ABSTRACT

Bayesian and Likelihood-Based Interval Estimation for the Risk Ratio Using Double Sampling with Misclassified Binomial Data

Dewi Gabriela Rahardja, Ph.D.

Chairperson: Dean M. Young, Ph.D.

We consider the problem of point and interval estimation for the risk ratio using double sampling with two-sample misclassified binary data. For such data, it is well-known that the actual data model is unidentifiable. To achieve model identifiability, then, we obtain additional data via a double-sampling scheme.

For the Bayesian paradigm, we devise a parametric, straight-forward algorithm for sampling from the joint posterior density for the parameters, given the data. We then obtain Bayesian point and interval estimators of the risk ratio of two-proportion parameters. We illustrate our algorithm using a real data example and conduct two Monte Carlo simulation studies to demonstrate that both the point and interval estimators perform well.

Additionally, we derive three likelihood-based confidence intervals (CIs) for the risk ratio. Specifically, we first obtain closed-form maximum likelihood estimators (MLEs) for all parameters. We then derive three CIs for the risk ratio: a naive Wald interval, a modified Wald interval, and a Fieller-type interval. For illustration purposes, we apply the three CIs to a real data example. We also perform various Monte Carlo simulation studies to assess and compare the coverage probabilities and average lengths of the three CIs. A modified Wald CI performs the best of the three CIs and has near-nominal coverage probabilities.

Bayesian and Likelihood-Based Interval Estimation for the Risk Ratio Using Double Sampling with Misclassified Binomial Data

by

Dewi Gabriela Rahardja, B.S., M.Eng., M.S., M.M., Ph.D.

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

_____

Dennis A. Johnston, Ph.D.

_____

James D. Stamey, Ph.D.

_____

Jack D. Tubbs, Ph.D.

_____

David J. Ryden, Ph.D.

Accepted by the Graduate School
December 2010

_____

J. Larry Lyon, Ph.D., Dean

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Gaby Rahardja deeply appreciates Dr. Jack D. Tubbs, Dr. Tom L. Bratcher, and other faculty members for all their support and encouragement that enable her to obtain this statistics doctoral degree from the Baylor University.

She also would like to sincerely thank her dissertation advisor, Dr. Dean M. Young, for his guidance throughout her dissertation journey, Mrs. Joy Young for her careful editing of this dissertation, Dr. Philip Young for his help on several occasions, and Dr. James D. Stamey, Dr. Dennis A. Johnston, and Dr. David J. Ryden for their assistance as dissertation committee members.

Last but not least, she is thankful to her family members, who always support, encourage, and love her unconditionally.

Soli Deo Gloria

CHAPTER ONE

Introduction

1.1 Overview

Most binary devices or binary classifiers are fallible and sometimes misclassify units into one of two mutually exclusive categories. The resulting observations are referred to as misclassified binary data. For example, in the medical field Hildesheim, Mann, Brinton, Szklo, Reeves, and Rawls (1991) have reported that a western blot procedure may misclassify an uninfected individual to have herpes simplex virus or vice versa.

Generally, two types of misclassified observations exist: false-positive and false-negative observations. For example, visual inspection by a midwife or obstetrician may misclassify a normal child as having Down's syndrome (false-positive error) or vice versa (false-negative error). In some situations, only one form of misclassification occurs. For instance, Perry, Vakil, and Cutler (2000) have analyzed blood testing data that had only false-positive or over-reported error. Also, Moors, van der Genugten, and Strijbosch (2000) have presented auditing data where only false-negative or under-reported errors occurred.

For binary data, Bross (1954) and Goldberg (1975) have demonstrated that classical estimators that ignore misclassification can be very biased. In addition to the bias problem, a model that incorporates misclassification has an unidentifiability issue. Hence, one needs additional data to achieve model identifiability and to correct for estimation bias.

Several information-addition methods are popular in the statistical literature. One method is the Bayesian approach that uses sufficiently informative priors specified by expert opinion or by previously collected data. Another information-addition

method is the use of multiple fallible classifiers. However, the information-addition method we utilize incorporates training data via a double-sampling scheme.

Statistical inferences on binary data with misclassification has been ongoing. For one-binomial-parameter problems, several researchers have considered the case where only false-negative errors occur. For example, Lie, Heuch, and Irgens (1994) have used a maximum likelihood approach with multiple fallible classifiers to correct false-negative errors. Also, York, Madigan, Heuch, and Lie (1995) have considered the same problem from a Bayesian perspective. Using a double-sampling scheme, Moors, van der Genugten, and Strijbosch (2000) have derived method of moment and maximum likelihood estimators and a one-sided interval estimator, and Boese, Young, and Stamey (2006) have derived several likelihood-based confidence intervals (CIs) for a single-proportion parameter using double sampling. For the same data, Lee and Byun (2008) have used noninformative priors to provide Bayesian credible intervals.

Additionally, several researchers have studied one-sample problems with both types of misclassification errors. Using double sampling, Tenenbein (1970) has proposed a maximum likelihood estimator for the proportion parameter and has derived its asymptotic variance approximation. Gaba and Winkler (1992) and Viana, Ramakrishnan, and Levy (1993) have developed Bayesian approaches using sufficiently informative priors for the case when training data are unavailable.

Also, Bayesian inference methods using informative priors have been developed for the case when training data were unavailable for two-sample problems with misclassification errors of both types. For example, Evans, Guttman, Haitovsky, and Swartz (1996) have derived a Bayesian approach for the risk difference, and Gustafson, Le, and Saskin (2001) have proposed a Bayesian method for the odds ratio. For the case when training data are obtained using a double-sampling scheme, Boese (2003) has derived several likelihood-based methods for the risk difference.

To date, we have found no inference methods for the risk ratio of two-proportion parameters using two-sample binomial data subject to misclassification. Therefore, the objective of this dissertation is to derive point and interval estimators for the risk ratio of two-proportion parameters using data from fallible and infallible classifiers.

The remainder of this chapter is organized as follows. In Section 1.2 we describe the double-sampling scheme pioneered by Tenenbein (1970). In Section 1.3 we describe a Bayesian approach to obtain point estimators and credible sets for a risk ratio. In Section 1.4 we discuss likelihood-based methods for obtaining point and interval estimators for the risk ratio. Specifically, we derive three likelihood-based interval estimators for the risk ratio by utilizing a double-sampling scheme. Finally, in Section 1.5 we summarize the organization of this dissertation.

1.2 Double-Sampling Scheme

One can apply Tenenbein's double-sampling scheme when both fallible and infallible classifiers are available. A fallible classifier is usually inexpensive but subject to errors in classified units, while an infallible (true) classifier is often more expensive but much more accurate. The double-sampling approach utilizes two different data sets: the original data, where only the fallible classifier is applied, and the smaller training data, where both the infallible classifier and fallible classifier are applied. Therefore, the use of both fallible and infallible classification procedures via double sampling is an economically viable method that yields model identifiability and reduces parameter estimation bias.

1.3 Bayesian Inference

We first consider a Bayesian inference approach for the risk ratio using misclassified binary data and training data obtained using a double-sampling scheme. Bayesian inference relies on the posterior distribution of parameters conditioned on the data. Once a posterior sample for a parameter is obtained, inference on this

parameter, such as point and interval estimation, is straightforward. For example, we can use the median of the posterior sample as a point estimator and the $(\alpha/2)$ and $(1-\alpha/2)$ quantiles as the lower and upper limits of an approximate $100(1-\alpha)\%$ credible interval. Bayesian inference usually does not depend on large-sample approximations and, hence, is appealing.

Bayesian inference typically requires complex Markov Chain Monte Carlo algorithms for sampling from the posterior distribution. One popular posterior sampling algorithm is the Gibbs sampler, which sequentially draws each parameter after conditioning on all other parameters and the data (full conditionals). If one cannot find an explicit algorithm for sampling from one of the full conditionals, one can use a Metropolis-Hastings algorithm to sample from that full conditional distribution of interest.

In general, technical challenges are associated with using Gibbs sampling for Bayesian inference. In some applications deriving all of the full conditional densities or probability mass functions can be difficult. Also, a Metropolis-Hastings algorithm can be time-consuming. Additionally, specifying initial values may be challengin,g and determining a Markov-Chain convergence may be a nontrivial task.

In our Bayesian approach, we reparameterize our model and then derive a simple parametric algorithm for sampling from the marginal posterior distribution of the model parameters, given the data. Once a posterior sample is drawn for the proportion parameters of interest, we obtain a posterior sample for the risk ratio by dividing the corresponding sample proportion parameters. We then obtain a credible interval for the risk ratio based on this marginal posterior sample of the risk ratio.

Several advantages of our reparameterization-based algorithms exist:

(1) Because we sample directly from the posterior distributions, we need not specify initial values, and we have no burn-in period or convergence issues.

(2) Because posterior samples are available for each parameter, inferences on the risk difference and the odds ratio are straightforward.

(3) The algorithm can accommodate zero counts.

(4) We require no asymptotic theory.

(5) We can generalize our posterior sampling algorithm to three or more proportion parameters.

1.4 Likelihood-Based Inference

Likelihood-based inference methods are used in many statistical applications. Under certain regularity conditions, the maximum likelihood estimator (MLE) is asymptotically unbiased and efficient and has an approximate multivariate normal distribution with the true parameters as the mean vector and inverse information matrix as the covariance matrix. This property can be combined with the delta method for variance approximation to construct CIs for functions of parameters.

Although MLEs are often desirable, they may be very difficult to compute, especially when the number of parameters is large. Obtaining an MLE usually requires iterative numerical algorithms, such as the Newton-Raphson algorithm. Associated with such algorithms are the challenges of specifying initial values and achieving convergence.

In this dissertation, through a reparameterization, we derive a parametric, straightforward formula for computing MLEs of all the model parameters. Then, by the invariance property of MLEs, the MLE for the risk ratio is obtained by dividing the MLEs of the proportion parameters of interest. In addition, we derive a closed-form for the inverse Fisher information matrix. We then approximate the variance of this MLE by using the delta method and compute an approximate $100(1-\alpha)\%$ Wald type CI for the risk ratio.

It is well-known that Wald-type CIs, when applied to multiple parameters, generally have lower-than-nominal coverage probabilities, especially for relatively small sample sizes. To improve small-sample performance, we propose first constructing an approximate $100(1-\alpha)\%$ Wald-type CI for the log risk ratio via the delta method. Then, we exponentiate this CI to obtain an approximate $100(1-\alpha)\%$ CI for the risk ratio. We term this CI as a modified Wald CI.

Also, because the MLEs of the proportion parameters of interest have an asymptotic normal distribution, one can use Fieller's method (Fieller, 1954) to construct an approximate $100(1-\alpha)\%$ CI for the risk ratio. We describe the Fieller's interval estimation method as follows. Let Z_1 and Z_2 be two random variables having a bi-variate normal distribution with mean vector $(\mu_1, \mu_2)'$, marginal variances σ_1^2 and σ_2^2, and covariance σ_{12}. Here, the variables Z_1 and Z_2 are observed, parameters σ_1^2, σ_2^2, and σ_{12} are known, and $(\mu_1, \mu_2)'$ is unknown. The objective is to construct an approximate $100(1-\alpha)\%$ CI for the ratio of $r = \mu_1/\mu_2$. The Fieller's method is based on an approximate pivotal quantity

$$\frac{X_1 - rX_2}{\sqrt{\sigma_1^2 + r^2\sigma_2^2 - 2r\sigma_{12}}} \dot{\sim} N(0,1).$$

Therefore, confidence limits for an approximate $100(1-\alpha)\%$ CI for r are obtained by solving

$$\left(\frac{X_1 - rX_2}{\sqrt{\sigma_1^2 + r^2\sigma_2^2 - 2r\sigma_{12}}}\right)^2 = Z_{\alpha/2}^2$$

for r. Let $\Delta \equiv (X_1X_2 - Z_{\alpha/2}^2\sigma_{12})^2 - (X_1^2 - Z_{\alpha/2}^2\sigma_1^2)(X_2^2 - Z_{\alpha/2}^2\sigma_2^2)$. When $\Delta > 0$, an approximate $100(1-\alpha)\%$ Fieller CI for r is

$$\frac{X_1X_2 - Z_{\alpha/2}^2\sigma_{12} \pm \sqrt{\Delta}}{X_2^2 - Z_{\alpha/2}^2\sigma_2^2};$$

otherwise, a $100(1-\alpha)\%$ Fieller CI for r does not exist.

1.5 Dissertation Organization

In this dissertation we derive point and interval estimators for the risk ratio of the proportion parameters using double sampling with misclassified two-sample binary data. The risk ratio, also known as the relative risk, is defined as the ratio of two-proportion parameters. Each chapter of this dissertation is self-contained.

In Chapter 2, we consider point and interval estimation for the risk ratio based on two independent samples of binomial data subject to only false-positive misclassification. Using both the fallible and infallible data, we propose Bayesian point and interval estimators of the risk ratio. In particular, we derive an easy-to-implement algorithm for sampling from the marginal posterior distribution of the risk ratio.

In Chapter 3, we generalize our Bayesian approach in Chapter 2 to data having two types of misclassification errors. To obtain model identifiability, we apply a double-sampling scheme and propose a Bayesian method for statistical inference for a two-proportion risk ratio for the identifiable model.

In Chapter 4, we consider two-sample binary data subject to only false-positive misclassification and use training data obtained using a double-sampling scheme. By maximizing the full likelihood function, we derive closed-form maximum likelihood estimators for all model parameters. In addition, we derive three confidence intervals: a naive Wald interval, a modified Wald interval, and a Fieller-type interval.

In Chapter 5, we generalize our likelihood-based approach in Chapter 4 to data with two types of misclassification errors.

CHAPTER TWO

Credible Sets for the Risk Ratio in Over-Reported Two-Sample Binomial Data Using a Double Sampling Scheme

2.1 Introduction

Because of imprecise diagnostic procedures or human error, misclassification can occur when recording binary data. Usually, both possible types of misclassification, false-positive or false-negative errors, occur. For example, for an imperfect blood test, a healthy patient may be incorrectly diagnosed as diseased and vice versa. In some cases, only one misclassification type occurs. For instance, Moors, van der Genugten, and Strijbosch (2000) have presented auditing data where only false-negative or under-reported errors occurred. Also, Perry, Vakil, and Cutler (2000) have considered blood testing data that had only false-positive or over-reported errors.

Among others, Bross (1954) has reported that classical estimators that ignore misclassification can be extremely biased when used to analyze misclassified binary data. Therefore, one needs additional information or data to promote correct-model identifiability and to correct the bias. Several information-addition methods for this purpose are popular in the statistics literature. One is to include training data via a double-sampling scheme as suggested by Tenenbein (1970); another is to use informative priors in the Bayesian paradigm that are specified by expert opinion or by previous data. The rationale for Tenenbein's double-sampling scheme is straightforward. Fallible classification procedures result in misclassification but are inexpensive, while infallible classification procedures result in correct classification or labeling but are usually much more expensive. Therefore, the use of both fallible and infallible classification procedures is an economically viable method that yields model identi-

fiability. In some cases, an infallible test or procedure is unavailable or prohibitively expensive; informative priors can then be used to yield model identifiability in the Bayesian paradigm. Another information-addition method is to use multiple fallible classifiers. We next review the research literature of research on binomial parameter estimation with misclassified data. The objective of this research is to statistically infer on the proportion parameters using possibly misclassified data and some form of additional information.

For one-binomial-parameter problems, several researchers have considered the case where only false-negative errors were present. Lie, Heuch, and Irgens (1994) have used a maximum likelihood approach where false-negative errors are corrected using multiple fallible classifiers. York, Madigan, Heuch, and Lie (1995) have considered this same problem from a Bayesian perspective. Using data obtained via a double-sampling scheme, Moors et al. (2000) have discussed maximum likelihood estimation and one-sided interval estimation, and applied these methods to auditing data. Boese, Young, and Stamey (2006) have derived several likelihood-based CIs for a single proportion parameter using double sampling. Also, Lee and Byun (2008) have provided Bayesian credible intervals using noninformative priors.

In addition, several authors also have examined one-sample problems with both types of misclassification errors. Using a double sampling method, Tenenbein (1970) has proposed a maximum likelihood estimator and has derived an asymptotic variance for the proportion parameter. Gaba and Winkler (1992) and Viana, Ramakrishnan, and Levy (1993) have developed Bayesian approaches with sufficiently informative priors for the case when training data are unavailable in one-sample problems.

For two-sample problems with misclassification errors of both types, Bayesian inference methods using informative priors have been developed for the case when training data were unavailable. For example, Evans, Guttman, Haitovsky, and

Swartz (1996) have derived a Bayesian approach for the difference of two proportion parameters and Gustafson, Le, and Saskin (2001) have proposed a Bayesian inference method for the odds ratio.

To date, no inference methods for the risk ratio of two proportion parameters have been developed using two-sample binomial data subject to misclassification. In this article, we limit our scope to data subject to only one error type. That is, we consider data with false-positive errors only. We derive Bayesian approaches for point and interval estimation for the risk ratio. In Section 2.2 we describe the data and in Section 2.3 we develop Bayesian models and algorithms for binomial parameter estimation with false-positives. In Section 2.4 we illustrate our algorithms using real data. Also, we examine the performance of our Bayesian inference method in Section 2.5 and give a brief discussion in Section 2.6.

2.2 The Data

In this section we consider two-sample binomial data subject to misclassification. The data are obtained using a fallible test or fallible classification method that can yield false-positive but not false-negative observations. For example, suppose a study is conducted to assess whether a certain infection type has the same prevalence rates for men and women. A positive blood test outcome is used to determine whether a subject in the study is infected. This blood test is fallible with only false-positive counts if a false-positive classification is the only incorrect result given by the blood test.

To describe the data, let F_{ij} be the observed classification by the fallible classifier for the jth individual in the ith sample, where $i = 1, 2$, $j = 1, \ldots, M_i$, and

$$F_{ij} = \begin{cases} 1, & \text{if the fallible classifier yields positive result,} \\ 0, & \text{otherwise.} \end{cases}$$

Denoting the numbers of individuals with positive and negative labels by X_i and Y_i,

Table 2.1. Fallible Data for Sample i ($i = 1, 2$)

Classification	0	1	Total
Count	Y_i	X_i	M_i

respectively, the observed data using the fallible classifier for sample i, $i = 1, 2$, are displayed in Table 2.1.

Similarly, we define the unobserved true classification of the jth individual in the ith sample as T_{ij},

$$T_{ij} = \begin{cases} 1, & \text{if the classifier result is actually positive} \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, misclassification occurs when $T_{ij} \neq F_{ij}$.

Next, we introduce the following notation, for the ith sample, $i = 1, 2$:

$$\begin{aligned} p_i &\equiv \Pr(T_{ij} = 1), \\ \pi_i &\equiv \Pr(F_{ij} = 1), \\ \phi_i &\equiv \Pr(F_{ij} = 1 | T_{ij} = 0). \end{aligned}$$

Thus, p_i is the actual proportion parameter of interest, π_i is the proportion parameter of the fallible classifier, and ϕ_i is the false-positive rate of the fallible classifier. Here, we allow the false-positive rates to be different for the two samples, i.e., $\phi_1 \neq \phi_2$. Note that

$$\begin{aligned} \pi_i &= \Pr(T_i = 1) \Pr(F_i = 1 | T_i = 1) + \Pr(T_i = 0) \Pr(F_i = 1 | T_i = 0) \\ &= p_i + q_i \phi_i, \end{aligned} \tag{2.1}$$

where $q_i = 1 - p_i$, $i = 1, 2$.

As stated in Section 2.1, we are interested in statistical inference on the risk ratio

$$r = p_1 / p_2. \tag{2.2}$$

Because π_i is determined through p_i and ϕ_i, effectively there are four model parameters: p_1, ϕ_1, p_2, ϕ_2. However, $(X_1, X_2)'$ is a vector of sufficient statistics for this model. Because the dimension of the sufficient statistic is less than the number of parameters, the model is unidentifiable and, therefore, additional data or information is needed for model identifiability. In this paper we consider a double-sampling scheme with training data to provide the necessary additional information for parameter estimation.

Tenenbein (1970) has used additional training data obtained by double sampling to infer a single proportion parameter using one-sample binomial data subject to misclassification. Specifically, in addition to the original fallible data classified only by the fallible method, Tenenbein (1970) has used a second smaller training data set obtained by classifying each individual in the training data by both the fallible classifier and the infallible classifier. This sampling design enables the assessment of the misclassification rate or rates of the fallible classifier. Other applications of double-sampling include Tenenbein (1972), Hochberg (1977) and Boese et al. (2006).

We apply Tenenbein's double-sampling scheme to our two-sample problem and obtain n_i training data in addition to the original M_i fallible data for the ith sample, $i = 1, 2$. The combined data are presented in Table 2.2. In this table we use n_{ijk} to denote the number of individuals classified as j and k by the infallible and fallible methods, respectively. For example, n_{i01} is the number of individuals in the ith sample classified as negative by the infallible classifier but positive by the fallible classifier. We remark that for the case we consider here, n_{i10} is not available because false-negative errors cannot occur. For easy reference, we give the cell probabilities for Table 2.2 in Table 2.3.

Table 2.2. Data for Sample i, $i = 1, 2$.

		Fallible Method		
Data	Infallible Method	0	1	Total
Training	0	n_{i00}	n_{i01}	$n_{i0\cdot}$
	1	NA	n_{i11}	n_{i11}
	Total	n_{i00}	$n_{i\cdot 1}$	n_i
Original	NA	Y_i	X_i	M_i

NA: Not Available

Table 2.3. Cell Probabilities for Sample i, $i = 1, 2$.

		Fallible Method		
Data	Infallible Method	0	1	Total
Training	0	$q_i(1-\phi_i)$	$q_i\phi_i$	q_i
	1	NA	p_i	p_i
Original	NA	$1-\pi_i$	π_i	1

NA: Not Available

2.3 The Model

In this section we develop Bayesian point and interval estimators for the risk ratio using the data model described in the previous section. Our aim is to derive explicit algorithms for sampling from the posterior distribution of all the parameters, given the data. After posterior samples are drawn for both p_1 and p_2, a posterior sample for risk ratio r is readily obtained by dividing the sample of p_1 by the sample of p_2, elementwise. We obtain point and interval estimators for r based on this sample draw.

For sample i, $i = 1, 2$, in Table 2.2, the observed counts $(n_{i00}, n_{i01}, n_{i11})'$ of the training data have a trinomial distribution with total size n_i and the probabilities displayed in an upper right 2×2 submatrix in Table 2.3, i.e.,

$$(n_{i00}, n_{i01}, n_{i11})|p_i, \phi_i \sim \text{Trin}[n_i, (q_i(1-\phi_i), q_i\phi_i, p_i)].$$

In addition, the observed counts $(X_i, Y_i)'$ have the binomial distribution

$$(X_i, Y_i)|p_i, \phi_i \sim \text{Bin}[M_i, (\pi_i, 1-\pi_i)],$$

$i = 1, 2$. Because $(n_{i00}, n_{i01}, n_{i11})'$ and $(X_i, Y_i)'$ are independent for sample i, $i = 1, 2$, and because sample 1 is independent of sample 2, the sampling distribution of the data vector, given the parameter vector, is

$$f(\boldsymbol{d}|\boldsymbol{\eta}) \propto \Pi_{i=1}^{2}\{[q_i(1-\phi_i)]^{n_{i00}}(q_i\phi_i)^{n_{i01}}p_i^{n_{i11}}\pi_i^{X_i}(1-\pi_i)^{Y_i}\}, \tag{2.3}$$

where

$$\boldsymbol{d} = (n_{100}, n_{101}, n_{111}, X_1, Y_1, n_{200}, n_{201}, n_{211}, X_2, Y_2)'$$

and

$$\boldsymbol{\eta} = (p_1, \phi_1, p_2, \phi_2)'.$$

In our Bayesian framework, we choose a non-informative proper prior for $\boldsymbol{\eta}$. Specifically, we choose a uniform prior for each component of $\boldsymbol{\eta}$ and assume that these priors are independent; i.e., the joint prior distribution is

$$p(\boldsymbol{\eta}) = 1. \tag{2.4}$$

Combining (2.3) and (2.4), we obtain the joint posterior

$$f(\boldsymbol{\eta}|\boldsymbol{d}) \propto \Pi_{i=1}^{2}\{[q_i(1-\phi_i)]^{n_{i00}}(q_i\phi_i)^{n_{i01}}p_i^{n_{i11}}\pi_i^{X_i}(1-\pi_i)^{Y_i}\}, \tag{2.5}$$

that has the same functional form as the sampling distribution in (2.3).

To draw from the posterior density in (2.5), we first transform the parameter vector $\boldsymbol{\eta}$. We note that

$$1 - \pi_i = q_i(1-\phi_i) \tag{2.6}$$

and let

$$\lambda_i \;=\; p_i/\pi_i. \tag{2.7}$$

Incorporating (2.6) and (2.7), we find the posterior density in (2.5) is

$$f(\boldsymbol{\eta}|\boldsymbol{d}) \propto \Pi_{i=1}^{2} \lambda_i^{n_{i11}} (1-\lambda_i)^{n_{i01}} \pi_i^{X_i+n_{i.1}} (1-\pi_i)^{Y_i+n_{i00}}. \tag{2.8}$$

Because the transformed parameters

$$(\lambda_1, \pi_1, \lambda_2, \pi_2)'$$

are now separable, one can straightforwardly draw λ_i and π_i from (2.8) by using

$$\lambda_i \sim \text{Beta}(n_{i11}+1, n_{i01}+1) \tag{2.9}$$

and

$$\pi_i \sim \text{Beta}(X_i+n_{i.1}+1, Y_i+n_{i00}+1). \tag{2.10}$$

We obtain p_i and ϕ_i by solving (2.1) and (2.7) so that

$$p_i = \pi_i \lambda_i \tag{2.11}$$

and

$$\phi_i = (1-\lambda_i)\pi_i / q_i \tag{2.12}$$

for $i = 1, 2$.

We summarize our parametric Monte Carlo sampling algorithm for the posterior density in (2.5) as follows. First, choose a large number J, say, 10,000, for the posterior draw sample size. For $i = 1, 2$,

(1) Obtain a size J sample of λ_i and π_i using (2.9), and (2.10).

(2) Obtain a size J sample of p_i and ϕ_i using (2.11) and (2.12).

(3) Obtain a size J sample of the risk ratio r using (2.2).

We then use the sample median as a point estimator of r because the distribution of the posterior sample of r is skewed. Finally, we obtain an approximate $100(1-\alpha)\%$ CI for r by using the $\alpha/2$ and $(1-\alpha/2)$ quantiles of the posterior sample of r.

2.4 An Example

In this section we apply our Bayesian point and interval estimators to real data first described in Hildesheim, Mann, Brinton, Szklo, Reeves, and Rawls (1991) and later used in Boese (2003). The original study examined the relationship between exposure to the herpes simplex virus (HSV) and invasive cervical cancer (ICC). They used the western blot procedure as a fallible detector of HSV. A sub-sample of the women was also tested with the refined western blot procedure, which is a relatively accurate procedure and, thus, is treated here as infallible. We regard this sub-sample as the training data in the double-sampling scheme. Actually, both false-positive and false-negative misclassification errors occurred in this study. However, for the sake of illustration, we consider only the occurrence of the false positives and absorb the n_{i10} false-negative observations into the n_{i11} negative observations. We display the data in Table 2.4.

Using the posterior sampling algorithm developed in the previous section with a posterior sample size of $J = 10,000$, the estimated posterior median for r is 1.45 and an approximate 90% Bayesian credible interval is $[1.18, 1.80]$. Because the lower limit of the Bayesian credible interval exceeds 1.0, we conclude that a larger proportion of women have been exposed to HSV in the case group than in the control group. Thus, a positive association could exist between exposure to HSV and having the disease ICC.

2.5 Monte Carlo Simulations

We conducted two Monte Carlo simulation studies to examine the performance of our posterior sampling algorithm and to evaluate the impact of various proportion parameter values, sample sizes, and false-positive rates on the Bayesian credible intervals for r. For the sake of simplifying the design of the simulation and the presentation of the simulation results, we used $N = N_1 = N_2$, $n = n_1 = n_2$, and

Table 2.4. Hildesheim Example Data

			Fallible Method	
Group	Data	Infallible Method	0	1
Case	Training	0	13	3
		1	NA	23
	Original	NA	318	375
Control	Training	0	33	11
		1	NA	32
	Original	NA	701	535

NA: Not Available

$\phi = \phi_1 = \phi_2$, although these restrictions are not required to apply our credible interval.

We considered 32 simulation scenarios resulting from combinations of the following parameter and sample-size values:

(1) Proportion parameters of interest (p_1, p_2): (.1, .2), (.4, .6).

(2) False-positive rates ϕ: .1, .2.

(3) Ratios of training-sample size versus the total sample size n/N: 0.2, 0.4.

(4) Total sample sizes N: 100, 200, 300, 400.

For each simulation scenario, we simulated $K = 10,000$ data sets. Within each data set, we drew a posterior sample of r of size $J = 10,000$ using the posterior sampling algorithm described in Section 2.3. We then computed the posterior sample median and an approximate 90% credible interval for r. Finally, we generated boxplots of the K estimated posterior medians of r and calculated the coverage probability and the average length of the K credible intervals.

In Figures 2.1 and 2.2 we present the boxplots of K estimated posterior medians of r plotted against total sample size N. The actual proportion parameters (p_1, p_2) used were $(.1, .2)$ and $(.4, .6)$ for Figure 2.1 and Figure 2.2, respectively. In

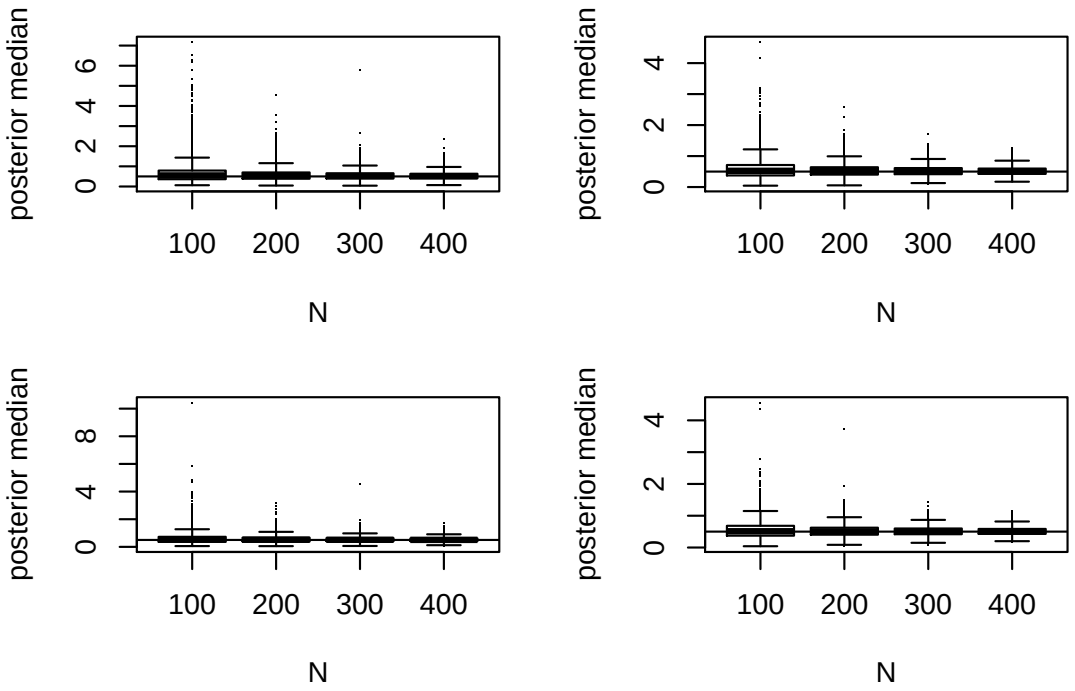

Figure 2.1: Boxplots of posterior medians versus total sample size N, where $(p_1, p_2) = (.1, .2)$. We assumed $\phi = .2$ for the simulations described in the top two panels and $\phi = .1$ for the simulation results shown in the bottom two panels; we used $n/N = .2$ in the simulations summarized in the left two panels and $n/N = .4$ in the simulations summarized in the right two panels.

each figure, the over-reporting parameter was $\phi = .2$ for the simulation results described in the top two panels and $\phi = .1$ for the simulation results shown in the bottom two panels. In addition, we used a training sample size to total sample size of $n/N = .2$ for the simulations in the left two panels and $n/N = .4$ for the simulations in the right two panels.

From the 32 simulation scenarios in both figures, the sample posterior medians performed well as a point estimator of the parameter r. Moreover, we make the following additional observations:

(1) For each panel of four boxplots, the variation of the posterior medians decreased as N increased.

(2) For each figure, the variation of the posterior medians with larger values of

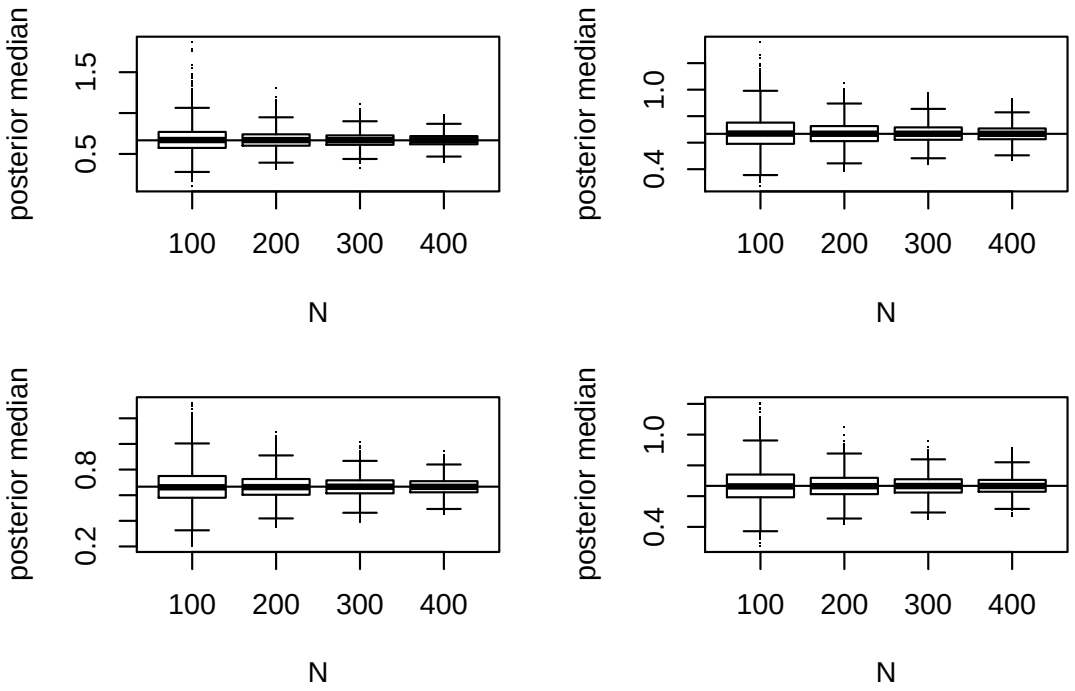

Figure 2.2: Boxplots of posterior medians versus total sample size N, where $(p_1, p_2) = (.4, .6)$. We assumed $\phi = .2$ for the simulations described in the top two panels and $\phi = .1$ for the simulation results shown in the bottom two panels; we used $n/N = .2$ in the simulations summarized in the left two panels and $n/N = .4$ in the simulations summarized in the right two panels.

ϕ, shown in the top two panels, was greater than posterior median variation with smaller values of ϕ, shown in the bottom two panels.

(3) For each figure, the variation of the posterior medians of the left two panels with smaller n/N was greater than that of the right two panels with larger n/N.

(4) For the same values of ϕ, n/N, N, the boxplot in Figure 2.1 with (p_1, p_2) close to (0, 0) has greater variation than the boxplot in Figure 2.2 with (p_1, p_2) close to (.5, .5).

In Table 2.5 we present the coverage probabilities and average lengths of the K credible intervals for r under each simulation scenario. The coverage probabilities for the three interval estimators were all close to the nominal 90% level. We make

Table 2.5. Coverage Probabilities and Average Lengths of 90% CIs for risk ratio r

					N			
r	(p_1, p_2)	ϕ	n/N		100	200	300	400
1/2	(.1, .2)	.2	.2	CP	92.00	90.80	90.31	90.21
				AL	1.97	0.97	0.74	0.61
			.4	CP	91.04	90.53	90.43	90.83
				AL	1.04	0.65	0.51	0.44
		.1	.2	CP	92.49	91.39	90.99	90.68
				AL	1.60	0.83	0.64	0.53
			.4	CP	90.85	90.26	90.52	90.14
				AL	0.94	0.59	0.47	0.40
2/3	(.4, .6)	.2	.2	CP	91.68	90.88	90.48	90.60
				AL	0.56	0.37	0.30	0.26
			.4	CP	90.54	90.68	89.65	90.00
				AL	0.42	0.29	0.23	0.20
		.1	.2	CP	93.34	91.28	90.92	90.95
				AL	0.50	0.33	0.26	0.23
			.4	CP	90.97	90.46	90.65	89.95
				AL	0.39	0.27	0.22	0.19

CP: Coverage Probability; AL: Average Length

the following observations on Table 2.5:

(1) For fixed (p_1, p_2), ϕ, n/N, the average length of the credible intervals decreased as N increased.

(2) For fixed (p_1, p_2), n/N, N, the average length of the credible intervals decreased as ϕ decreased.

(3) For fixed (p_1, p_2), ϕ, N, the average length of the credible intervals decreased as n/N increased.

(4) For the same values of ϕ, n/N, N, the average length of the credible intervals decreased as (p_1, p_2) approached $(.5, .5)$.

2.6 Discussion

In this paper we have proposed Bayesian credible interval estimators for the risk ratio using binomial data subject to false-positive misclassification. Monte Carlo simulations haven shown that our Bayesian credible intervals produced close-to-nominal coverage probabilities. The posterior median, our point estimator for r, also was well-behaved.

There are several advantages of our proposed posterior sampling algorithm for drawing from the full posterior distribution:

(1) We need not specify initial values and we had no burn-in period or convergence issues because we draw directly from the posterior distributions.

(2) Because posterior draws are available for each parameter, inferences on the risk difference, the odds ratio, and other functions of p_1 and p_2 are straightforward.

(3) As shown in (2.9) and (2.10), the algorithm can accommodate zero counts.

(4) We use no asymptotic theory.

(5) We can generalize our posterior sampling algorithm to data model with three or more proportion parameters.

CHAPTER THREE

Bayesian Inference of Risk Ratio of Two Proportions Using a Double Sampling Scheme

3.1 Introduction

In many disciplines, such as medical research, one obtains binary data that are possibly misclassified. For example, a healthy patient may be wrongly diagnosed as having a certain disease or vice versa. The consequence of ignoring misclassification in statistical inference from binary data was first reported by Bross (1954), who has shown that classical estimators based on only data subject to misclassification can be extremely biased. Also, the actual data model suffers from unidentifiability. Therefore, one requires additional data to obtain model identifiability and to correct the bias. One can utilize at least two methods to achieve model identifiability in the Bayesian paradigm. The first method, pioneered by Tenenbein (1970), is to collect training data using a double-sampling scheme; the other method is to use sufficiently informative priors specified by expert opinion or previous data. The rationale for Tenenbein's double-sampling scheme is as follows. Fallible tests or classification procedures result in misclassification but are usually inexpensive, while infallible tests or classification procedures result in errorless classification but are generally much more expensive. Therefore, the use of both fallible and infallible procedures yields not only model identifiability, but also economically viable bias correction. When an infallible procedure is unavailable or prohibitively expensive, one can use sufficiently informative priors to create an identifiable model in the Bayesian paradigm. For example, see Gustafson (2005).

We next review the literature on statistical inference using binary data with possible misclassification. For the single-proportion parameter problem, when data

are obtained using a double-sampling scheme, Tenenbein (1970) has proposed a maximum likelihood estimator and derived its asymptotic variance for the proportion parameter of interest. Also, Boese, Young, and Stamey (2006) have derived several likelihood-based confidence intervals for a single-proportion parameter using data subject to only false-positive misclassification. For the single-proportion problem using misclassified data with no training data, Gaba and Winkler (1992) and Viana, Ramakrishnan, and Levy (1993) have developed Bayesian approaches with highly informative priors. Bayesian inference using informative priors was also developed for two-sample problems for two-proportion parameters. For example, see Evans, Guttman, Haitovsky, and Swartz (1996) for the risk difference, that is, the difference of two-proportion parameters, and see Gustafson, Le, and Saskin (2001) for the odds ratio.

Although misclassified binary data are common in epidemiology, we remark that such data also frequently occur in device testing. For example, Zhong (2002) has studied the specificity and sensitivity of a fallible diagnostic test together with a gold standard. In addition, Stamey, Seaman, and Young (2007) have developed a Bayesian estimator of an intervention effect with pre- and post-misclassified binomial data. Also, one can obtain clinical trial binary data that contain misclassified observations. For example, Lyles, Lin, and Williamson (2004) have provided design and analytic considerations for single-armed clinical trial studies with misclassification of a repeated binary outcome.

The risk ratio, also known as the relative risk, is defined as the ratio of the proportion parameters from two groups. The risk ratio is commonly used in clinical trials and epidemiological studies to measure the relative frequency between two groups for a certain event of interest, such as an adverse event of interest. To date, in the literature, we have found no inference methods for the risk ratio for proportions using two-sample binary data subject to misclassification. Hence, we intend to fill

in this gap and therefore, in this article, we propose a Bayesian approach to solve this problem.

In Section 3.2 we describe the data, and in Section 3.3 we develop Bayesian models and relatively simple posterior sampling algorithms. In Section 3.4 we illustrate our sampling algorithms using real data. We then examine the performance of our Bayesian point and interval estimators in Section 3.5 and conclude with a brief discussion in Section 3.6.

3.2 The Data

In this section we consider the structure and parameterization for two-sample binary data subject to misclassification. The fallible data are obtained by using an imperfect classification device that yields both false-positive and false-negative observations. For example, suppose a study's objective is to assess if a certain disease has the same prevalence rates between men and women where a positive result of an assay indicates that a subject in the study has the disease. Because this assay does not have perfect sensitivity and specificity, both false-positive and false-negative observations can occur.

To describe the data, let F_{ij} be the observed classification by the fallible test or classification method for the jth individual in the ith sample, where $i = 1, 2$, $j = 1, \ldots, M_i$, and

$$F_{ij} = \begin{cases} 1, & \text{if the result is positive by the fallible classifier} \\ 0, & \text{if the result is negative by the fallible classifier.} \end{cases}$$

If one denotes the number of positive observations by X_i, the observed fallible data for sample i, $i = 1, 2$, are displayed in Table 3.1. Similarly, we define the unobserved true classification of the jth individual in the ith sample as T_{ij}, where

$$T_{ij} = \begin{cases} 1, & \text{if the result is truly positive} \\ 0, & \text{otherwise.} \end{cases}$$

Table 3.1. Fallible Data for Sample i, $i = 1, 2$.

Classification	0	1	Total
Count	$M_i - X_i$	X_i	M_i

Clearly, misclassification occurs when $T_{ij} \neq F_{ij}$. Next, we introduce the following notation for the ith sample, $i = 1, 2$. Let

$$\begin{aligned} p_i &\equiv \Pr(T_{ij} = 1), \\ \pi_i &\equiv \Pr(F_{ij} = 1), \\ \phi_i &\equiv \Pr(F_{ij} = 1 | T_{ij} = 0), \end{aligned}$$

and

$$\theta_i \equiv \Pr(F_{ij} = 0 | T_{ij} = 1).$$

We see that p_i is the actual proportion parameter of interest, π_i is the proportion parameter of the fallible test or classification method, and ϕ_i and θ_i are the false-positive rate and the false-negative rate for the fallible classification method, respectively. Here, we allow the false-positive and false-negative rates to differ between the two samples, i.e., $\phi_1 \neq \phi_2$ and $\theta_1 \neq \theta_2$. Note that π_i, $i = 1, 2$, are functions of the parameters p_i, θ_i, and ϕ_i. In particular,

$$\begin{aligned} \pi_i &= \Pr(T_{ij} = 1) \Pr(F_{ij} = 1 | T_{ij} = 1) + \Pr(T_{ij} = 0) \Pr(F_{ij} = 1 | T_{ij} = 0) \\ &= p_i(1 - \theta_i) + q_i \phi_i, \end{aligned} \tag{3.1}$$

where $q_i = 1 - p_i$.

As stated in Section 3.1, we are interested in statistical inference on the risk ratio

$$r \equiv p_1 / p_2. \tag{3.2}$$

Because π_i is determined through p_i, ϕ_i, and θ_i, $i = 1, 2$, six effective model parameters result. However, the sufficient-statistic dimension is only two because X_1 and

X_2, given in Table 3.1, are the sufficient statistics for the parameters in (3.1). Because the dimension of the sufficient statistics is less than the parameter dimension, (3.1) is unidentifiable. Therefore, we need additional data or information about the parameters to obtain model identifiability. For the Bayesian paradigm, two primary methods can provide additional information: the first is to obtain training data by using a double-sampling scheme, and the second is to incorporate prior knowledge about model parameters through sufficiently informative priors. In this paper we focus on data obtained via a double-sampling scheme.

To create model identifiability when one-sample binary data are subject to misclassification, Tenenbein (1970) has used additional training data obtained by double sampling. Specifically, in addition to the original data generated by only the fallible device, he has used training data obtained by classifying each individual in the training data using both the fallible and the infallible devices. The additional data enable the assessment of the false-positive rates ϕ_i and false-negative rates θ_i of the fallible device, $i = 1, 2$. One can find other applications of double sampling in Tenenbein (1972), Hochberg (1977), and Boese et al. (2006).

We apply the same double-sampling scheme to our two-proportion risk ratio problem to obtain n_i training data in addition to the original M_i fallible data for the ith sample, $i = 1, 2$. The combined data are presented in Table 3.2, where we use n_{ijk} to denote the number of individuals classified as j and k by the infallible and fallible labeling methods, respectively, where $j = 0, 1$ and $k = 0, 1$. For example, n_{i01} is the number of individuals classified as negative by the infallible method but positive by the fallible method in the ith sample. Using the combined fallible and training data, we can now estimate all parameters. For easy reference, we present the cell probabilities for Table 3.2 in Table 3.3.

Table 3.2. Data for Sample i

		Fallible Method		
Data	Infallible Method	0	1	Total
Training	0	n_{i00}	n_{i01}	$n_{i0\cdot}$
	1	n_{i10}	n_{i11}	$n_{i1\cdot}$
	Total	$n_{i\cdot 0}$	$n_{i\cdot 1}$	n_i
Original	NA	$M_i - X_i$	X_i	M_i

NA: Not Available

Table 3.3. Cell Probabilities for Sample i

		Fallible Method		
Data	Infallible Method	0	1	Total
Training	0	$q_i(1-\phi_i)$	$q_i\phi_i$	q_i
	1	$p_i\theta_i$	$p_i(1-\theta_i)$	p_i
Original	NA	$1-\pi_i$	π_i	1

NA: Not Available

3.3 Bayesian Inference

We now develop Bayesian point and interval estimators for the data described in the previous section. In particular, we derive algorithms to sample from the posterior distributions of each of the parameters. Once we draw a posterior sample for p_1 and p_2, we can readily obtain a posterior sample for the risk ratio r using (3.2). We determine a point estimator of r, which is the sample median from the posterior sample of r, and a credible interval for r. We next derive explicit algorithms for sampling from the marginal posterior distributions of interest.

For sample i, $i = 1, 2$, in Table 3.2, the observed counts $\boldsymbol{n}_i \equiv (n_{i00}, n_{i01}, n_{i10}, n_{i11})'$ of the training data have a quadrinomial distribution with total size n_i and cell probabilities displayed in the upper-right 2×2 submatrix in Table 3.3, i.e.,

$$\boldsymbol{n}_i | p_i, \phi_i, \theta_i \sim \text{Quad}\left[n_i, (q_i(1-\phi_i), q_i\phi_i, p_i\theta_i, p_i(1-\theta_i))\right].$$

In addition,

$$X_i|M_i, p_i, \phi_i, \theta_i \sim \text{Bin}(M_i, \pi_i), i = 1, 2.$$

Because $\boldsymbol{n}_i$ and X_i are independent for sample i, $i = 1, 2$, and $X_i, i = 1, 2$ are independent, the distribution of the combined data is

$$f(\boldsymbol{d}|\boldsymbol{\eta}) \propto \prod_{i=1}^{2} \{[q_i(1-\phi_i)]^{n_{i00}}(q_i\phi_i)^{n_{i01}}(p_i\theta_i)^{n_{i10}}[p_i(1-\theta_i)]^{n_{i11}}\pi_i^{X_i}(1-\pi_i)^{M_i-X_i}\}, \quad (3.3)$$

where

$$\boldsymbol{d} \equiv (n_{100}, n_{101}, n_{110}, n_{111}, X_1, n_{200}, n_{201}, n_{210}, n_{211}, X_2)' \quad (3.4)$$

is the data vector and

$$\boldsymbol{\eta} \equiv (p_1, \phi_1, \theta_1, p_2, \phi_2, \theta_2)'$$

is the parameter vector. For our Bayesian framework, we choose a uniform prior for each component of $\boldsymbol{\eta}$ and assume these priors were independent; therefore, the joint prior distribution is the noninformative prior

$$p(\boldsymbol{\eta}) = 1, \quad (3.5)$$

where $\eta_j \in (0, 1], j = 1, 2, \ldots, 6$. Combining (3.3) and (3.5), we obtain the joint posterior

$$f(\boldsymbol{\eta}|\boldsymbol{d}) \propto \Pi_{i=1}^{2}\{[q_i(1-\phi_i)]^{n_{i00}}(q_i\phi_i)^{n_{i01}}(p_i\theta_i)^{n_{i10}}[p_i(1-\theta_i)]^{n_{i11}}\pi_i^{X_i}(1-\pi_i)^{M_i-X_i}\}, \quad (3.6)$$

which has the same functional form as (3.3).

To sample from (3.6), we reparameterize $\boldsymbol{\eta}$ to obtain a parametric, straightforward posterior sampling algorithm. Specifically, let

$$\lambda_{i1} = p_i(1-\theta_i)/\pi_i \quad (3.7)$$

and

$$\lambda_{i2} = p_i\theta_i/(1-\pi_i). \quad (3.8)$$

Using (3.7) and (3.8), we see that the posterior density (3.6) becomes

$$f(\boldsymbol{\eta}^*|\boldsymbol{d}) \propto \Pi_{i=1}^{2} \lambda_{i1}^{n_{i11}}(1-\lambda_{i1})^{n_{i01}} \lambda_{i2}^{n_{i10}}(1-\lambda_{i2})^{n_{i00}} \pi_i^{X_i+n_{i.1}}(1-\pi_i)^{M_i-X_i+n_{i.0}}, \quad (3.9)$$

where $\boldsymbol{d}$ is defined in (3.4) and

$$\boldsymbol{\eta}^* \equiv (\lambda_{11}, \lambda_{12}, \pi_1, \lambda_{21}, \lambda_{22}, \pi_2)' \quad (3.10)$$

is the reparameterized parameter vector. Because the parameters in (3.10) are now separable, we can straightforwardly draw λ_{i1}, λ_{i2}, and π_i from the posterior (3.9) by using

$$\lambda_{i1} \sim \text{Beta}(n_{i11}+1, n_{i01}+1), \quad (3.11)$$

$$\lambda_{i2} \sim \text{Beta}(n_{i10}+1, n_{i00}+1), \quad (3.12)$$

and

$$\pi_i \sim \text{Beta}(X_i+n_{i.1}+1, M_i-X_i+n_{i.0}+1), \quad (3.13)$$

$i = 1, 2$. Once λ_{i1}, λ_{i2}, and π_i are available, we obtain p_i, ϕ_i, and θ_i by solving (3.1), (3.7), and (3.8) so that

$$p_i = \pi_i\lambda_{i1} + (1-\pi_i)\lambda_{i2}, \quad (3.14)$$

$$\phi_i = (1-\lambda_{i1})\pi_i/q_i, \quad (3.15)$$

and

$$\theta_i = \lambda_{i2}(1-\pi_i)/p_i, \quad (3.16)$$

$i = 1, 2$.

In summary, the following is a parametric-based algorithm to sample from the marginal posterior density $p(r|\boldsymbol{d})$. First, choose a large number J, say 10,000, for the posterior draw sample size. For $i = 1, 2$,

Table 3.4. Hildesheim Example Data

			Fallible Method	
Group	Data	Infallible Method	0	1
Case	Training	0	13	3
		1	5	18
	Original	NA	318	375
Control	Training	0	33	11
		1	16	16
	Original	NA	701	535

NA: Not Available

(1) Obtain a sample of λ_{i1}, λ_{i2}, and π_i with a sample size of J using (3.11) – (3.13).

(2) Obtain a sample of p_i, ϕ_i, and θ_i with a sample size of J using (3.14) – (3.16).

(3) Obtain a sample of the risk ratio r with a sample size of J using (3.2).

We then use the median of the J observations from $p(r|\boldsymbol{d})$ as a point estimator of r because the marginal posterior density $p(r|\boldsymbol{d})$ is skewed. Finally, we obtain an approximate $100(1-\alpha)\%$ credible interval for r by using the $\alpha/2$ and $(1-\alpha/2)$ quantiles of the sample drawn from $p(r|\boldsymbol{d})$.

3.4 An Example

In this section we apply our Bayesian point and interval estimators to real data that were first described in Hildesheim et al. (1991) and later were analyzed in Boese (2003). This study explored the association between exposure to herpes simplex virus (HSV) and having invasive cervical cancer (ICC). The data are displayed in Table 3.4. This data consist of a total of 2044 women with 732 women in the case group and 1312 women in the control group. The western blot procedure is a fallible detector of HSV. A sub-sample of the women was also tested with the refined western

blot procedure, which is a relatively accurate procedure and, thus, treated here as an infallible classifier. We regard this sub-sample as the training data in the double-sampling scheme. Both false-positive and false-negative misclassification errors for testing for HSV using the western blot procedure occurred in this study. Here, p_1 = Pr(exposed to HSV | has ICC) is the probability that a patient has been exposed to HSV given she has ICC (case group), p_2 = Pr(exposed to HSV | does not have ICC) is the probability that a patient has been exposed to HSV given she does not have ICC (control group), and the risk ratio r is given in (3.2). Using the sampling algorithm developed in the previous section with a posterior sample size J=10,000, we determine that $\hat{r}_{Bayes} = 1.34$, and an approximate 90% Bayesian credible interval for r is

$$[1.01, 1.75]\,. \tag{3.17}$$

We interpret (3.17) to indicate that $\Pr(r \in [1.01, 1.75]|\boldsymbol{d}) \approx 0.90$. Because the lower limit of (3.17) exceeds 1.0, we believe that statistical evidence indicates that a larger proportion of women have been exposed to HSV in the case group than in the control group. Thus, a positive association could conceivably exist between exposure to HSV and actually having ICC.

3.5 Monte Carlo Simulations

In this section we describe the results of Monte Carlo simulation studies that examine the performance of our sampling algorithms for varying sample sizes, false-positive rates, and false-negative rates. For the sake of simplifying the presentation of simulation results, we assumed the sample size and parameter configurations of the form $N = N_1 = N_2$, $n = n_1 = n_2$, $\phi = \phi_1 = \phi_2$, and $\theta = \theta_1 = \theta_2$. We remark that these assumptions are not required by our posterior sampling algorithms.

We considered 32 simulation scenarios resulting from combinations of the following configurations:

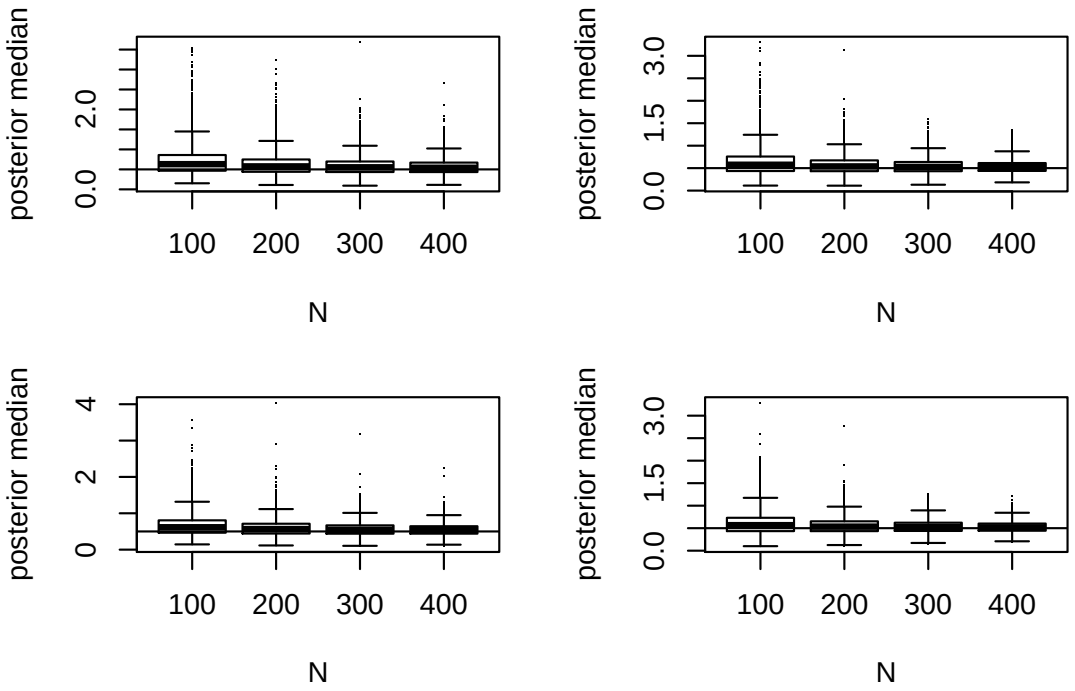

Figure 3.1: Boxplots of posterior medians versus total sample size N where $(p_1, p_2) = (.1, .2)$. We used $(\phi, \theta) = (.2, .2)$ for the simulations summarized in the top two panels and $(\phi, \theta) = (.1, .1)$ for the simulations described in the bottom two panels; we assumed $n/N = .2$ for the simulations summarized in the left two panels and $n/N = .4$ for the simulations described in the right two panels.

(1) Proportion parameters of interest (p_1, p_2): (.1, .2), (.4, .6).

(2) False-positive and false-negative rates (ϕ, θ): (.1, .1), (.2, .2).

(3) Ratio of training-sample size versus the total sample size (n/N): 0.2, 0.4.

(4) Total sample sizes (N): 100, 200, 300, 400.

For each simulation scenario, we simulated $K = 10,000$ data sets. For each data set, we drew $J = 10,000$ samples of r from the marginal posterior density using the algorithm described in Section 3.3. We then computed the posterior sample median to estimate r and calculated an approximate 90% credible interval for r. We next generated boxplots of the K estimated posterior medians of r to examine their behavior as a point estimator of r. In addition, we calculated the coverage probability and the average length of the K credible intervals.

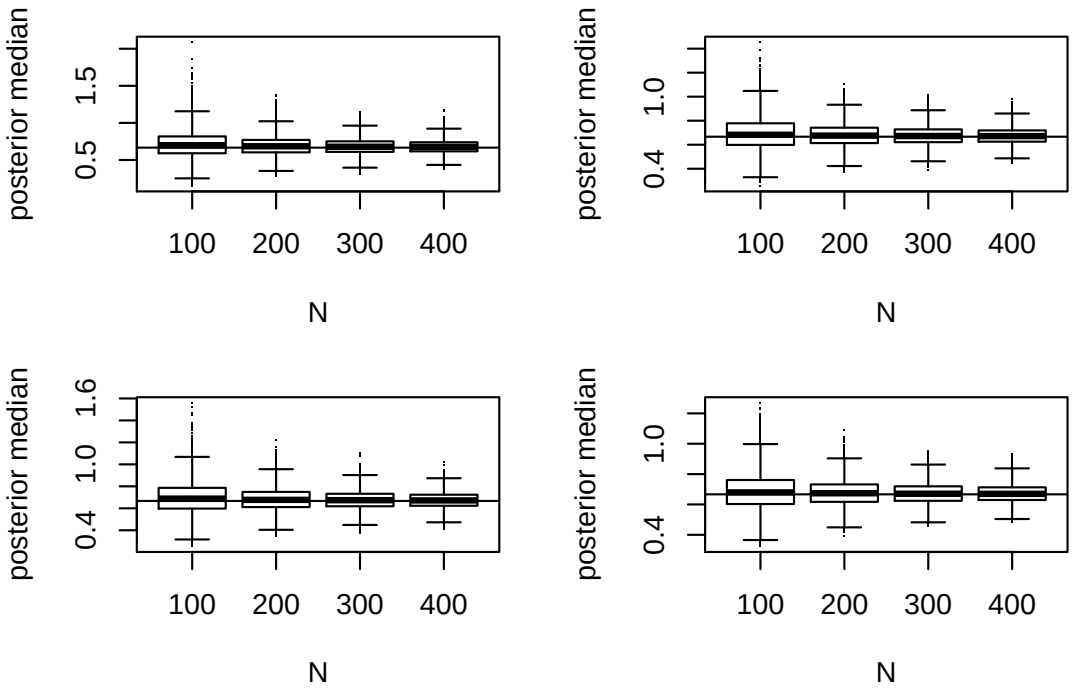

Figure 3.2: Boxplots of posterior medians versus total sample size N where $(p_1, p_2) = (.4, .6)$. We used $(\phi, \theta) = (.2, .2)$ for the simulations summarized in the top two panels and $(\phi, \theta) = (.1, .1)$ for the simulations described in the bottom two panels; we assumed $n/N = .2$ for the simulations summarized in the left two panels and $n/N = .4$ for the simulations described in the right two panels.

In Figures 3.1 and 3.2, we present the boxplots of K posterior sample medians of r versus the total sample size N. The actual proportion parameters were $(p_1, p_2) = (.1, .2)$ and $(.4, .6)$ for Figures 3.1 and 3.2, respectively. In each figure, we used $(\phi, \theta) = (.2, .2)$ in the top two panels and $(\phi, \theta) = (.1, .1)$ in the bottom two panels; we assumed $n/N = .2$ for the simulations results in the left two panels and $n/N = .4$ for the simulation results in the right two panels.

For the 32 simulation configurations for both figures, we have the following observations:

(1) For each panel of 4 boxplots, the variation of the posterior medians decreased as N increased.

(2) For each figure, the variation of the posterior medians, shown in the top

two panels, which have larger misclassification probabilities ϕ and θ, was greater than that of the bottom two panels with smaller misclassification probabilities.

(3) For each figure, the variation of the posterior medians shown in the left two panels for smaller n/N was greater than the variation of the posterior medians depicted in the right two panels for larger n/N.

(4) For the same values of ϕ, θ, n/N, and N, the boxplots in Figure 3.1 had greater variation than the corresponding boxplots in Figure 3.2.

In Table 3.5 we present the coverage probabilities, average lengths, and standard deviations of the K credible intervals for r under each simulation parameter configuration. The coverage probabilities were all close to the nominal 90% level, thus indicating that our credible interval estimator performed well for the parameter and sample-size configurations considered here. Also, we have the following observations concerning Table 3.5:

(1) For fixed p_1, p_2, ϕ, θ, and n/N, the average length and standard deviation of the credible intervals decreased as N increased.

(2) For fixed p_1, p_2, n/N, and N, the average length and standard deviation of the credible intervals decreased as ϕ and θ decreased.

(3) For fixed p_1, p_2, ϕ, θ, and N, the average length and standard deviation of the credible intervals decreased as n/N increased.

(4) For the same ϕ, θ, n/N, and N, the average length and standard deviation of the credible intervals decreased when (p_1, p_2) approached $(.5, .5)$.

Table 3.5. The CP, AL, and SD of the ALs of 90% CIs for the risk ratio r

					N			
r	(p_1, p_2)	(ϕ, θ)	n/N		100	200	300	400
1/2	(.1, .2)	(.2, .2)	.2	CP	92.84	91.77	91.33	90.96
				AL	1.66	1.03	0.79	0.67
				SD	1.23	0.54	0.35	0.22
			.4	CP	91.90	90.84	90.67	90.93
				AL	1.07	0.69	0.54	0.46
				SD	0.58	0.24	0.15	0.10
		(.1, .1)	.2	CP	93.72	93.17	92.19	91.69
				AL	1.46	0.90	0.70	0.58
				SD	0.95	0.44	0.27	0.17
			.4	CP	92.64	91.62	91.00	90.65
				AL	0.97	0.63	0.49	0.42
				SD	0.48	0.19	0.12	0.09
2/3	(.4, .6)	(.2, .2)	.2	CP	91.84	90.49	90.16	90.80
				AL	0.65	0.45	0.36	0.31
				SD	0.18	0.08	0.05	0.04
			.4	CP	90.78	90.58	89.82	89.99
				AL	0.47	0.33	0.26	0.23
				SD	0.09	0.04	0.03	0.02
		(.1, .1)	.2	CP	93.46	92.31	91.56	91.23
				AL	0.57	0.38	0.31	0.26
				SD	0.13	0.06	0.04	0.03
			.4	CP	91.98	91.00	90.42	89.99
				AL	0.42	0.29	0.24	0.20
				SD	0.07	0.03	0.02	0.02

3.6 Discussion

In this paper we have derived Bayesian point and interval estimators for the risk ratio of two proportions using binary data subject to both false-positive and false-negative misclassification. Monte Carlo simulations have shown that our inference algorithms produced credible intervals with near-nominal coverage probabilities. Also, the estimated posterior median performed well as a point estimator of r.

Several advantages of our straightforward posterior sampling algorithms are apparent for sampling from the joint posterior distribution:

(1) Because we draw directly from the posterior distributions, we need not specify initial values, and no burn-in period or convergence issues occur.

(2) Because posterior draws are available for each parameter, inferences on the risk difference, the odds ratio, and other functions of p_1 and p_2 are straightforward.

(3) As shown in (3.11)-(3.13), the posterior sampling algorithms can accommodate zero counts.

(4) We use no asymptotic theory.

(5) We can generalize our posterior sampling algorithms to data for more than two groups of observations.

Some applications exist where no training data are available. In these cases, if sufficiently informative priors are obtainable, then more sophisticated Monte Carlo Markov Chain (MCMC) methods may be needed to sample from the posterior of interest instead of the straightforward Monte Carlo posterior sampling method proposed here.

CHAPTER FOUR

Likelihood-Based Confidence Intervals for the Risk Ratio Using Double Sampling with Over-Reported Binary Data

4.1 Introduction

Binary data are sometimes obtained when experimental units are classified into two mutually exclusive categories. Generally, a classifier is not perfect and, therefore, misclassified binary data can occur. Two types of misclassified binary data exist: false-positive and false-negative observations. For example, visual inspection by a midwife or obstetrician may erroneously classify a normal child as having Down's syndrome (false-positive), or it may classify a child with Down's syndrome as being healthy (false-negative). In other cases, only one type of misclassification may exist. For instance, Perry, Vakil, and Cutler (2000) have displayed blood testing data that had only false-positive or over-reported errors, and Moors, van der Genugten, and Strijbosch (2000) have presented auditing data indicating only false-negative or under-reported errors to be present.

Many researchers, including Bross (1954), have demonstrated that classical estimators that ignore misclassifications are biased when applied to misclassified binary data. Therefore, additional external information or additional data are needed to correct the bias. Several methods in the literature are dedicated to this purpose. In the Bayesian paradigm, when an infallible procedure is unavailable or prohibitively expensive, one can use informative priors to obtain model identifiability. Another information-producing method is to use multiple fallible classifiers. The method we focus on in this article uses additional training data via the double-sampling scheme first proposed by Tenenbein (1970). Tenenbein's double-sampling method is used when infallible and fallible classification procedures are available. Usually, the infal-

lible procedure is very expensive; the fallible procedure is usually cheap but prone to error. Therefore, the use of an infallible procedure on only a small portion of the data and the use of a fallible procedure on all the data are an economically feasible means of promoting model identifiability.

A rich literature of research is available on binary data subject to misclassification that provides point and interval estimation methods on various functions of the proportion parameters of interest. For one-sample problems, several researchers have considered the case in which only one type of error is present. Lie, Heuch, and Irgens (1994) have used a maximum likelihood approach, where false-negative errors were corrected using multiple 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-based confidence intervals (CIs) for a single-proportion parameter, while Lee and Byun (2008) have provided Bayesian credible intervals using noninformative priors for the same problem.

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

For two-sample problems with both types of misclassification errors, Bayesian inference methods using sufficiently informative priors have also been developed when training data are unavailable. For example, see Evans, Guttman, Haitovsky,

and Swartz (1996) for risk difference estimation, that is, the difference of two-proportion parameters, and Gustafson, Le, and Saskin (2001) for odds ratios. When training data are obtained through a double-sampling scheme, Boese (2003) has derived several likelihood-based CIs for the risk difference.

To date, no methods for inference on risk ratio exist in the literature for two-sample misclassified binary data. The risk ratio (RR) is defined as the ratio of two-proportion parameters and is also known as the relative risk. In this article we focus on data subject to only one type of misclassification error. Without loss of generality, we consider data with false-positive errors only.

The remainder of this paper is organized as follows. In Section 4.2 we describe the data, and in Section 4.3 we propose three likelihood-based methods for interval estimation of a risk ratio using double sampling with over-reported data. In Section 4.4 we illustrate the newly derived interval estimation methods using real cervical cancer data. We examine the performance of three proposed interval estimation methods in Section 4.5 using Monte Carlo simulation, and we provide a brief discussion in Section 4.6.

4.2 The Data

In this section we describe two-sample misclassified binary data with one type of misclassification error. We assume that the data is obtained using a fallible classification procedure that yields false-positive but not false-negative counts.

We next introduce notation useful for describing the data. Let F_{ij} be the observed classification by the fallible classification for the jth individual in the ith sample, where $i = 1, 2$, $j = 1, \ldots, M_i$, and

$$F_{ij} = \begin{cases} 1, & \text{if the result is positive by the fallible classifier} \\ 0, & \text{otherwise.} \end{cases}$$

Let $X_i = \sum_j F_{ij}$ and $Y_i = M_i - X_i$ be the observed number of positive and negative

Table 4.1. Data from the Fallible Method for Sample i, $i = 1, 2$

Classification	0	1	Total
Count	Y_i	X_i	M_i

classifications, respectively. The data obtained by the fallible classification for sample i, $i = 1, 2$, are displayed in Table 4.1. Similarly, we define the unobserved true classification of the jth individual in the ith sample as

$$T_{ij} = \begin{cases} 1, & \text{if the result is truly positive} \\ 0, & \text{otherwise.} \end{cases}$$

We remark that misclassification occurs when $T_{ij} \neq F_{ij}$.

Also, we let

$$\begin{aligned} p_i &\equiv \Pr(T_{ij} = 1), \\ \pi_i &\equiv \Pr(F_{ij} = 1), \end{aligned}$$

and

$$\phi_i \equiv \Pr(F_{ij} = 1 | T_{ij} = 0).$$

Here, p_i is the actual proportion parameter of interest, π_i is the proportion parameter of the fallible procedure, and ϕ_i is the false-positive rate for the fallible procedure. We allow the false-positive rates to be different between the two samples, i.e., $\phi_1 \neq \phi_2$. Note that π_1 and π_2 are not additional unique parameters because

$$\begin{aligned} \pi_i &= \Pr(T_i = 1) \Pr(F_i = 1 | T_i = 1) + \Pr(T_i = 0) \Pr(F_i = 1 | T_i = 0) \\ &= p_i + q_i \phi_i, \end{aligned} \tag{4.1}$$

where $q_i = 1 - p_i$, $i = 1, 2$.

As noted in Section 4.1, we desire to determine point and interval estimators of the risk ratio

$$r = p_1 / p_2. \tag{4.2}$$

Because π_i is a function of p_i and ϕ_i, effectively four unique parameters exist in the model: p_1, ϕ_1, p_2, ϕ_2. However, the sufficient statistic dimension is only two because (X_1, X_2) are sufficient statistics for this model. Thus, because the dimension of the sufficient statistics is less than the number of parameters, the four parameters in model (4.1) are unidentifiable and, therefore, additional data are needed for model identifiability. In this paper we use double sampling to provide additional information.

Tenenbein (1970) has initiated a double-sampling scheme to promote model identifiability while controlling the cost. Specifically, in addition to the original fallibly classified data, a new, smaller training data set is obtained by classifying each individual using both the fallible and the infallible procedures. The double-sampling scheme has attracted researchers' attention because of its practicality. Other applications of double-sampling schemes have been given in Tenenbein (1972), Hochberg (1977), and Boese et al. (2006).

In this paper we assume that training data of size n_i are obtained using a double-sampling scheme in addition to the original fallible data of size M_i for the ith sample. Hence, the size of the combined data is $N_i = M_i + n_i$. Table 4.2 presents the combined data. In Table 4.2 we use n_{ijk} to denote the number of individuals classified as j and k by the infallible and fallible classification procedures, respectively. For example, n_{i01} is the number of individuals in the ith sample classified as negative by the infallible procedure but classified as positive by the fallible procedure. Note that n_{i10} is not available because false-negative errors are assumed non-existent. With the additional training data, the dimension of sufficient statistics for the combined data is sufficient to estimate all parameters and, therefore, the model is now identifiable. For future estimation methodology development, we give the cell probabilities corresponding to Table 4.2 in Table 4.3.

Table 4.2. Data for Sample i

		Fallible Procedure		
Data	Infallible Procedure	0	1	Total
Training	0	n_{i00}	n_{i01}	$n_{i0\cdot}$
	1	NA	n_{i11}	n_{i11}
	Total	n_{i00}	$n_{i\cdot 1}$	n_i
Original	NA	Y_i	X_i	M_i

NA: Not Available

Table 4.3. Cell Probabilities for Sample i

		Fallible Procedure		
Data	Infallible Procedure	0	1	Total
Training	0	$q_i(1-\phi_i)$	$q_i\phi_i$	q_i
	1	NA	p_i	p_i
Original	NA	$1-\pi_i$	π_i	1

NA: Not Available

4.3 The Model

In this section we develop point and interval estimators for the risk ratio (4.2) of two-proportion parameters using double sampling with over-reported data. In particular, we provide formulas for maximum likelihood estimators (MLE) based on the full likelihood function. In addition, we develop two Wald-based CIs and a Fieller-type CI based on the full likelihood.

4.3.1 The Full Likelihood Function

The data, obtained using a double-sampling scheme, are presented in Table 4.2. For sample i, the observed counts $(n_{i00}, n_{i01}, n_{i11})$ of the training data have a trinomial distribution with total size n_i and probabilities displayed in an upper-right 2×2 submatrix in Table 4.3, i.e.,

$$(n_{i00}, n_{i01}, n_{i11})|p_i, \phi_i \sim \text{Trin}[n_i, (q_i(1-\phi_i), q_i\phi_i, p_i)]. \tag{4.3}$$

In addition, the observed counts (X_i, Y_i) have the binomial distribution

$$(X_i, Y_i)|p_i, \phi_i \sim \text{Bin}[M_i, (\pi_i, 1-\pi_i)]. \tag{4.4}$$

Because $(n_{i00}, n_{i01}, n_{i11})$ and (X_i, Y_i) are independent for sample i, $i = 1, 2$, and sample 1 is independent of sample 2, the density function of the data vector given the parameter vector is

$$f(\boldsymbol{d}|\boldsymbol{\eta}) \propto \prod_{i=1}^{2} \{[q_i(1-\phi_i)]^{n_{i00}} (q_i\phi_i)^{n_{i01}} p_i^{n_{i11}} \pi_i^{X_i} (1-\pi_i)^{Y_i}\}, \tag{4.5}$$

where

$$\boldsymbol{d} = (n_{100}, n_{101}, n_{111}, X_1, Y_1, n_{200}, n_{201}, n_{211}, X_2, Y_2)' \tag{4.6}$$

and

$$\boldsymbol{\eta} = (p_1, \phi_1, p_2, \phi_2)'.$$

Therefore, we can express the full likelihood as

$$L_f(\boldsymbol{\eta}) \propto \prod_{i=1}^{2} \{[q_i(1-\phi_i)]^{n_{i00}} (q_i\phi_i)^{n_{i01}} p_i^{n_{i11}} \pi_i^{X_i} (1-\pi_i)^{Y_i}\}. \tag{4.7}$$

4.3.2 Full Likelihood MLEs

To determine the MLE of the risk ratio (4.2), we first perform a reparameterization of parameters $\boldsymbol{\eta}$ and then derive closed-form solutions. Let

$$\lambda_i \equiv p_i/\pi_i, \tag{4.8}$$

and let $\boldsymbol{\gamma} \equiv (\lambda_1, \pi_1, \lambda_2, \pi_2)'$, where π_i, $i = 1, 2$, is given in (4.1). Using (4.1) and (4.8), we see that the full likelihood in (4.7) can be re-expressed as

$$L_f(\boldsymbol{\gamma}) \propto \prod_{i=1}^{2} \left[\lambda_i^{n_{i11}} (1-\lambda_i)^{n_{i01}} \pi_i^{X_i+n_{i.1}} (1-\pi_i)^{Y_i+n_{i00}}\right]. \tag{4.9}$$

Therefore, the full log likelihood is

$$\begin{aligned} l_f(\boldsymbol{\gamma}) \propto & \sum_{i=1}^{2} [n_{i11} \log \lambda_i + n_{i01} \log(1-\lambda_i) + (X_i + n_{i.1}) \log \pi_i \\ & + (Y_i + n_{i00}) \log(1-\pi_i)], \end{aligned} \tag{4.10}$$

and the corresponding score vector is

$$\begin{aligned} \boldsymbol{s}_f(\boldsymbol{\gamma}) &\equiv \frac{\partial l_f(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \\ &= \left[\frac{n_{111}}{\lambda_1} - \frac{n_{101}}{1-\lambda_1}, \frac{X_1 + n_{1.1}}{\pi_1} - \frac{Y_1 + n_{100}}{1-\pi_1}, \right. \\ &\quad \left. \frac{n_{211}}{\lambda_2} - \frac{n_{201}}{1-\lambda_2}, \frac{X_2 + n_{2.1}}{\pi_2} - \frac{Y_2 + n_{200}}{1-\pi_2}\right]'. \end{aligned} \tag{4.11}$$

We obtain the MLE for $\boldsymbol{\gamma}$ by setting $\boldsymbol{s}_f(\boldsymbol{\gamma}) = \mathbf{0}$ and solving for λ_i and π_i so that

$$\hat{\lambda}_i = \frac{n_{i11}}{n_{i\cdot 1}}$$

and

$$\hat{\pi}_i = \frac{X_i + n_{i\cdot 1}}{N_i}.$$

By the MLE invariance property,

$$\hat{p}_i = \hat{\pi}_i \hat{\lambda}_i,$$

$$\hat{\phi}_i = (1 - \hat{\lambda}_i)\hat{\pi}_i / \hat{q}_i,$$

and

$$\hat{r} = \hat{p}_1 / \hat{p}_2. \tag{4.12}$$

4.3.3 The Full Likelihood Information Matrix

From (4.11), the Hessian matrix is

$$\begin{aligned} \boldsymbol{H}_f(\boldsymbol{\gamma}) = \text{Diag} \left[-\frac{n_{111}}{\lambda_1^2} - \frac{n_{101}}{(1-\lambda_1)^2}, -\frac{X_1 + n_{1.1}}{\pi_1^2} - \frac{Y_1 + n_{100}}{(1-\pi_1)^2}, \right. \\ \left. -\frac{n_{211}}{\lambda_2^2} - \frac{n_{201}}{(1-\lambda_2)^2}, -\frac{X_2 + n_{2.1}}{\pi_2^2} - \frac{Y_2 + n_{200}}{(1-\pi_2)^2}\right]. \end{aligned} \tag{4.13}$$

Thus, the expected Fisher information matrix is

$$\boldsymbol{I}_f(\boldsymbol{\gamma}) = \text{Diag} \left[\frac{n_1 \pi_1}{\lambda_1(1-\lambda_1)}, \frac{N_1}{\pi_1(1-\pi_1)}, \frac{n_2 \pi_2}{\lambda_2(1-\lambda_2)}, \frac{N_2}{\pi_2(1-\pi_2)}\right].$$

Because the necessary regularity conditions are satisfied for this model, the MLE vector $\hat{\boldsymbol{\gamma}} = (\hat{\lambda}_1, \hat{\pi}_1, \hat{\lambda}_2, \hat{\pi}_2)'$ has an asymptotic multivariate normal distribution with asymptotic mean $\boldsymbol{\gamma}$ and asymptotic covariance matrix

$$\boldsymbol{I}_f^{-1}(\boldsymbol{\gamma}) = \text{Diag}\left[\frac{\lambda_1(1-\lambda_1)}{n_1\pi_1}, \frac{\pi_1(1-\pi_1)}{N_1}, \frac{\lambda_2(1-\lambda_2)}{n_2\pi_2}, \frac{\pi_2(1-\pi_2)}{N_2}\right].$$

Thus, for $i = 1, 2$, we have

$$\begin{aligned} V(\hat{\lambda}_i) &= \frac{\lambda_i(1-\lambda_i)}{n_i\pi_i}, \\ V(\hat{\pi}_i) &= \frac{\pi_i(1-\pi_i)}{N_i}, \end{aligned}$$

and that $\hat{\lambda}_i, \hat{\pi}_i$, $i = 1, 2$, are mutually asymptotically uncorrelated.

4.3.4 A Full Likelihood Naive Wald CI

We begin with constructing a naive Wald (nWald) confidence interval for the risk ratio r. Note that $\hat{p}_i = \hat{\pi}_i\hat{\lambda}_i$ and $\hat{\pi}_i$ and $\hat{\lambda}_i$ are uncorrelated, $i = 1, 2$. Thus, by the delta method, we have

$$\begin{aligned} \sigma_i^2 \equiv V(\hat{p}_i) &\approx \left(\frac{\partial p_i}{\partial \pi_i}\right)^2 V(\hat{\pi}_i) + \left(\frac{\partial p_i}{\partial \lambda_i}\right)^2 V(\hat{\lambda}_i) \\ &= \frac{\lambda_i^2\pi_i(1-\pi_i)}{N_i} + \frac{\pi_i\lambda_i(1-\lambda_i)}{n_i}. \end{aligned} \tag{4.14}$$

The MLEs $\hat{\lambda}_i$ and $\hat{\pi}_i$ are consistent estimators of λ_i and π_i. Because continuous functions of consistent estimators are consistent, we have that a consistent estimator of (4.14) is

$$\hat{\sigma}_i^2 = \frac{\hat{\lambda}_i^2\hat{\pi}_i(1-\hat{\pi}_i)}{N_i} + \frac{\hat{\pi}_i\hat{\lambda}_i(1-\hat{\lambda}_i)}{n_i}. \tag{4.15}$$

Recall that the MLE of r is $\hat{r} = \hat{p}_1/\hat{p}_2$. Again using the delta method, we have

$$\begin{aligned} \sigma_r^2 \equiv V(\hat{r}) &\approx \left(\frac{\partial r}{\partial p_1}\right)^2 V(\hat{p}_1) + \left(\frac{\partial r}{\partial p_2}\right)^2 V(\hat{p}_2) \\ &= \frac{\sigma_1^2}{p_2^2} + \frac{p_1^2\sigma_2^2}{p_2^4}, \end{aligned} \tag{4.16}$$

and a consistent estimator of (4.16) is

$$\hat{\sigma}_r^2 = \frac{\hat{\sigma}_1^2}{\hat{p}_2^2} + \frac{\hat{p}_1^2\hat{\sigma}_2^2}{\hat{p}_2^4}. \tag{4.17}$$

Therefore, an approximate $100(1-\alpha)\%$ nWald CI for r is

$$\hat{r} \pm Z_{\alpha/2}\hat{\sigma}_r, \tag{4.18}$$

where $Z_{\alpha/2}$ is the upper $(\alpha/2)^{th}$ quantile of the standard normal distribution. We refer to (4.18) as a naive Wald CI because the lower limit of the CI can be negative, especially when the sample size is small and r is close to zero. If the lower limit of the CI is negative, we replace the lower limit by zero.

4.3.5 A Full Likelihood Modified Wald CI

To alleviate under-coverage and negative-endpoint problems with the nWald CI, we propose a modified Wald (mWald) CI by constructing an approximate $100(1-\alpha)\%$ Wald CI for $\tau = \log r$ that we exponentiate to obtain an approximate $100(1-\alpha)\%$ CI for r. Hong, Meeker, and Escobar (2008) also suggested using transformation of parameters when constructing Wald-type CIs. Specifically, let $\hat{\tau} = \log \hat{r}$. Then, using the delta method, we have

$$\begin{aligned}
\sigma_\tau^2 \equiv V(\hat{\tau}) &\approx V(\log\hat{\pi}_1 + \log\hat{\lambda}_1 - \log\hat{\pi}_2 - \log\hat{\lambda}_2) \\
&= \pi_1^{-2}V(\hat{\pi}_1) + \lambda_1^{-2}V(\hat{\lambda}_1) + \pi_2^{-2}V(\hat{\pi}_2) + \lambda_2^{-2}V(\hat{\lambda}_2) \\
&= \frac{1-\pi_1}{N_1\pi_1} + \frac{1-\lambda_1}{n_1\lambda_1\pi_1} + \frac{1-\pi_2}{N_2\pi_2} + \frac{1-\lambda_2}{n_2\lambda_2\pi_2}.
\end{aligned} \tag{4.19}$$

A consistent estimator of (4.19) is

$$\hat{\sigma}_\tau^2 = \frac{1-\hat{\pi}_1}{N_1\hat{\pi}_1} + \frac{1-\hat{\lambda}_1}{n_1\hat{\lambda}_1\hat{\pi}_1} + \frac{1-\hat{\pi}_2}{N_2\hat{\pi}_2} + \frac{1-\hat{\lambda}_2}{n_2\hat{\lambda}_2\hat{\pi}_2}. \tag{4.20}$$

Therefore, an approximate $100(1-\alpha)\%$ Wald-type CI for τ is

$$\hat{\tau} \pm Z_{\alpha/2}\hat{\sigma}_\tau. \tag{4.21}$$

Finally, we determine an approximate $100(1-\alpha)\%$ mWald CI for r by exponentiating the endpoints of (4.21) so that we obtain

$$\left[\hat{r}/\exp(Z_{\alpha/2}\hat{\sigma}_\tau), \hat{r}\exp(Z_{\alpha/2}\hat{\sigma}_\tau)\right]. \tag{4.22}$$

Note that the mWald CI guarantees that the lower limit of (4.22) is nonnegative.

4.3.6 A Full Likelihood Fieller-Type CI

We next derive a CI for r based on an interval estimation concept introduced in Fieller (1954). As noted previously, asymptotically, we have

$$\hat{p}_i \sim N(p_i, \sigma_i^2).$$

Because

$$\frac{\hat{p}_1 - r\hat{p}_2}{\sqrt{\sigma_1^2 + r^2\sigma_2^2}} \sim N(0,1)$$

is an asymptotic pivotal quantity, we can obtain an approximate $100(1-\alpha)\%$ Fieller CI by solving

$$\frac{(\hat{p}_1 - r\hat{p}_2)^2}{\hat{\sigma}_1^2 + r^2\hat{\sigma}_2^2} = Z_{\alpha/2}^2$$

for r. Let

$$\Delta \equiv \hat{p}_1^2\hat{p}_2^2 - (\hat{p}_1^2 - Z_{\alpha/2}^2\hat{\sigma}_1^2)(\hat{p}_2^2 - Z_{\alpha/2}^2\hat{\sigma}_2^2).$$

Because

$$\hat{p}_1^2 \geq \hat{p}_1^2 - Z_{\alpha/2}^2\hat{\sigma}_1^2 \text{ and } \hat{p}_2^2 \geq \hat{p}_2^2 - Z_{\alpha/2}^2\hat{\sigma}_2^2,$$

we have $\Delta \geq 0$. Moreover, $\Delta = 0$ if and only if $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = 0$. Clearly, when $\Delta = 0$, Fieller CI does not exist, which is a well-known limitation of the Fieller method. However, this circumstance occurs rarely. One instance is when $n_{111} = n_{211} = 0$. When $\Delta > 0$, an approximate $100(1-\alpha)\%$ Fieller CI for r is

$$\frac{\hat{p}_1\hat{p}_2 \pm \sqrt{\Delta}}{|\hat{p}_2^2 - Z_{\alpha/2}^2\hat{\sigma}_2^2|}.$$

Table 4.4. Hildesheim Example Data

			Fallible Procedure	
Group	Data	Infallible Procedure	0	1
Case	Training	0	13	3
		1	NA	23
	Original	NA	318	375
Control	Training	0	33	11
		1	NA	32
	Original	NA	701	535

NA: Not Available

4.4 An Example

In this section we compute an MLE point estimate and three CI estimates, the nWald interval, the mWald interval, and the Fieller interval, for the risk ratio r using real data. This data, displayed in Table 4.4, was first described in Hildesheim, Mann, Brinton, Szklo, Reeves, and Rawls (1991) and was later used in Boese et al. (2006). The original study explored the relationship between exposure to herpes simplex virus (HSV) and invasive cervical cancer (ICC). For the ease of description, we define the variable HSV=1 for women exposed to HSV and HSV=0, otherwise. In addition, we define the variable ICC=1 for women in the first group, or the case group, that have ICC and ICC=2 for women in the second group, or the control group, that do not have ICC. This study included a total of 2044 women with 732 women in the case group and 1312 women in the control group. The western blot procedure was treated as a fallible detector of HSV. A sub-sample of 6% of the women were also tested using the refined western blot procedure, which is a relatively accurate procedure and, thus, was treated as infallible. We regard this sub-sample as the training data in the double-sampling scheme. Both false-positive and false-negative misclassification errors of HSV using the western blot procedure occurred in this study. However, for the sake of illustration, we consider only the false-positives and absorb the n_{i10} (false-negative) observations into the n_{i11} observations, $i = 1, 2$.

Table 4.5. nWald, mWald, and Fieller CIs for the Hildesheim et. al Data

Method	CI	Length
nWald	(1.15, 1.81)	0.66
mWald	(1.18, 1.85)	0.67
Fieller	(1.19, 1.86)	0.67

Here, $p_1 = \Pr(\text{exposed to HSV} \mid \text{has ICC})$ is the probability that a patient truly has been exposed to HSV given she has ICC (case group), and $p_2 = \Pr(\text{exposed to HSV} \mid \text{does not have ICC})$ is the probability that a patient truly has been exposed to HSV, given she does not have ICC (control group). Recall that $r = p_1/p_2$.

The MLE for r is $\hat{r} = 1.48$, and we give the nWald, mWald, and Fieller approximate 90% CIs and their corresponding lengths in Table 4.5. For this particular example, all three interval estimation methods produce similar CIs. Because the lower limit of the CIs for all three intervals exceeds one, we conclude that a higher proportion of women in the case group have been exposed to HSV than in the control group. Thus, a relationship between exposure to HSV and having ICC could exist.

4.5 Simulations

In this section we describe and present the results of two Monte Carlo simulation studies to assess and compare the coverage probabilities and average lengths of our proposed CIs under various parameter and sample-size scenarios. In particular we considered two-sided approximate 90% CIs. Although equal sample sizes from each group are not required for the interval estimators, for the sake of simplifying the conduct of the simulation studies and presentation of simulation results, we used the total sample sizes $N_1 = N_2 = N$, training data sample sizes $n_1 = n_2 = n$, and false-positive rates $\phi_1 = \phi_2 = \phi$.

We first studied the performance of our three proposed CI methods by varying total sample sizes. In these simulations, we chose the following parameter and

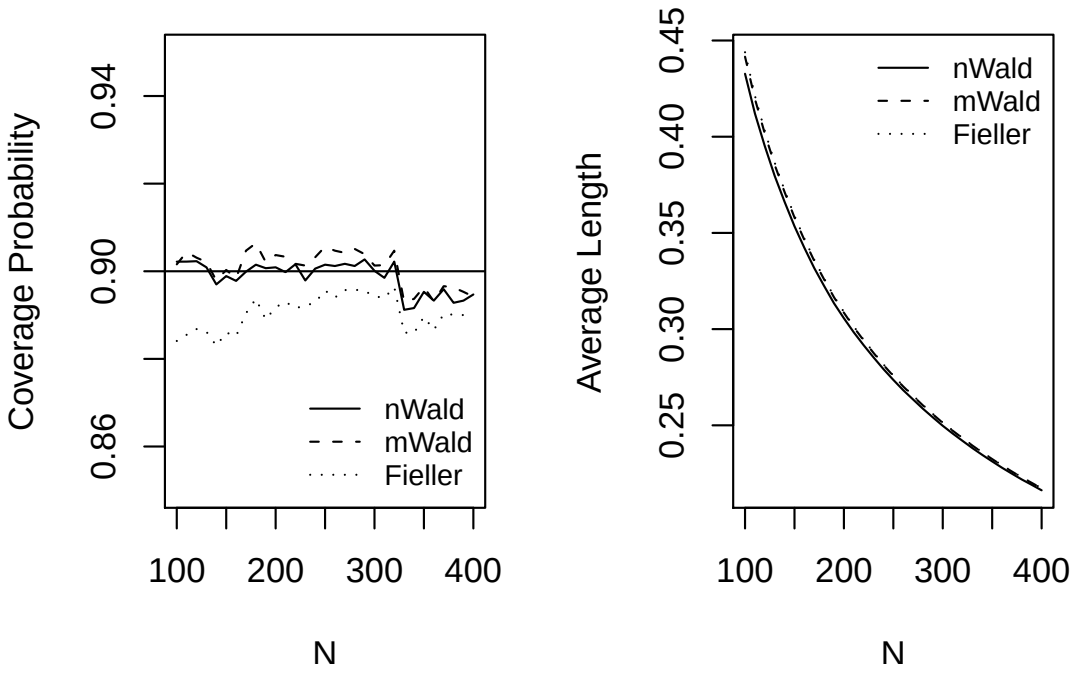

Figure 4.1: Coverage probabilities and average lengths versus total sample size N where $(p_1, p_2) = (.4, .6)$. The false-positive rate is $\phi = .1$ and $s = n/N = 0.2$.

sample-size configurations:

(1) False-positive rate: $\phi = .1$,

(2) Ratio of the training-sample size versus the total sample size: $s = n/N = 0.2$,

(3) Total sample size N: from 100 to 400 with increments of 10,

(4) Proportion parameters of interest (p_1, p_2): (.4, .6), (.1, .2), corresponding to risk ratios of 2/3 and 1/2, respectively.

For each simulation parameter and sample-size configuration for (p_1, p_2), ϕ, n/N, and N, we generated $K = 10,000$ data sets. To generate a data set, for $i = 1, 2$, we sampled $(n_{i00}, n_{i01}, n_{i11})$ using (4.3) and (X_i, Y_i) using (4.4). Then, we obtained the complete data $\boldsymbol{d}$ by using (4.6). After a data set was generated, we used the three competing CI methods developed in Section 4.3 to compute CIs for r. Once the K

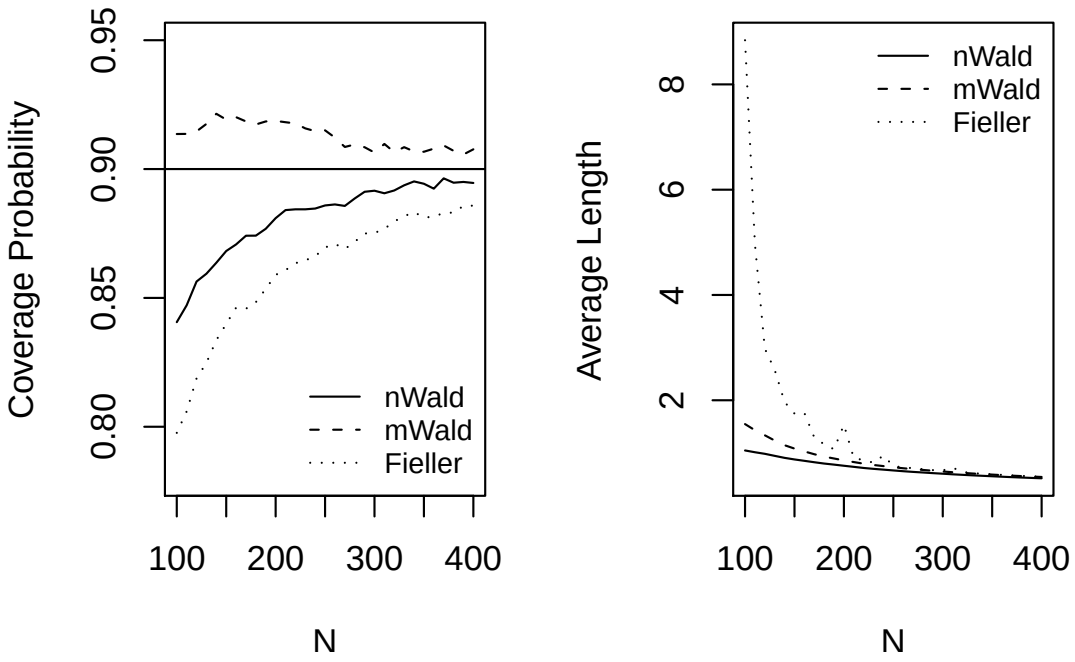

Figure 4.2: Coverage probabilities and average lengths versus total sample size N where $(p_1, p_2) = (.1, .2)$. The false-positive rate is $\phi = .1$ and $s = n/N = 0.2$.

CIs were obtained for each type of CI, we calculated the coverage probability(CP) and average length (AL) of the K CIs. Finally, we plotted the CP and AL versus N for each CI method.

Figures 4.1 and 4.2 display plots of CPs and ALs of the three CI methods versus N for $(p_1, p_2) = (.4, .6)$ and $(p_1, p_2) = (.1, .2)$, respectively. We remark that when $(p_1, p_2) = (.4, .6)$, binomial distributions can be well approximated by normal distributions and, therefore, we expected the proposed CI methods to perform well. In fact, Figure 4.1 demonstrates that both the nWald and mWald CIs had similar, close-to-nominal CPs even for small samples with $N = 100$. The Fieller CIs had reasonable CPs when sample sizes were small ($N < 200$) and close-to-nominal CPs when sample sizes were larger ($N \geq 200$). The ALs were similar for all three CIs with the nWald CIs being the narrowest and the Fieller CIs being the widest. On the other hand, when $(p_1, p_2) = (.1, .2)$, binomial distributions were skewed and, therefore, not very well behaved. Therefore, we did not expect the proposed CIs to

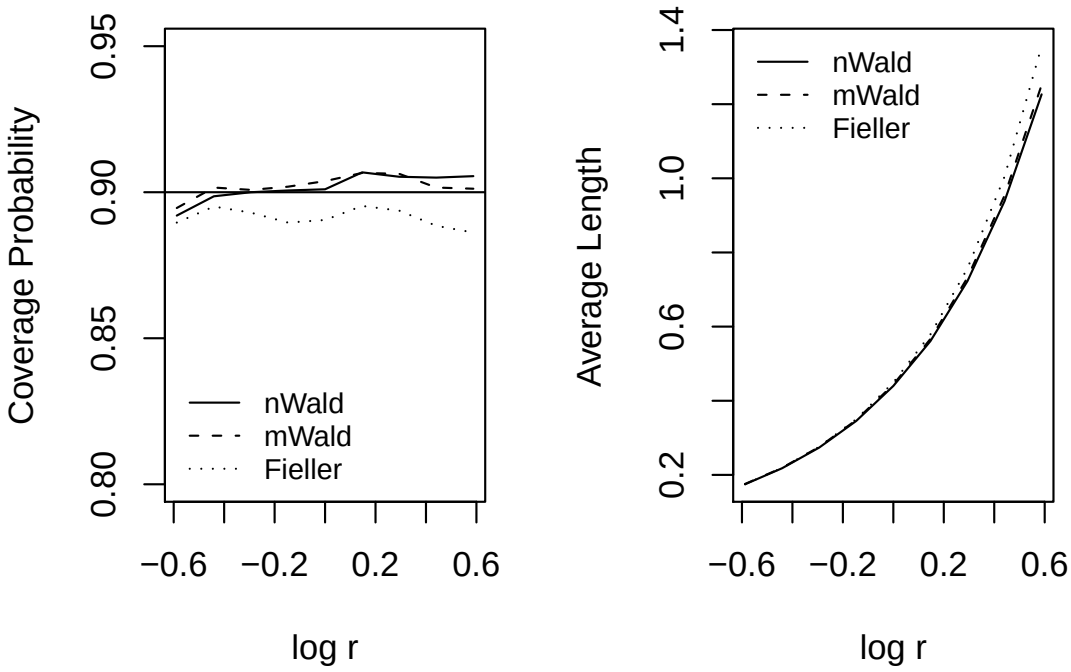

Figure 4.3: Coverage probabilities and average lengths versus the log risk ratios r, where $p_1 = .5$. The false-positive rate is $\phi = .1$, total sample size $N = 200$, and $s = n/N = 0.2$.

perform well for small sample sizes ($N < 200$). In fact, Figure 4.2 shows that both the nWald and Fieller CIs had poor coverage where $N < 200$. The coverage was close to nominal when sample sizes were larger ($N > 300$). Remarkably, the mWald CIs had good coverage for the sample sizes considered here. When comparing ALs, we were not surprised that nWald CIs were narrower on average than mWald CIs because naive Wald intervals tended to be consistently too narrow. However, the Fieller CIs were the widest, especially when the sample sizes were small. Thus, even though the Fieller CIs were wide, they often did not cover the true risk ratios.

We then studied the performance of the nWald, mWald, and Fieller CIs by varying the risk ratio r. Here, we chose the following parameter configurations:

(1) False-positive rate: $\phi = .1$,

(2) Ratio of the training-sample size versus the total sample size: $s = n/N = 0.2$, and

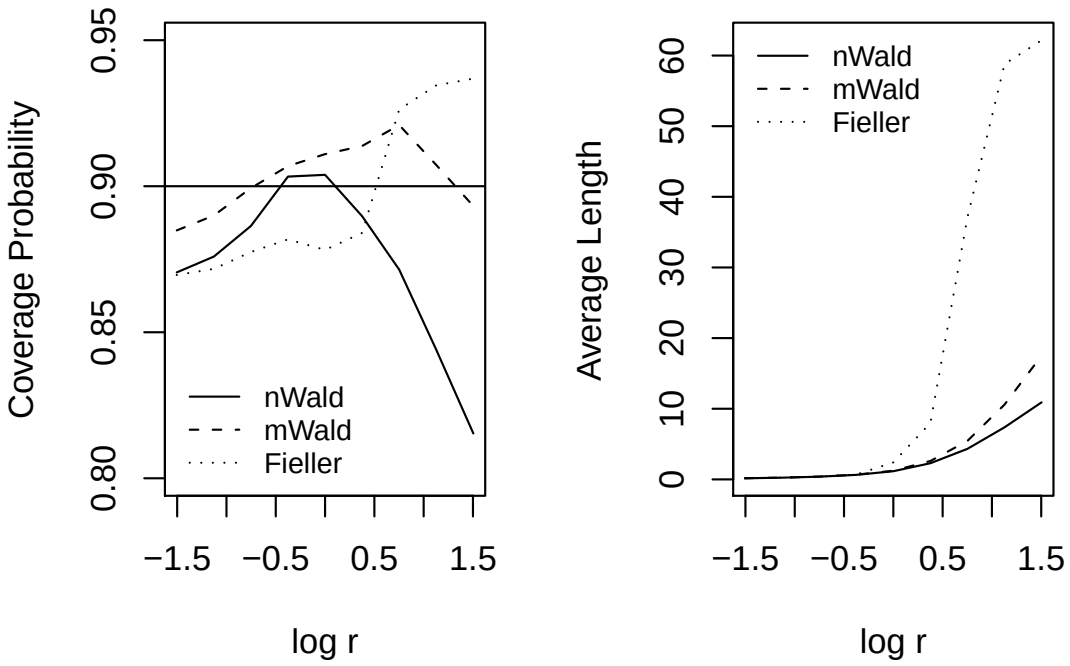

Figure 4.4: Coverage probabilities and average lengths versus the log risk ratios r, where $p_1 = .2$. The false-positive rate is $\phi = .1$, total sample size $N = 200$, and $s = n/N = 0.2$.

(3) Total sample size: $N = 200$.

We considered two simulation configurations for p_1 and p_2 with fixed values of $p_1 = .5$ and $p_1 = .2$ for the first and second configurations, respectively. For each configuration, we chose 9 values of p_2, $\{p_{2,1}, \ldots, p_{2,9}\}$, in an increasing order, such that $\log(r_1)$ and $\log(r_9)$ were symmetric about 0, and $\{\log(r_1), \ldots, \log(r_9)\}$ were equally spaced, where $r_t = p_1/p_{2,t}, t = 1, \ldots, 9$. We let $p_{2,9} = .9$ for both configurations. Using the assumption that $\log(r_1)$ and $\log(r_9)$ are symmetric about 0, we obtained $p_{2,1} \approx 0.278$ and $p_{2,1} \approx 0.044$ for the two configurations, respectively. Note that in this way, we ensured that the values of the parameters $\{p_{2,1}, \ldots, p_{2,9}\}$ were between 0 and 1. For each configuration, we then determined $p_{2,2}, \ldots, p_{2,8}$ such that $\{\log(r_1), \ldots, \log(r_9)\}$ were equally spaced.

For each simulation scenario of known (p_1, p_2), ϕ, n/N, and N, we generated $K = 10,000$ data sets. The generation of one data set was described previously

in this section. Once the K CIs for each interval were obtained, we calculated the coverage probabilities (CPs) and average lengths (ALs). Finally, we plotted the CPs and ALs versus $\log r$ for each type of CI.

Figures 4.3 and 4.4 depict plots of the CPs and ALs of all CI methods versus $\log r$ for both configurations of p_1 and p_2, respectively. Figure 4.3 shows that both the nWald and the mWald CIs had close-to-nominal coverage for the range of $\log r$ studied here. The Fieller CI also had close-to-nominal coverage for the range of $\log r$ studied here, although the coverage was consistently slightly below the nominal level. Figure 4.3 also displays that the Fieller CI was slightly wider than the other two CIs. Figure 4.4 shows that the mWald CI had close-to-nominal coverage for the range of $\log r$ studied here. The nWald CI had close-to-nominal coverage when $\log r \in (-\log .5, \log .5)$ and below-nominal coverage otherwise. The Fieller CI had below-nominal coverage when $\log r < .5$ and above-nominal coverage when $\log r > .5$. Figure 4.4 also displays that the mWald CI was slightly wider than the nWald CI. The Fieller CI was the widest and was much wider than the other two CIs when $\log r > .5$.

4.6 Discussion

In this paper we have derived three CIs: the nWald, mWald, and Fieller intervals for risk ratios using two-sample binary data subject to false-positive misclassification. The nWald CI is obtained using a naive application of the Wald interval estimation method, the mWald CI is based on a modified Wald method that guarantees nonnegative CI limits, and the Fieller CI is constructed using a pivotal quantity. All three interval estimators are easy to compute.

The three proposed CIs were applied to cervical cancer data and produced similar CIs for the risk ratio r. We expected this outcome because all three CIs perform well when sample sizes are large, as in the cervical cancer data.

We have conducted Monte Carlo simulation studies to examine the CPs and ALs of the three proposed CIs for the risk ratio r under various simulation scenarios. In general, all three interval estimators performed well for large samples. That is, the CPs were close to the nominal level. Also, the ALs decreased as sample sizes increased. The mWald interval produced CIs that were close to the nominal level under all simulation scenarios considered here. Compared with the mWald CI, the nWald CI was generally narrower but had CPs considerably less than the nominal level, especially when p_1 and p_2 were close to zero or one and the sample sizes were small ($N \approx 100$). The Fieller CI was sometimes much wider than the other two intervals and had CPs larger or smaller than the nominal levels, especially when p_1 and p_2 were close to zero or one and the sample sizes were small ($N \approx 100$). In summary, the mWald CI performed well for the parameter configurations and sample sizes considered here and, hence, is preferred to the nWald and Fieller intervals.

CHAPTER FIVE

Confidence Intervals for the Risk Ratio Using Double Sampling with Misclassified Binomial Data

5.1 Introduction

Several researchers, including Bross (1954), have studied the effect of misclassification on the classical proportion estimators. In general, two types of misclassification for binary misclassified observations exist: false-positive and false-negative binary observations. For example, visual inspection by a midwife or obstetrician may erroneously classify a normal child as having Down's syndrome (false-positive), or one may classify a child with Down's syndrome as being healthy (false-negative). In many applications with misclassified binary data, both misclassification types are present.

Because classical estimators that ignore misclassification are biased, one needs additional data to correct the bias and achieve model identifiability. Various methods in the statistical literature have been proposed for this purpose. For the Bayesian paradigm, when an infallible classifier is unavailable or prohibitively expensive, one can use sufficiently informative priors to obtain model identifiability. Another information-producing method is to use multiple fallible classifiers. This article focuses on an information-addition method first proposed by Tenenbein (1970) that includes training data obtained by double sampling.

One can apply Tenenbein's double sampling scheme when both fallible and infallible measuring devices or classifiers are available. Usually, the fallible classifier is relatively inexpensive but may misclassify units, while the infallible classifier is generally much more expensive but is infallible. Tenenbein's approach was to compromise between the two extremes by using the infallible classifier on only a small

portion of the data and using the fallible classifier on all of the data. This approach, called double sampling, not only enables model identifiability but is also economical.

A number of researchers have used misclassified binary data to provide point and interval estimation methods for various functions of the proportion parameters of interest. For one-sample binomial problems where only one type of error or misclassification is present, Lie, Heuch, and Irgens (1994) have used a maximum likelihood approach, where false-negative errors are corrected with multiple fallible classifiers, whereas York, Madigan, Heuch, and Lie (1995) have considered the same problem from a Bayesian approach. Using data obtained by double sampling, Moors, van der Genugten, and Strijbosch (2000) have discussed method of moments and maximum likelihood estimation, in addition to one-sided interval estimation. Also, Boese, Young, and Stamey (2006) have derived several likelihood-based confidence intervals (CIs) for a single proportion parameter, while Lee and Byun (2008) have provided Bayesian credible intervals using noninformative priors for the same problem.

Additionally, several researchers have studied one-sample problems with both types of binary misclassification errors. In conjunction with double sampling, Tenenbein (1970) has proposed a maximum likelihood estimator for a single proportion parameter and has derived an expression for the estimator's asymptotic variance. For the case when training data are unavailable in the one-sample problem, Gaba and Winkler (1992) and Viana, Ramakrishnan, and Levy (1993) have developed Bayesian approaches using sufficiently informative priors.

For the two-sample problem with both types of binary misclassification errors, Bayesian inference methods using sufficiently informative priors have also been developed when training data are unavailable. For example, see Evans, Guttman, Haitovsky, and Swartz (1996) for risk-difference estimation, that is, the difference of two proportion parameters, and Gustafson, Le, and Saskin (2001) for estimation of

Table 5.1. Data from the Fallible Classifier for Sample i, $i = 1, 2$

Classification	0	1	Total
Count	Y_i	X_i	M_i

odds ratios. When training data is obtained through double sampling, Boese (2003) has derived several likelihood-based CIs for the risk difference.

So far, no inference methods for the risk ratio of two proportion parameters have been published for two-sample misclassified binary data. In this article, we develop point and interval estimators for this problem. The remainder of this paper is organized as follows. In Section 5.2 we describe the data, and in Section 5.3 we derive three likelihood-based interval estimators of a risk ratio using double sampling with misclassified data containing both false-negative and false-positive observations. In Section 5.4 we illustrate the newly derived interval estimators using real cervical cancer data. We examine and compare the performance of three interval estimators in Section 5.5, and we give a brief discussion in Section 5.6.

5.2 The Data

In this section we introduce notation and rigorously describe two-sample misclassified binomial data. The original data are obtained with a fallible classifier that produces both false-positive and false-negative observations.

We first introduce notation necessary for describing the data. Let F_{ij} be the observed classification by the fallible classifier for the jth individual in the ith sample, where $i = 1, 2$, $j = 1, \ldots, M_i$, and $F_{ij} = 1$ if the result by the fallible classifier is positive, and $F_{ij} = 0$ otherwise. Let $X_i = \sum_j F_{ij}$ and $Y_i = M_i - X_i$ be the observed number of positive and negative observations, respectively. The data obtained by the fallible classifier for sample i, $i = 1, 2$, are displayed in Table 5.1. Similarly, we define the unobserved true classification of the jth individual in the ith sample

as $T_{ij} = 1$ if the classifier result is truly positive, and $T_{ij} = 0$ otherwise. Clearly, misclassification occurs when $T_{ij} \neq F_{ij}$.

Also, we let

$$\begin{aligned} p_i &\equiv \Pr(T_{ij} = 1), \\ \pi_i &\equiv \Pr(F_{ij} = 1), \\ \phi_i &\equiv \Pr(F_{ij} = 1 | T_{ij} = 0), \end{aligned}$$

and

$$\theta_i \equiv \Pr(F_{ij} = 0 | T_{ij} = 1).$$

Here, p_i is the actual proportion parameter of interest, π_i is the proportion parameter of the fallible classifier, ϕ_i and θ_i are the false-positive and the false-negative rates, respectively, for the fallible classifier. Note that we allow the false-positive rates and false-negative rates to be different between the two samples, i.e., we allow $\phi_1 \neq \phi_2$ and $\theta_1 \neq \theta_2$. Also we remark that π_1 and π_2 are not additional unique parameters because

$$\begin{aligned} \pi_i &= \Pr(T_{ij} = 1) \Pr(F_{ij} = 1 | T_{ij} = 1) + \Pr(T_{ij} = 0) \Pr(F_{ij} = 1 | T_{ij} = 0) \\ &= p_i(1 - \theta_i) + q_i \phi_i, \end{aligned} \tag{5.1}$$

where $q_i = 1 - p_i$. As noted in Section 5.1, we wish to develop point and interval estimators of the risk ratio

$$r = p_1 / p_2. \tag{5.2}$$

Because π_i is determined through p_i, ϕ_i, and θ_i, $i = 1, 2$, effectively six parameters result in the model: p_1, ϕ_1, θ_1, p_2, ϕ_2, θ_2. However, the sufficient statistics dimension is only two because X_1 and X_2 are the minimal sufficient statistics for this model. Therefore, six parameters in model (5.1) are unidentifiable because the dimension of the sufficient statistics is less than the number of parameters and,

Table 5.2. Data for Sample i

		Fallible Classifier		
Data	Infallible Classifier	0	1	Total
Training	0	n_{i00}	n_{i01}	$n_{i0\cdot}$
	1	n_{i10}	n_{i11}	$n_{i1\cdot}$
	Total	$n_{i\cdot 0}$	$n_{i\cdot 1}$	n_i
Original	NA	Y_i	X_i	M_i

NA: Not Available

therefore, additional data are needed for model identifiability. In this paper we use double sampling to provide additional information. Specifically, in addition to the original fallible data classified only by the fallible classifier, new but smaller training data are obtained when we classify each individual in this training data by both the fallible and the infallible classifiers. The double sampling paradigm has attracted researchers' interests due to its practicality. Other applications of double-sampling schemes have been addressed in Tenenbein (1972), Hochberg (1977), and Boese et al. (2006).

In this paper we assume that for the ith sample, training data of size n_i are obtained using double sampling in addition to the original fallible data of size M_i, $i = 1, 2$. Hence, the size of the combined data is $N_i = M_i + n_i$ for sample i. Table 5.2 presents the combined data by concatenating the original and training data. In Table 5.2 we use n_{ijk} to denote the number of individuals classified as j and k by the infallible and fallible classifiers, respectively. For example, n_{i01} is the number of individuals in the ith sample classified as negative by the infallible classifier but classified as positive by the fallible classifier. With the additional training data, the dimension of the sufficient statistic for the combined data is sufficient for estimating all parameters and, therefore, the full model is identifiable. For future estimation methodology development, we present the cell probabilities corresponding to Table 5.2 in Table 5.3.

Table 5.3. Cell Probabilities for Sample i

		Fallible Classifier		
Data	Infallible Classifier	0	1	Total
Training	0	$q_i(1-\phi_i)$	$q_i\phi_i$	q_i
	1	$p_i\theta_i$	$p_i(1-\theta_i)$	p_i
Original	NA	$1-\pi_i$	π_i	1

NA: Not Available

5.3 The Model

For data described in the previous section, we derive point and interval estimators for the risk ratio (5.2) of two proportion parameters using double sampling on possibly misclassified data. In particular, we derive closed-form maximum likelihood estimators (MLEs). In addition, we obtain an asymptotic covariance matrix of the vector of MLEs by computing the inverse of the Fisher information matrix. Finally, we develop two closed-form Wald-based CIs and a Fieller-type CI for the risk ratio r based on the full likelihood.

5.3.1 The Full Likelihood Function

Table 5.2 presents the data for the inference problem under consideration. For sample i, the observed counts $(n_{i00}, n_{i01}, n_{i10}, n_{i11})'$ of the training data have a quadrinomial distribution with total size n_i and probabilities displayed in an upper right 2×2 submatrix in Table 5.3, i.e.,

$$(n_{i00}, n_{i01}, n_{i10}, n_{i11})|p_i, \phi_i, \theta_i \sim \text{Quad}[n_i, (q_i(1-\phi_i), q_i\phi_i, p_i\theta_i, p_i(1-\theta_i))]. \quad (5.3)$$

In addition, the observed counts (X_i, Y_i) have the binomial distribution

$$(X_i, Y_i)|p_i, \phi_i, \theta_i \sim \text{Bin}[M_i, (\pi_i, 1-\pi_i)]. \quad (5.4)$$

Because $(n_{i00}, n_{i01}, n_{i10}, n_{i11})'$ and $(X_i, Y_i)'$ are independent for sample i and because sample 1 is independent of sample 2, the probability density function of the data

vector given the parameter vector is

$$f(\boldsymbol{d}|\boldsymbol{\eta}) \propto \prod_{i=1}^{2}\{[q_i(1-\phi_i)]^{n_{i00}}(q_i\phi_i)^{n_{i01}}(p_i\theta_i)^{n_{i10}}[p_i(1-\theta_i)]^{n_{i11}}\pi_i^{X_i}(1-\pi_i)^{Y_i}\}, \quad (5.5)$$

where

$$\boldsymbol{d} = (n_{100}, n_{101}, n_{110}, n_{111}, X_1, Y_1, n_{200}, n_{201}, n_{210}, n_{211}, X_2, Y_2)' \quad (5.6)$$

and

$$\boldsymbol{\eta} = (p_1, \phi_1, \theta_1, p_2, \phi_2, \theta_2)'.$$

Finally, we can express the full likelihood function as

$$L_f(\boldsymbol{\eta}) \propto \prod_{i=1}^{2}\{[q_i(1-\phi_i)]^{n_{i00}}(q_i\phi_i)^{n_{i01}}(p_i\theta_i)^{n_{i10}}[p_i(1-\theta_i)]^{n_{i11}}\pi_i^{X_i}(1-\pi_i)^{Y_i}\}. \quad (5.7)$$

5.3.2 MLEs Based on the Full Likelihood Function

We now derive the maximum likelihood estimators (MLEs) of all parameters of interest. Generally, directly maximizing (5.7) with respect to $\boldsymbol{\eta}$ requires such numerical methods as the Newton-Raphson algorithm. These numerical methods are computationally expensive and may have convergence issues. Instead of using these numerical methods, we first perform a reparameterization of parameters $\boldsymbol{\eta}$ and then derive closed-form solutions. Let

$$\lambda_{i1} \equiv p_i(1-\theta_i)/\pi_i, \quad (5.8)$$

$$\lambda_{i2} \equiv p_i\theta_i/(1-\pi_i), \quad (5.9)$$

and $\boldsymbol{\gamma} \equiv (\lambda_{11}, \lambda_{12}, \pi_1, \lambda_{21}, \lambda_{22}, \pi_2)'$, $i = 1, 2$. Using (5.1), (5.8), and (5.9), we see that (5.7) can be reexpressed as

$$L_f(\boldsymbol{\gamma}) \propto \prod_{i=1}^{2}\left[\lambda_{i1}^{n_{i11}}(1-\lambda_{i1})^{n_{i01}}\lambda_{i2}^{n_{i10}}(1-\lambda_{i2})^{n_{i00}}\pi_i^{X_i+n_{i\cdot1}}(1-\pi_i)^{Y_i+n_{i\cdot0}}\right]. \quad (5.10)$$

Therefore, the full log likelihood is

$$\begin{aligned} l_f(\boldsymbol{\gamma}) \propto \sum_{i=1}^{2}[&n_{i11}\log\lambda_{i1} + n_{i01}\log(1-\lambda_{i1}) + n_{i10}\log\lambda_{i2} + n_{i00}\log(1-\lambda_{i2}) + \\ &(X_i + n_{i\cdot1})\log\pi_i + (Y_i + n_{i\cdot0})\log(1-\pi_i)], \end{aligned} \quad (5.11)$$

and the corresponding score vector is

$$\begin{aligned}
\boldsymbol{s}_f(\boldsymbol{\gamma}) &\equiv \frac{\partial l_f(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \\
&= \left[\frac{n_{111}}{\lambda_{11}} - \frac{n_{101}}{1-\lambda_{11}}, \frac{n_{110}}{\lambda_{12}} - \frac{n_{100}}{1-\lambda_{12}}, \frac{X_1 + n_{1.1}}{\pi_1} - \frac{Y_1 + n_{1\cdot 0}}{1-\pi_1}, \right. \\
&\quad \left. \frac{n_{211}}{\lambda_{21}} - \frac{n_{201}}{1-\lambda_{21}}, \frac{n_{210}}{\lambda_{22}} - \frac{n_{200}}{1-\lambda_{22}}, \frac{X_2 + n_{2.1}}{\pi_2} - \frac{Y_2 + n_{2\cdot 0}}{1-\pi_2}\right]'. \quad (5.12)
\end{aligned}$$

We obtain the MLE for $\boldsymbol{\gamma}$ by setting $\boldsymbol{s}_f(\boldsymbol{\gamma}) = \mathbf{0}$ and solving for λ_{i1}, λ_{i2}, and π_i, so that

$$\hat{\lambda}_{i1} = \frac{n_{i11}}{n_{i\cdot 1}},$$

$$\hat{\lambda}_{i2} = \frac{n_{i10}}{n_{i\cdot 0}},$$

and

$$\hat{\pi}_i = \frac{X_i + n_{i\cdot 1}}{N_i},$$

$i = 1, 2$. By solving (5.1), (5.8), and (5.9) and applying the invariance property of MLEs, we find the MLEs for $\boldsymbol{\eta}$ are

$$\hat{p}_i = \hat{\pi}_i \hat{\lambda}_{i1} + (1 - \hat{\pi}_i)\hat{\lambda}_{i2},$$

$$\hat{\phi}_i = (1 - \hat{\lambda}_{i1})\hat{\pi}_i / \hat{q}_i,$$

$$\hat{\theta}_i = \hat{\lambda}_{i2}(1 - \hat{\pi}_i)/\hat{p}_i,$$

$i = 1, 2$, and

$$\hat{r} = \hat{p}_1 / \hat{p}_2. \quad (5.13)$$

5.3.3 The Full Likelihood Information Matrix

From (5.12), the Hessian matrix is

$$\begin{aligned}\boldsymbol{H}_f(\boldsymbol{\gamma}) = \text{Diag}\Bigg[&-\frac{n_{111}}{\lambda_{11}^2} - \frac{n_{101}}{(1-\lambda_{11})^2}, -\frac{n_{110}}{\lambda_{12}^2} - \frac{n_{100}}{(1-\lambda_{12})^2},\\ &-\frac{X_1 + n_{1.1}}{\pi_1^2} - \frac{Y_1 + n_{1\cdot 0}}{(1-\pi_1)^2}, -\frac{n_{211}}{\lambda_{21}^2} - \frac{n_{201}}{(1-\lambda_{21})^2},\\ &-\frac{n_{210}}{\lambda_{22}^2} - \frac{n_{200}}{(1-\lambda_{22})^2}, -\frac{X_2 + n_{2.1}}{\pi_2^2} - \frac{Y_2 + n_{2\cdot 0}}{(1-\pi_2)^2}\Bigg].\end{aligned} \tag{5.14}$$

Thus, the expected Fisher information matrix is

$$\begin{aligned}\boldsymbol{I}_f(\boldsymbol{\gamma}) = \text{Diag}\Bigg[&\frac{n_1\pi_1}{\lambda_{11}(1-\lambda_{11})}, \frac{n_1(1-\pi_1)}{\lambda_{12}(1-\lambda_{12})}, \frac{N_1}{\pi_1(1-\pi_1)},\\ &\frac{n_2\pi_2}{\lambda_{21}(1-\lambda_{21})}, \frac{n_2(1-\pi_2)}{\lambda_{22}(1-\lambda_{22})}, \frac{N_2}{\pi_2(1-\pi_2)}\Bigg].\end{aligned}$$

Because the necessary regularity conditions are satisfied for this model, the MLE vector $\hat{\boldsymbol{\gamma}} = (\hat{\lambda}_{11}, \hat{\lambda}_{12}, \hat{\pi}_1, \hat{\lambda}_{21}, \hat{\lambda}_{22}, \hat{\pi}_2)'$ has an asymptotic multivariate normal distribution with asymptotic mean $\boldsymbol{\gamma}$ and asymptotic covariance matrix

$$\begin{aligned}\boldsymbol{I}_f^{-1}(\boldsymbol{\gamma}) = \text{Diag}\Bigg[&\frac{\lambda_{11}(1-\lambda_{11})}{n_1\pi_1}, \frac{\lambda_{12}(1-\lambda_{12})}{n_1(1-\pi_1)}, \frac{\pi_1(1-\pi_1)}{N_1},\\ &\frac{\lambda_{21}(1-\lambda_{21})}{n_2\pi_2}, \frac{\lambda_{22}(1-\lambda_{22})}{n_2(1-\pi_2)}, \frac{\pi_2(1-\pi_2)}{N_2}\Bigg].\end{aligned}$$

Thus, for $i = 1, 2$, asymptotically we have

$$\begin{aligned} V(\hat{\lambda}_{i1}) &= \frac{\lambda_{i1}(1-\lambda_{i1})}{n_i\pi_i},\\ V(\hat{\lambda}_{i2}) &= \frac{\lambda_{i2}(1-\lambda_{i2})}{n_i(1-\pi_i)},\\ V(\hat{\pi}_i) &= \frac{\pi_i(1-\pi_i)}{N_i},\end{aligned}$$

and that $\hat{\lambda}_{11}, \hat{\lambda}_{12}, \hat{\pi}_1, \hat{\lambda}_{21}, \hat{\lambda}_{22}, \hat{\pi}_2$ are asymptotically mutually independent.

5.3.4 A Full Likelihood Naive Wald CI

We begin with constructing a naive Wald-type confidence interval for the risk ratio r. Note that $\hat{p}_i = \hat{\pi}_i\hat{\lambda}_{i1} + (1-\hat{\pi}_i)\hat{\lambda}_{i2}$ and that $\hat{\lambda}_{i1}$, $\hat{\lambda}_{i2}$, and $\hat{\pi}_i$ are independent,

$i = 1, 2$. Thus, using the delta method, we have

$$\begin{aligned}
\sigma_i^2 &\equiv V(\hat{p}_i) \\
&\approx \left(\frac{\partial p_i}{\partial \lambda_{i1}}\right)^2 V(\hat{\lambda}_{i1}) + \left(\frac{\partial p_i}{\partial \lambda_{i2}}\right)^2 V(\hat{\lambda}_{i2}) + \left(\frac{\partial p_i}{\partial \pi_i}\right)^2 V(\hat{\pi}_i) \\
&= \frac{\pi_i \lambda_{i1}(1-\lambda_{i1})}{n_i} + \frac{(1-\pi_i)\lambda_{i2}(1-\lambda_{i2})}{n_i} + \frac{(\lambda_{i1}-\lambda_{i2})^2 \pi_i (1-\pi_i)}{N_i}. \quad (5.15)
\end{aligned}$$

The MLEs $\hat{\lambda}_{i1}$, $\hat{\lambda}_{i2}$, and $\hat{\pi}_i$ are consistent estimators of λ_{i1}, λ_{i2}, and π_i, respectively. Because a continuous function of consistent estimators is consistent, we have that a consistent estimator of (5.15) is

$$\hat{\sigma}_i^2 = \frac{\hat{\pi}_i \hat{\lambda}_{i1}(1-\hat{\lambda}_{i1})}{n_i} + \frac{(1-\hat{\pi}_i)\hat{\lambda}_{i2}(1-\hat{\lambda}_{i2})}{n_i} + \frac{(\hat{\lambda}_{i1}-\hat{\lambda}_{i2})^2 \hat{\pi}_i (1-\hat{\pi}_i)}{N_i}. \quad (5.16)$$

Recall that the MLE of r is $\hat{r} = \hat{p}_1/\hat{p}_2$. Again using the delta method, we have

$$\begin{aligned}
\sigma_r^2 \equiv V(\hat{r}) &\approx \left(\frac{\partial r}{\partial p_1}\right)^2 V(\hat{p}_1) + \left(\frac{\partial r}{\partial p_2}\right)^2 V(\hat{p}_2) \\
&= \frac{\sigma_1^2}{p_2^2} + \frac{p_1^2 \sigma_2^2}{p_2^4}, \quad (5.17)
\end{aligned}$$

and a consistent estimator of (5.17) is

$$\hat{\sigma}_r^2 = \frac{\hat{\sigma}_1^2}{\hat{p}_2^2} + \frac{\hat{p}_1^2 \hat{\sigma}_2^2}{\hat{p}_2^4}. \quad (5.18)$$

Therefore, an approximate $100(1-\alpha)\%$ naive Wald (nWald) CI for r is

$$\hat{r} \pm Z_{\alpha/2} \hat{\sigma}_r, \quad (5.19)$$

where $Z_{\alpha/2}$ is the upper $(\alpha/2)$th quantile of the standard normal distribution. This interval estimator is referred to as a naive Wald CI because it results from a naive application of the Wald interval estimation method. We remark that the lower limit of the CI can be negative, especially when sample sizes are small and r is close to zero. In the case where the lower limit of the CI is negative, we replace the lower limit by zero.

5.3.5 A Full Likelihood Modified Wald CI

To alleviate the problem with the nWald CI, we propose a modified Wald (mWald) CI by first constructing an approximate $100(1-\alpha)\%$ CI for $\tau = \log r$. Then, we exponentiate this CI to obtain an approximate $100(1-\alpha)\%$ CI for r. Hong, Meeker, and Escobar (2008) also suggested using transformation of parameters when constructing Wald-type CIs. Specifically, we let $\hat{\tau} = \log \hat{r}$. Then, using the delta method, we compute

$$\begin{aligned} \sigma_\tau^2 \equiv V(\hat{\tau}) &= V(\log \hat{p}_1 - \log \hat{p}_2) \\ &\approx \frac{V(\hat{p}_1)}{p_1^2} + \frac{V(\hat{p}_2)}{p_2^2} \\ &= \frac{\sigma_1^2}{p_1^2} + \frac{\sigma_2^2}{p_2^2}. \end{aligned} \tag{5.20}$$

Clearly, a consistent estimator of (5.20) is

$$\hat{\sigma}_\tau^2 = \frac{\hat{\sigma}_1^2}{\hat{p}_1^2} + \frac{\hat{\sigma}_2^2}{\hat{p}_2^2}. \tag{5.21}$$

Then, a $100(1-\alpha)\%$ CI for τ is

$$\hat{\tau} \pm Z_{\alpha/2}\hat{\sigma}_\tau. \tag{5.22}$$

Finally, an approximate $100(1-\alpha)\%$ mWald CI for r is obtained by exponentiating (5.22):

$$\left[\hat{r}/\exp(Z_{\alpha/2}\hat{\sigma}_\tau), \hat{r}\exp(Z_{\alpha/2}\hat{\sigma}_\tau)\right]. \tag{5.23}$$

Note that the mWald CI guarantees the lower limit of (5.23) is nonnegative.

5.3.6 A Full Likelihood Fieller-Type CI

We next develop a CI for r based on an interval estimation concept introduced in Fieller (1954). As noted previously, asymptotically, we have

$$\hat{p}_i \sim N(p_i, \sigma_i^2)$$

and $\hat{p}_1$ and $\hat{p}_2$ are independent. Because

$$\frac{\hat{p}_1 - r\hat{p}_2}{\sqrt{\sigma_1^2 + r^2\sigma_2^2}} \sim N(0,1)$$

is an asymptotic pivotal quantity, we can obtain an approximate $100(1-\alpha)\%$ Fieller CI by solving

$$\frac{(\hat{p}_1 - r\hat{p}_2)^2}{\hat{\sigma}_1^2 + r^2\hat{\sigma}_2^2} = Z_{\alpha/2}^2$$

for r. Let

$$\Delta \equiv \hat{p}_1^2\hat{p}_2^2 - (\hat{p}_1^2 - Z_{\alpha/2}^2\hat{\sigma}_1^2)(\hat{p}_2^2 - Z_{\alpha/2}^2\hat{\sigma}_2^2).$$

Because

$$\hat{p}_1^2 \geq \hat{p}_1^2 - Z_{\alpha/2}^2\hat{\sigma}_1^2 \text{ and } \hat{p}_2^2 \geq \hat{p}_2^2 - Z_{\alpha/2}^2\hat{\sigma}_2^2,$$

we have $\Delta \geq 0$. Moreover, $\Delta = 0$ if and only if $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = 0$. This phenomenon occurs rarely, for example, when $n_{111} = n_{211} = 0$. When $\Delta > 0$, an approximate $100(1-\alpha)\%$ Fieller CI for r is

$$\frac{\hat{p}_1\hat{p}_2 \pm \sqrt{\Delta}}{|\hat{p}_2^2 - Z_{\alpha/2}^2\hat{\sigma}_2^2|}.$$

Clearly, when $\Delta = 0$, a $100(1-\alpha)\%$ Fieller CI does not exist, which is a well-known limitation of the Fieller method.

5.4 An Example

In this section we use a real data set to compute an MLE point estimate and three CI estimates using the nWald interval, mWald interval, and Fieller interval for the risk ratio r. This dataset, displayed in Table 5.4, was first described in Hildesheim, Mann, Brinton, Szklo, Reeves, and Rawls (1991) and was later used in Boese et al. (2006). The original study explored the relationship between exposure to herpes simplex virus (HSV) and invasive cervical cancer (ICC). A total of 2044

Table 5.4. Hildesheim et. al Data

			Fallible Classifier	
Group	Data	Infallible Classifier	0	1
Case	Training	0	13	3
		1	5	18
	Original	NA	318	375
Control	Training	0	33	11
		1	16	16
	Original	NA	701	535

NA: Not Available

women participated in this study with 732 women in the case group and 1312 women in the control group. The western blot procedure was treated as a fallible detector of HSV. A sub-sample of 6% of the women were also tested using the refined western blot procedure, which is a relatively accurate procedure and, thus, was treated as infallible. We regard this sub-sample as the training data in the double sampling scheme. Both false-positive and false-negative misclassification errors of HSV using the western blot procedure occurred in this study.

In this example, we have the p_1=Pr(exposed to HSV | has ICC) is the probability that a patient truly has been exposed to HSV, given that she has ICC (case group), and p_2=Pr(exposed to HSV | does not have ICC) is the probability that a patient truly has been exposed to HSV, given that she does not have ICC (control group). Recall that $r = p_1/p_2$.

The MLE for r is $\hat{r} = 1.34$, and we give approximate 90% nWald, mWald, and Fieller CIs and their ALs in Table 5.5. For this example, all three interval estimators produced similar CIs. Because the lower limits of two CIs (mWald and Fieller) exceed one, we conclude there is a higher proportion of women exposed to HSV in the case group than in the control group. Thus, an association between exposure to HSV and having ICC could exist. However, the evidence for drawing this conclusion is relatively weak because the lower limits of the CIs are close to one.

Table 5.5. nWald, mWald, and Fieller CIs for the Hildesheim et. al Data

Method	CI	Length
nWald	(0.98, 1.71)	0.73
mWald	(1.02, 1.76)	0.74
Fieller	(1.02, 1.78)	0.76

5.5 Simulations

In this section, we describe and present the results of two Monte Carlo simulation studies to assess and compare the performance of our proposed CIs under various parameter and sample-size scenarios. The performance was evaluated in terms of CI coverage probabilities and average lengths. In particular, we considered two-sided approximate 90% CIs. Although equal sample sizes from each group were not required by these interval estimation methods, we assumed the total sample size $N_1 = N_2 = N$, training data sample size $n_1 = n_2 = n$, false-positive rate $\phi_1 = \phi_2 = \phi$, and false-negative rate $\theta_1 = \theta_2 = \theta$, to simplify the simulation studies and presentation of simulation results.

We first investigated the performance of our three proposed CI methods by varying total sample size. In this simulation, we chose the following parameter and sample-size configurations:

(1) False-positive rate: $\phi = .1$,

(2) False-negative rate: $\theta = .1$,

(3) Ratio of the training sample size versus the total sample size: $s = n/N = 0.2$,

(4) Total sample size N: from 100 to 400 with increments of 10,

(5) True proportion parameters of interest (p_1, p_2): (.4, .6) and (.1, .2), corresponding to risk ratios of 2/3 and 1/2, respectively.

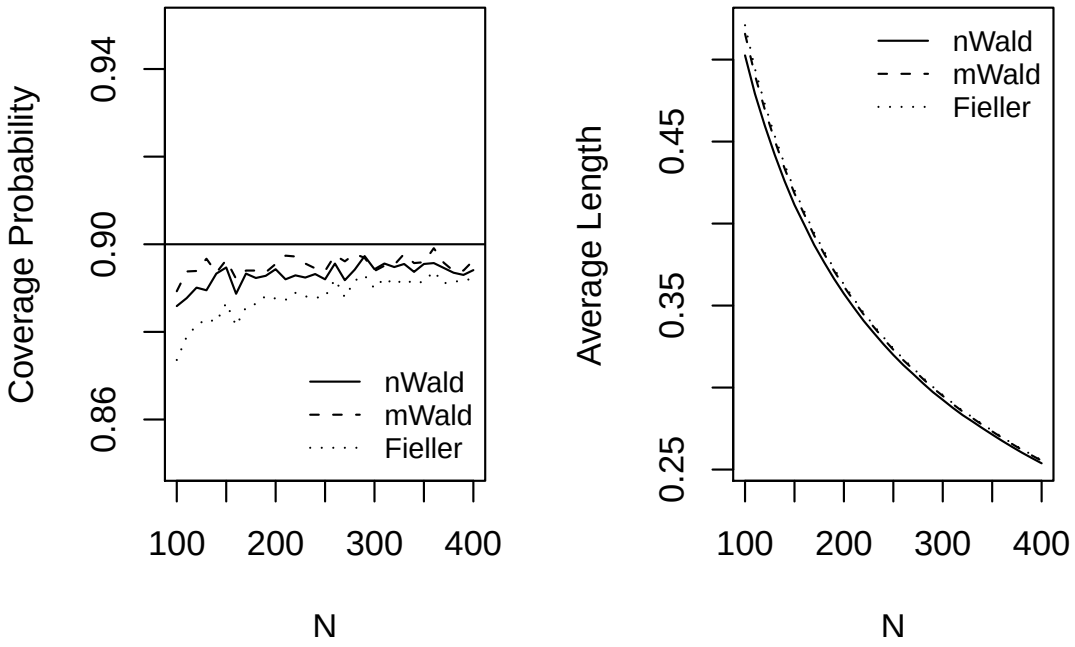

Figure 5.1: Coverage probabilities and average lengths versus total sample size N where $(p_1, p_2) = (.4, .6)$. The false-positive rate is $\phi = .1$, the false-negative rate is $\theta = .1$, and $s = n/N = 0.2$.

For each configuration of p_1, p_2, ϕ, θ, n/N, and N, we simulated $K = 10,000$ data sets. To simulate a data set, for $i = 1, 2$, we sampled $(n_{i00}, n_{i01}, n_{i10}, n_{i11})'$ using (5.3) and (X_i, Y_i) using (5.4). Then, we created the complete data $\boldsymbol{d}$ using (5.6). After a data set was created, we computed the three competing CIs for r. Once the K CIs were available for each type of CI, we computed the coverage probabilities (CPs) and the average lengths (ALs). Finally, we plotted the CPs and ALs versus sample sizes N for each type of CI.

Figures 5.1 and 5.2 display curves of CPs and ALs of the three CI estimators versus N for $(p_1, p_2) = (.4, .6)$ and $(p_1, p_2) = (.1, .2)$, respectively. When $(p_1, p_2) = (.4, .6)$, the corresponding binomial distributions are approximately symmetric about their means and, therefore, we expected the proposed CIs to perform well. Not surprisingly, Figure 5.1 demonstrates that both the nWald and mWald CIs had similar, close-to-nominal CPs, regardless of the sample sizes. The Fieller CIs had

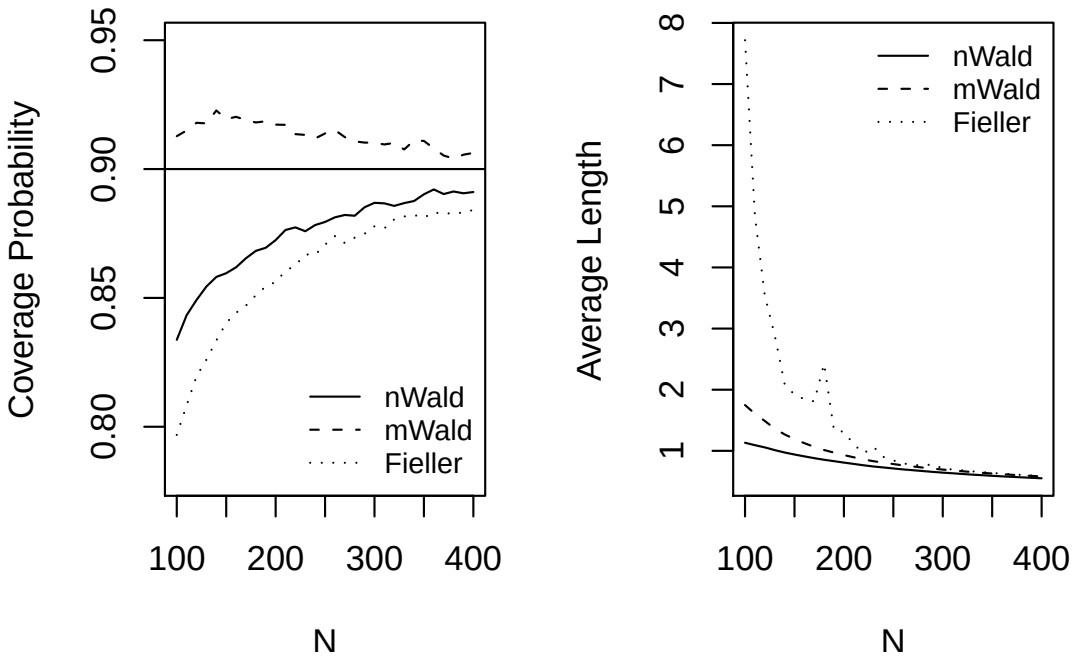

Figure 5.2: Coverage probabilities and average lengths versus total sample size N where $(p_1, p_2) = (.1, .2)$. The false-positive rate is $\phi = .1$, the false-negative rate is $\theta = .1$, and $s = n/N = 0.2$.

reasonable CPs for small samples ($N < 200$) and close-to-nominal CPs for large samples ($N \geq 200$). The ALs were similar for all three CIs with the nWald CIs being the narrowest and the Fieller CIs being the widest. On the other hand, when $(p_1, p_2) = (.1, .2)$, the corresponding binomial distributions were skewed and, therefore, not very well-behaved. Therefore, in this case, we did not expect the proposed CIs to perform as well for small sample sizes ($N < 200$). In fact, Figure 5.2 shows that both the nWald and Fieller CIs had very poor coverage for small samples ($N < 200$). However, the coverage for nWald and Fieller CIs was close to nominal when sample sizes were large ($N > 300$). Impressively, the mWald CIs had good coverage properties for all of the sample sizes considered here. For the comparison of ALs, we expected that the nWald CIs would be narrower than mWald CIs on average because naive Wald intervals commonly tend to be consistently too narrow. The Fieller CIs were generally the widest and were much wider than the

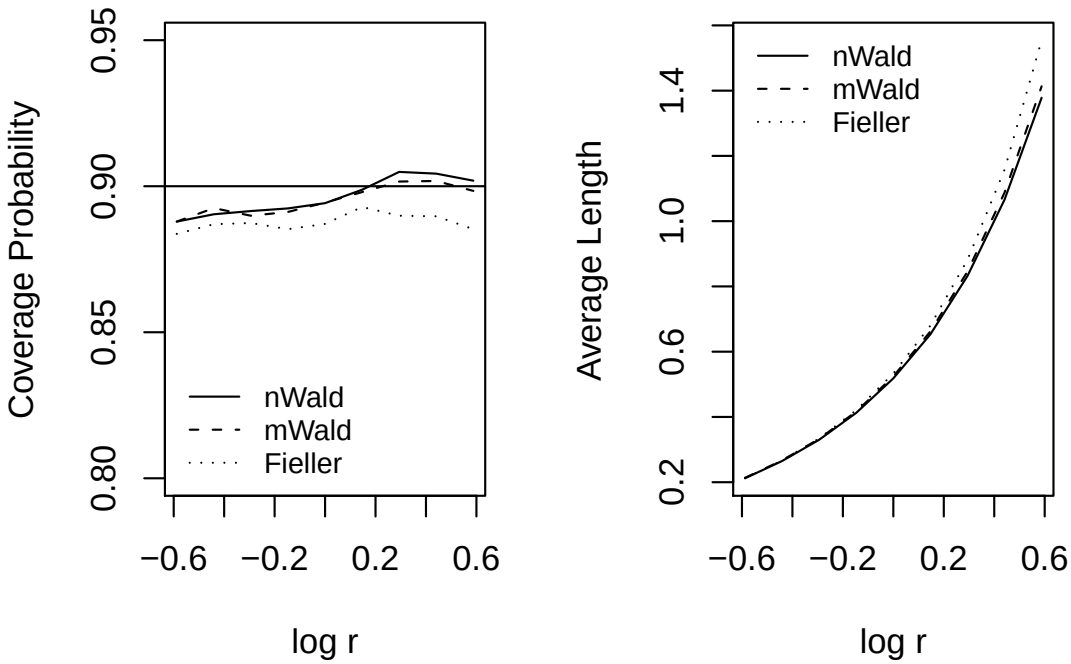

Figure 5.3: Coverage probabilities and average lengths versus the log risk ratios r where $p_1 = .5$. The false-positive rate is $\phi = .1$, the false-negative rate is $\theta = .1$, the total sample size $N = 200$, and $s = n/N = 0.2$.

other two interval estimators when the sample sizes were small. This property is very undesirable, especially with the fact that the Fieller CIs had low coverage probabilities.

Secondly, we studied the performance of the nWald, mWald, and Fieller CIs by varying the risk ratio r. In these simulations, we chose the following parameter configurations:

(1) False-positive rate: $\phi = .1$,

(2) False-negative rate: $\theta = .1$,

(3) Ratio of the training sample size versus the total sample size: $s = n/N = 0.2$,

(4) Total sample size: $N = 200$.

We considered two simulation configurations for p_1 and p_2 with fixed values of $p_1 = .5$ and $p_1 = .2$ for the first and second simulation configurations, respectively. For each

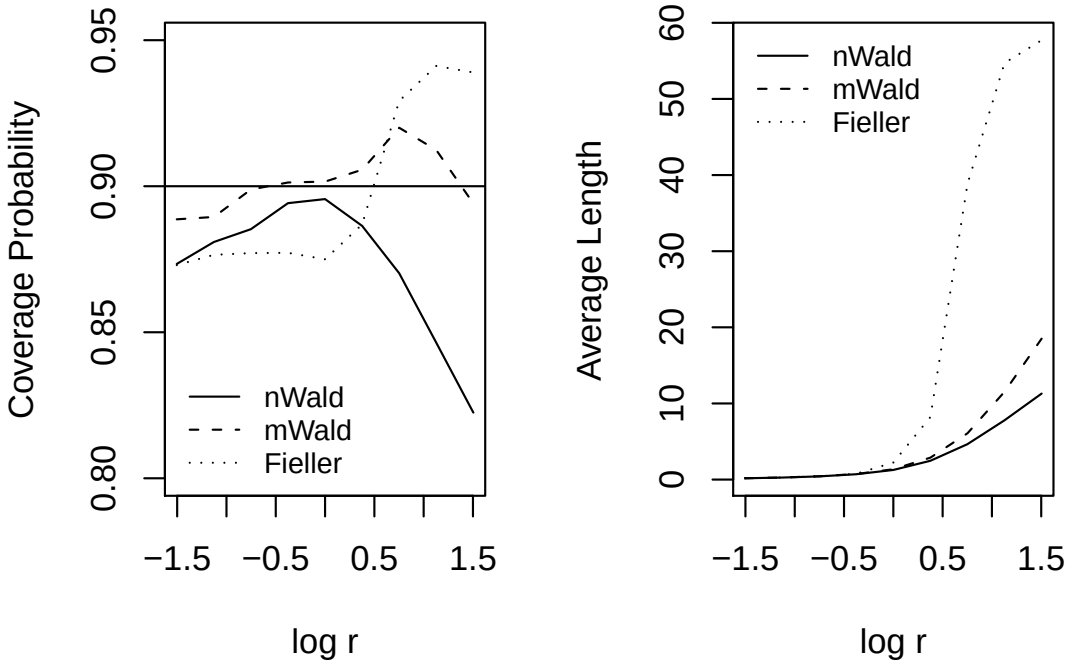

Figure 5.4: Coverage probabilities and average lengths versus the log risk ratios r, where $p_1 = .2$. The false-positive rate is $\phi = .1$, the false-negative rate is $\theta = .1$, the total sample size $N = 200$, and $s = n/N = 0.2$.

simulation configuration, we chose 9 values of p_2, $\{p_{2,1}, \ldots, p_{2,9}\}$, in an increasing order, such that $\log(r_1)$ and $\log(r_9)$ were symmetric about 0, and $\{\log(r_1), \ldots, \log(r_9)\}$ were equally spaced, where $r_t = p_1/p_{2,t}, t = 1, \ldots, 9$. We let $p_{2,9} = .9$ for both configurations. Using the assumption that $\log(r_1)$ and $\log(r_9)$ are symmetric about 0, we obtained $p_{2,1} \approx 0.278$ and $p_{2,1} \approx 0.044$ for the two configurations, respectively. Note that in this way, we ensured that the values of the parameters $\{p_{2,1}, \ldots, p_{2,9}\}$ were between 0 and 1. For each simulation configuration, we then determined $p_{2,2}, \ldots, p_{2,8}$ such that $\{\log(r_1), \ldots, \log(r_9)\}$ were equally spaced.

For each simulation scenario with known p_1, p_2, ϕ, θ, n/N, and N, we simulated $K = 10,000$ data sets. The simulation of one data set was described previously in this section. Once the K CIs for each interval method were obtained, we calculated the coverage probabilities (CPs) and average lengths (ALs). Finally, we plotted the CPs and ALs versus $\log r$ for each CI method.

Figures 5.3 and 5.4 display plots of the CPs and ALs of all CI methods versus $\log r$ for both configurations of p_1 and p_2, respectively. Figure 5.3 shows that both the nWald and the mWald CIs had close-to-nominal coverage for the range of $\log r$ studied here. The Fieller CI also had close-to-nominal coverage for the range of $\log r$, although the coverage was consistently slightly below the nominal level. Figure 5.3 also displays that the Fieller CI was slightly wider than the other two CIs. Figure 5.4 shows that the mWald CI had close-to-nominal coverage for the range of $\log r$ studied here. The nWald CI had close-to-nominal coverage when $\log r \in (-\log .5, \log .5)$ but much below-nominal coverage otherwise. The Fieller CI had below-nominal coverage when $\log r < .5$ and above-nominal coverage when $\log r > .5$. Figure 5.4 also displays that the mWald CI was slightly wider than the nWald CI. The Fieller CI was the widest and was much wider than the other two CIs when $\log r > .5$.

5.6 Discussion

In this article, we have considered interval estimation of the risk ratio of two binomial proportion parameters using two-sample misclassified binomial data. Because the original full likelihood function was difficult to work with, we have performed a reparameterization of the parameters. The transformed parameters in the new likelihood function were separable and, therefore, the maximum likelihood estimation was straightforward. As a result, we have derived closed-form formulas for the MLE and the nWald, the mWald, and the Fieller CIs, for the risk ratio. The nWald CI was computed using a naive application of the Wald method; the mWald CI was based on a modified Wald method that guarantees nonnegative CI limits; and the Fieller CI was constructed using an asymptotic pivotal quantity. All three CIs are easy to compute and require little computing resources.

To illustrate, all three CIs were applied to a cervical cancer data set. As expected, they produced similar CIs because the cervical cancer data have a large

sample size. To compare and evaluate these three CIs, we conducted several Monte Carlo simulation studies to examine the CPs and ALs of all three CIs for r under various parameter-configuration scenarios. Because the CI estimators were developed based on asymptotic theory, we expected all three methods to perform well for large samples. This assumption was confirmed in our simulations because the CPs were close to the nominal level for large samples and the ALs decreased as sample sizes increased.

Substantial differences in performance occurred among these three CIs. We remark that the mWald CIs had CPs close to nominal level under various parameter and sample-size scenarios. Compared with the mWald CIs, the nWald CIs were narrower but tended to have CPs less than the nominal level, especially when p_1 and p_2 were close to zero or one and the sample sizes were small ($N < 200$). The Fieller CIs generally were the widest and sometimes were much wider than the other two intervals. The behavior of the Fieller CIs was somewhat erratic because the CPs could be above or below the nominal levels, especially when p_1 and p_2 were close to zero or one and the sample sizes were small ($N < 200$). In summary, the mWald CIs consistently had nominal coverage and performed the best among three CI methods for parameter and sample-size configurations considered here and, therefore, are preferred to the nWald and Fieller intervals for the parameter and sample-size configurations considered here.

BIBLIOGRAPHY

Boese, D. H. (2003), "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.

Bross, I. (1954), "Misclassification in 2x2 tables," *Biometrics*, 10, 478–486.

Evans, M., Guttman, I., Haitovsky, Y., and Swartz, T. (1996), "Bayesian analysis of binary data subject to misclassification," in *In Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, eds. Berry, D., Chaloner, K., and Geweke, J., John Wiley, pp. 67–77.

Fieller, E. C. (1954), "Some problems in interval estimation." *Journal of the Royal Statistical Society, Series B*, 16, 175–185.

Gaba, A. and Winkler, R. L. (1992), "Implications of errors in survey data: a Bayesian model," *Management Science*, 38, 913–925.

Goldberg, J. D. (1975), "The Effects of Misclassification on the Bias in the Difference Between Two Proportions and the Relative Odds in the Fourfold Table," *Journal of the American Statistical Association*, 70, 561–567.

Gustafson, P. (2005), "On Model expansion, model contraction, identifiability and prior information: two Illustrative scenarios involving mismeasured variables," *Statistical Science*, 20, 111–140.

Gustafson, P., Le, N. D., and Saskin, R. (2001), "Case-control analysis with partial knowledge of exposure misclassification probabilities," *Biometrics*, 57, 598–609.

Hildesheim, A., Mann, V., Brinton, L. A., Szklo, M., Reeves, W. C., and Rawls, W. E. (1991), "Herpes simplex virus type 2: a possible interaction with human papillomavirus types 16/18 in the development of invasive cervical cancer," *International Journal of Cancer*, 49, 335–340.

Hochberg, Y. (1977), "On the use of double sampling schemes in analyzing categorical data with misclassification errors," *Journal of American Statistical Association*, 72, 914–921.

Hong, Y., Meeker, W., and Escobar, L. (2008), "Avoiding problems with normal approximation confidence intervals for probabilities," *Technometrics*, 50, 64–68.

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.

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., Lin, H.-M., and Williamson, J. M. (2004), "Design and Analytic Considerations for Single-Armed Studies with Misclassification of a Repeated Binary Outcome," *Journal of Biopharmaceutical Statistics*, 14, 229–247.

Moors, J. J. A., van der Genugten, B. B., and Strijbosch, L. W. G. (2000), "Repeated audit controls," *Statistica Neerlandica*, 54, 3–13.

Perry, M., Vakil, N., and Cutler, A. (2000), "Admixture With Whole Blood Does Not Explain False-Negative Urease Tests," *Journal of Clinical Gastroenterology*, 30, 64–65.

Stamey, J. D., Seaman, J. W., and Young, D. M. (2007), "Bayesian Estimation of Intervention Effect with Pre- and Post-Misclassified Binomial Data," *Journal of Biopharmaceutical Statistics*, 17, 93–108.

Tenenbein, A. (1970), "A double sampling scheme for estimating from binomial data with misclassifications," *Journal of American Statistical Association*, 65 (331), 1350–1361.

— (1972), "A double sampling scheme for estimating from multinomial data with applications to sampling inspection," *Technometrics*, 14, 187–202.

Viana, M., Ramakrishnan, V., and Levy, P. (1993), "Bayesian analysis of prevalence from results of small screening samples," *Commun. Statist. –Theory Meth.*, 22, 575–585.

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," *Applied Statistics*, 44, 227–242.

Zhong, B. (2002), "Evaluating Qualitative Assays Using Sensitivity and Specificity," *Journal of Biopharmaceutical Statistics*, 12, 409–424.