460.

Hochberg, Y. (1977), "On the Use of Double Sampling Schemes in Analyzing Categorical Data with Misclassification Errors," *Journal of the American Statistical Association*, 72, 914 - 921.

Inskip, H. M., Godfrey, K. M., Robinson, S. M., Law, C. M., Barker, D. J., Cooper, C., , and the SWS Study Group (February 2006), "Cohort Profile: The Southampton Women's Survey," *International Journal of Epidemiology*, 35, 42 - 48.

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," *American Journal of Epidemiology*, 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.

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

Kircher, T., Nelson, J., and Burdo, H. (1985), "The Autopsy as a Measure of Accuracy of the Death Certificate," *New England Journal of Medicine*, 313, 1263 - 1269, pMID: 4058507.

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," *International Journal of Epidemiology*, 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.

MacMahon, B. and Trichopoulos, D. (1996), *Epidemiology Principles and Methods*, Lippincot Williams and Wilkins, 2nd ed.

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.

Markova, D. G. (2011), "Topics in Odds Ratio Estimation in the Case Control Studies and the Bioequivalence Testing in the Crossover Studies," Ph.D. thesis, Baylor University.

McCullagh, P. and Tibshirani, R. (1990), "A Simple Method for the Adjustment of Profile Likelihoods," *Journal of the Royal Statistical Society. Series B (Methodological)*, 52, 325 - 344.

Moors, J. J. A., Van Der Genugten, B. B., and Strijbosch, L. W. G. (2000), "Repeated Audit Controls," *Statistica Neerlandica*, 54, 3 - 13.

Moreno, E. and Girón, J. (1998), "Estimating with Incomplete Count Data A Bayesian Approach," *Journal of Statistical Planning and Inference*, 66, 147 - 159.

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.

Pepe, M. S. (1992), "Inference Using Surrogate Outcome Data and a Validation Sample," *Biometrika*, 79, 355 - 365.

Pinder, R. J., Greenberg, N., Boyko, E. J., Gackstetter, G. D., Hooper, T. I., Murphy, D., Ryan, M. A., Smith, B., Smith, T. C., Wells, T. S., and Wessely, S. (2012), "Profile of two Cohorts: UK and US Prospective Studies of Military Health," *International Journal of Epidemiology*, 41, 1272 - 1282.

Power, C. and Elliott, J. (February 2006), "Cohort profile: 1958 British birth cohort (National Child Development Study)," *International Journal of Epidemiology*, 35, 34 - 41.

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

Rahardja, D. and Young, D. M. (2011), "Lilelihood-Based Confidence Intervals for the Risk Ratio using Double Sampling with Over-Reported Binary Data," *Computational Statistics and Data Analysis*, 55, 813 - 823.

Rahme, E., Joseph, L., and Gyorkos, T. W. (2000), "Bayesian Sample Size Determination for Estimating Binomial Parameters from Data Subject to Misclassification," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 49, 119 - 128.

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., Young, D., and Stamey, J. (2009), "Likelihood-Based Confidence Intervals for Complementary Poisson Rate Parameters with Misclassified Data," *Communications in Statistics - Theory and Methods*, 38, 159 - 172.

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

Robert H. Lyles, Li Tang, H. M. S. C. C. K. D. D. C. Y. L. J. D. S. (2011), "Validation Data-Based Adjustments for Outcome Misclassification in Logistic Regression: an Illustration." *Epidemiology*, 22, 589 - 597.

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

Rothman, K. J., Greenland, S., and Lash, T. L. (2008), *Modern Epidemiology*, Lippincott Williams & Wilkins, third edition 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.

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.

Snow, J. (1851), "On the Mode of Propagation of Cholera," *Medical Times*, 3, 559 - 562.

Sorahan, T. and Gilthorpe, M. S. (1994), "Non-Differential Misclassification of Exposure Always Leads to an Underestimate of Risk: an Incorrect Conclusion," *Occupational and Environmental Medecine*, 51, 839 - 840.

Sposto, R., Preston, D. L., Shimizu, Y., and Mabuchi, K. (1992), "The Effect of Diagnostic Misclassification on Non-Cancer and Cancer Mortality Dose Response in A-Bomb Survivors," *Biometrics*, 48, 605 - 617.

Stamey, J. (2000), "A Bayesian Analysis of Poisson Data with Misclassification," Ph.D. thesis, Baylor University.

Stamey, J., Young, D., and Stephens, D. (2005), "Maximum Likelihood Estimation of Two Inversely Related Poisson Rate Parameters with Misclassified Data," *American Journal of Mathematics Management and Science*, 25, 65 - 82.

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.

Stamey, J. D. and Young, D. M. (2005), "Maximum Likelihood Estimation for a Poisson Model with Misclassified Counts," *Australian and New Zealand Journal of Statistics*, 47, 163172.

Stamey, J. D., Young, D. M., and Cecchini, M. (2003a), "A Double-Sampling Approach For Maximum Likelihood Estimation For a Poisson Rate Parameter With Visibility-Biased Data," *Statistica, Anno LXIII*, n.1, 3 - 11.

— (2003b), "A Double-Sampling Approach For Maximum Likelihood Estimation For a Poisson Rate Parameter With Visibility-Biased Data," *Statistica, Anno LXIII*, 1, 3 - 11.

Szumilas, M. (2010), "Explaining Odds Ratios," *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19, 227 - 229.

Tang, M.-L., Qiu, S.-F., and Poon, W.-Y. (2012), "Comparison of Disease Prevalence in Two Populations in the Presence of Misclassification," *Biometrical Journal*, 54, 786 - 807.

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

— (1971), "A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications: Sample Size Determination," *Biometrics*, 27, 935 - 944.

— (1972), "A Double Sampling Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection," *Technometrics*, 14, 187 - 202.

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.

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.

Viana, M., Ramakrishnan, V., and Levy, P. (1993), "Bayesian Analysis of Prevalence from the Results of Small Screening Samples," *Communications in Statistics: Theory and Methods*, 22, 575 - 585.

Wadsworth, M., Kuh, D., Richards, M., and Hardy, R. (February 2006), "Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development)," *International Journal of Epidemiology*, 35, 49 - 54.

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.

Whittemore, A. S. and Gong, G. (1991), "Poisson Regression with Misclassified Counts: Application to Cervical Cancer Mortality Rates," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 40, 81 - 93.

York, J., Madigan, D., Heuch, I., and Lie, R. T. (1995), "Birth Defects Registered by Double Sampling: A Bayesian Approach Incorporating Covariates and Model Uncertainty," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44, 227 - 242.