Non-inferiority Testing for Correlated Ordinal Categorical Data with Misclassification

HAN, Yuanyuan

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Statistics

The Chinese University of Hong Kong
July 2011

Thesis/Assessment Committee

Professor Cheung Siu-Hung (Chair)
Professor Poon Wai-Yin(Thesis Supervisor)
Professor Wu Ka-Ho (Committee Member)
Professor Tang Man-Lai (External Examiner)

Abstract of thesis entitled:

Non-inferiority Testing for Correlated Ordinal Categorical Data with Misclassification

Submitted by Han, Yuanyuan

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in July 2011

When a new treatment comes out, it is likely to find benefits of the new one, such as fewer side effects, greater convenience of employment, or lower cost in terms of money and time. Therefore, the more appropriate research question is whether the new one is non-inferior or equivalent to, but not necessarily superior to the reference treatment. Consequently, the non-inferiority test or equivalence test is widely used in medical research, which is oriented towards showing that the difference of effect between the two treatments probably lies in a tolerance interval with the pre-defined lower or upper bounds. In this thesis, we consider non-inferiority tests when the data are ordinal categorical. In particular, we are interested in correlated data. We will develop non-inferiority testing procedures for data that are obtained by the paired design and three-armed design. We take advantage of a latent normal distribution approach to model ordinal categorical data.

Moreover, misclassification is frequently encountered in collecting ordinal categorical data. We also consider the non-inferiority test based on the data with misclassification. We have explored two different approaches. The first approach can be applied when misclassification probabilities are known or can be calibrated. The second approach deals with the case when we have partially validated data that provide the information on misclassification. The proposed approaches have wide applications that are not confined to tests in medical research. We design a substantive study to illustrate the practicality and applicability of the proposed approaches.

Keywords: Non-inferiority Test, Bootstrap, Misclassification, Partially Validated Data.

摘 要

當一種新的治療方法出現，我們可能發現它的好處主要來自于副作用較少、使用較方便、或成本較低。在此條件下，我們不一定需要新方法在效果上顯著好於現有方法，而只需要其在預先設定的範圍內不次於或等效於現有方法。因此，當給定一個可容許區間的情況下，非劣效性檢驗和等效檢驗在醫學領域得到了廣泛的應用。在本文中，我們考慮具有相關性的有序分類數據的非劣效性檢驗問題。實驗設計包括配對設計和三維塊設計。我們使用潛在正態分佈方法對具有相關關係的有序分類數據建模。

此外，錯誤分類在有序分類數據收集的過程中經常出現。我們也提出用于有分類錯誤的有序分類數據的非劣效性檢驗。我們探討兩種不同的方法。第一種方法可以處理分類錯誤概率已知或者可以被校準的情況；另一種是當錯誤分類概率未知的情況，对此，我們則可以採用部份驗證數據的方法。這兩種方法不僅在醫學領域有廣泛應用，也能被廣泛應用到其他领域。在本文中，我們基于一个实际问题设计问卷、收集数据并應用上述的方法加以分析，以說明所提出方法的實用性和適用性。

關鍵字： 非劣效性檢驗，錯誤分類，部份驗證數據，自助抽樣。

Acknowledgement

I would like to thank my supervisor, Professor Poon Wai-Yin, for her patience and generosity of encouragement and supervision during my research study. She truly cares for students whom she supervises and I am really grateful about this. Her invaluable guidance helped me with every aspect for writing of this thesis. And I would like to express my gratitude to Professor Poon Wai-Yin for her help to my 4-year PhD study.

I also thank Professor Tang Man-Lai, Professor Cheung Siu-Hung, and Professor Wu Ka-Ho for their constructive comments.

Finally, I would like to dedicate this thesis to my parents for their support and love.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

When a new treatment comes out, it is likely to find benefits of the new one, such as fewer side effects, greater convenience of employment, or lower cost in terms of money and time. Therefore, the more appropriate research question is whether the new one is non-inferior or equivalent to, but not necessarily superior to the reference treatment. Consequently, the non-inferiority test or equivalence test is widely used in medical research, which is oriented towards showing that the difference of effect between the two treatments probably lies in a tolerance interval with the pre-defined lower or upper bounds.

For continuous variables, non-inferiority is often assessed by comparing the two mean responses as presented in Ng (2001) and Liu et al. (2002). Tang & Tang (2004) and Munk et al. (2007) considered the methods for the assessment of non-inferiority with binary data. Tang & Poon (2007) developed methods to conduct equivalence and non-inferiority tests when there are two independent treatments with data that are in ordinal scale. However, little has been achieved for non-inferiority for correlated ordinal outcomes. We generalize the work of Tang & Poon (2007) and use the latent normal distribution model to account for non-inferiority tests with correlated ordinal data. The latent normal model assumes that the observed ordinal categorical variables are crude measurements of latent continuous variables which are multivariate normally distributed.

Although the two-armed non-inferiority test that compares two treatments including an active control and an experimental treatment is frequently employed, Tang & Tang (2004) and Hung et al. (2007) are of the view that the applications of a non-inferiority trial design that does not contain a placebo arm are controversial. Thus the results from such a trial design can be rather difficult to interpret with confidence. One remedy to this problem is to add a placebo arm, which is widely acknowledged, see D'Agostino et al. (2003) and Pigeot et al. (2003). Three-armed trial enables to addressing the important issue of assay sensitivity that refers to the ability of a test to detect differences between effective, less effective, and ineffective therapies.

Methods have been proposed in the literature to test the non-inferiority of the treatments in a three-armed design. For example, Koch and Tangen (1999) and Pigeot et al. (2003) have developed methods for three-armed non-inferiority tests. However, they operated on the assumption that the outcome variables from the placebo, active controlled, and the new experimental groups were all independently normally distributed with the same variance. In this thesis, we will relax both the independent and the continuous-scale normally-distributed-data assumptions and develop non-inferiority test for three-armed design data.

Another important consideration in this thesis is to develop non-inferiority testing procedures that can handle data sets with misclassification. It has been noted in Yiu & Poon (2008) that misclassification is frequently encountered in collecting ordinal categorical data. Misclassification occurs when the responses do not reflect the true state of a participant of the study and hence the state of the participant is misclassified to an inappropriate cell in the contingency table, producing a contingency table with misclassified data. Yiu & Poon (2008) proposed two approaches to analyze ordinal categorical data with misclassifications.

In the case with known misclassification probabilities, Yiu & Poon (2008) used the direct optimization method to get the maximum likelihood (ML) estimates of the parameters of a latent variable normal model. In the case with unknown misclassification probabilities, Yiu & Poon (2008) used a latent variable normal model and a two-stage estimation procedure. In the first stage, the ML estimates of the cell probabilities are obtained by maximizing the likelihood function. Based on these estimated cell probabilities, the model parameters are estimated through a modified minimum chi-squared approach. Poon & Wang (2010) also worked on a latent variable normal model. They developed a unified EM approach to find the ML estimates of the model parameters. Although EM approach is computationally more efficient for multivariate problem, the computational time may be longer than the two-stage estimation procedure in the low-dimensional designs. We will develop a non-inferiority testing procedure that can be applied to correlated ordinal categorical data.

Non-inferiority test in the three-armed design with misclassified categorical ordinal data has not been addressed in the literature. We will also develop procedure for the analysis of three-armed multivariate ordinal categorical data with misclassification. Non-inferiority testing procedures for ordinal categorical data with possible misclassification have wide applicability. We designed a substantive study to illustrate how these procedures can be applied in research studies. We have designed a study with the real data set on How Effective the Learning Paths (HELP) are obtained through questionnaire. The study investigated the effectiveness of four different learning approaches of university students. Students were asked to express their opinions on the four approaches by means of rating the degree of agreement on the effectiveness of the four different approaches. The data set consists of responses that correspond to the true states of the variables of interest and responses on surrogate variables with misclassifications. The study shows that the proposed partially validated data method to be introduced in Chapter 5 can be used in a flexible manner to facilitate a more comprehensive and informative data analysis.

Depending on different types of ordinal categorical data, we develop various tests for non-inferiority tests with different models for ordinal categorical data. The thesis is organized as follows. Chapter 2 addresses the non-inferiority test for the paired data, and in Chapter 3 we further discuss non-inferiority test for paired data with misclassification. In Chapter 4, non-inferiority test in three-armed block design is introduced, and in Chapter 5 testing procedures for three-dimensional data set with misclassification are introduced. Specifically, in Chapter 3 and Chapter 5, we present two approaches that can handle known and unknown misclassification probabilities, respectively. In each of the chapter from Chapters 2 to 5, the proposed methods have been applied to analyze real data sets, and simulation studies have been used to assess the proposed procedures. The assessment results are reported separately in each chapter.

It is worthy of note that most of the works in relation to three-armed designs in the literature have been focused on non-inferiority rather than equivalence test. In light of this, our focus will be on non-inferiority test. However, the nature of testing equivalence and testing non-inferiority are the same. We will discuss the equivalence test briefly in Chapter 6 that concludes the thesis with a discussion.

Chapter 2

Paired Data

2.1 Introduction

In many research studies, researchers need to deal with pairwise ordinal data. For example, in medical research with one treatment and one control, we may use paired data design to address the heterogeneity of the initial conditions of the test subjects and to enhance generalizability of the research results. When we have a sample of n pairs of observations, with responses classified by two ordinal variables, say reference group (Z_R) and treatment group (Z_T), then we may summarize the paired data into a contingency table. Assuming the numbers of categories are the same for the two variables, i.e., both have categories from 1 to K, where K is the number of categories, then the responses can be summarized in a K x K contingency table that is presented in Table 2.1.

Table 2.1: Two-way contingency table for matched pairs with ordinal data

Z_R \ Z_T	1	2	3	$\cdots$	K	Total
1	n_{11}	n_{12}	n_{13}	$\cdots$	n_{1K}	n_{1+}
2	n_{21}	n_{22}	n_{23}	$\cdots$	n_{2K}	n_{2+}
3	n_{31}	n_{32}	n_{33}	$\cdots$	n_{3K}	n_{3+}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
K	n_{K1}	n_{K2}	n_{K3}	$\cdots$	n_{KK}	n_{K+}
Total	n_{+1}	n_{+2}	n_{+3}	$\ddots$	n_{+K}	n

Let n_{ij} be the observed frequency in the (i, j)th cell, then $\sum_{i=1}^{K}\sum_{j=1}^{K} n_{ij} = n$. To analyze the ordinal data, one of the commonly used approaches is to consider a model with underlying continuous variables, say X_R and X_T. We also assume that $(X_R, X_T)^T$ are bivariate normally distributed as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_R^2 & \sigma_{RT} \\ \sigma_{RT} & \sigma_T^2 \end{pmatrix}.$$

The underlying continuous variables X_R and X_T are related to the observed ordinal categorical variables Z_R and Z_T. Specifically, the relationship is given by

$$Z_i = k \quad \text{if } \beta_k \le X_i < \beta_{k+1} \quad \text{for } i = R, T; k = 1, \ldots, K,$$

where β_1 = -∞, $\beta_{K+1} = \infty$ and $\boldsymbol{\beta} = (\beta_2, \ldots, \beta_K)^T$ is a vector of unknown thresholds. Here, to enable straightforward comparison between μ_T and μ_R, we follow Tang & Poon (2007) and assume that the two sets of thresholds are the same for X_R and X_T.

Since the location and dispersion of the underlying continuous variables are not defined, we fix the population mean and variance for the underlying continuous variable of the reference group. That is, we set E(X_R)=0 and Var(X_R)=1, and we have

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ \mu_T \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma_{RT} \\ \sigma_{RT} & \sigma_T^2 \end{pmatrix}. \tag{2.1}$$

The means and variances of the two underlying variables are then compared in a relative sense. Let $\boldsymbol{\theta} = (\mu_T, \sigma_{RT}, \sigma_T^2, \boldsymbol{\beta})^T$ be the unknown parameter vector of the model. The parameter σ_{RT} is the covariance of X_R and X_T, μ_T and σ_T^2 are the mean and variance of X_T, which can be interpreted in a relative sense to those of X_R. For example, if the value of μ_T is positive, the location of X_T lies to the right of X_R, and if σ_T^2 is less than 1, the variation of X_T is less than that of X_R. Please refer to Poon (2004) for more details.

Assuming p_{ij} to be the probability of an observation that falls in the (i,j)th cell, we have

$$\begin{aligned} p_{ij} &= \Pr(Z_R = i, Z_T = j) \\ &= \Phi_2(\beta_{i+1}, \beta_{j+1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) - \Phi_2(\beta_{i+1}, \beta_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\quad -\Phi_2(\beta_i, \beta_{j+1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \Phi_2(\beta_i, \beta_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned} \tag{2.2}$$

where $\Phi_2(\beta_i, \beta_j; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the bivariate normal distribution function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, evaluating at β_i and β_j. The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ can be obtained by maximizing the log likelihood function

$$\sum_{i=1}^{K} \sum_{j=1}^{K} n_{ij} \log p_{ij}(\boldsymbol{\theta}). \tag{2.3}$$

This latent normal distribution approach has been used extensively to analyze ordinal categorical data. However, its applications in the context of equivalence or non-inferiority tests have not been explored. Some initial works in this regard are available in Tang and Poon (2007).

2.2 Non-inferiority Test

When a new treatment is readily available to be applied, it is likely to find its benefits, such as fewer side effects, greater convenience of employment, or lower cost in terms of money and time. Therefore, to compare the newly developed treatment and the existing reference treatment, a more appropriate way to approach the problem is to examine whether the new treatment is non-inferior to the existing reference treatment. The non-inferiority test is oriented to show that the effect difference between the two treatments is probably to lie in a tolerance interval with pre-defined lower bound.

Based on the latent normal distribution model that is described in Section 2.1, we develop a procedure that can be applied to ordinal categorical data to assess the non-inferiority of two treatments. Specifically, let $\mu_T - \mu_R$ be the discrepancy in location between the two treatments under condition, the non-inferiority test is accomplished by testing

$$\begin{aligned} H_0 &: \mu_T - \mu_R \leq -\Delta \\ H_1 &: \mu_T - \mu_R > -\Delta. \end{aligned} \tag{2.4}$$

Here, Δ is a positive value, known as the non-inferiority margin, which can be predefined by the experience in practice and in need. The non-inferiority of two treatments can be concluded if the null hypothesis is rejected, *i.e.*, the lower bound of the $100(1-\alpha)\%$ confidence interval (CI) for $\mu_T - \mu_R$ is greater than $-\Delta$ given significance level α (Blackwelder, 1998).

Based on the model in (2.1) with $\mu_R = 0$, the non-inferiority test can be simplified as follows

$$\begin{aligned} H_0 &: \mu_T \leq -\Delta \\ H_1 &: \mu_T > -\Delta. \end{aligned} \tag{2.5}$$

2.3 Implementation

The MLE $\hat{\boldsymbol{\theta}} = (\hat{\mu}_T, \hat{\sigma}_{RT}, \hat{\sigma}_T^2, \hat{\boldsymbol{\beta}})^T$ of $\boldsymbol{\theta} = (\mu_T, \sigma_{RT}, \sigma_T^2, \boldsymbol{\beta})^T$ can be obtained by maximizing the log likelihood function in (2.3). To find the CI of μ_T for conducting the non-inferiority test, we also need a way to get the standard errors of $\hat{\mu}_T$. In practice, one can implement the program Mx developed by Neale et al. (1999) to produce parameter estimates and standard errors. Mx is a widely used software and it can be downloaded at no cost. It is recommended for two reasons. Firstly, it is widely available and is a free software available in the public domain. Secondly, it allows flexibility in setting the constraints for the parameters. However, the standard errors produced by the function "*SE*" in Mx are not stable. As the new feature of Mx enables the generation of bootstrap standard error, so we make use of the bootstrap method to find the standard errors (*SE*) and then construct one-sided $100(1-\alpha)\%$ CI for μ_T by $(\hat{\mu}_T - z_\alpha SE(\hat{\mu}_T), +\infty)$, where z_α is the upper α probability point of the standard normal distribution. To assess the performance of the CIs such constructed, we need to first assess the performance of the *SE*. The assessment is accomplished by simulation studies with results presented in the next two sections.

2.4 Simulation

To examine the performance of the proposed method by simulation, a number of data sets were generated based on a set of known parameter values. We set K equal to 3, and the thresholds for the two variables are set to be identical to $\beta_2 = -0.2$ and $\beta_3 = 0.5$. In other words, the true values are $(\mu, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2))^T = (0.2, 0.3, 1, (-0.2, 0.7))^T$. The number of simulation replications is 1000. We considered three different sample sizes: 300, 100 and 80.

Two statistics were computed for the simulation of each sample size as performance indicators. They are the mean of all the estimates for each parameter and the root mean squared error (RMS) compared to the true parameter value across the 1000 replications. Specifically, let $\hat{\theta}_j^{(i)}$ be the estimate of the j-th element in $\boldsymbol{\theta}$ in the i-th replication, we computed

$$\bar{\hat{\theta}}_j = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\theta}_j^{(i)},$$

and

$$RMS(\hat{\theta}_j) = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_j^{(i)} - \theta_{0j})^2},$$

where θ_{0j} is the true value of the j-th element θ_j of $\boldsymbol{\theta}$. Moreover, the ratio R is used to assess the accuracy of standard error estimate for the parameter, say, $\hat{\theta}_j$. The empirical sampling standard deviation based on the 1000 replications $SD(\hat{\theta}_j)$ and the mean across the 1000 replications of the standard error estimates $\bar{SE}(\hat{\theta}_j)$ were used to compute the ratio

$$R = SD(\hat{\theta}_j)/\bar{SE}(\hat{\theta}_j).$$

If R is close to 1, it indicates that the standard error estimates are accurate.

We have summarized the simulation results in the following three tables.

Table 2.2: Simulation result for sample size 300

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.2	0.7
Mean	0.200	0.301	1.040	-0.202	0.702
RMS	0.085	0.081	0.292	0.071	0.069
SE	0.083	0.084	0.307	0.071	0.068
SD	0.085	0.081	0.289	0.071	0.069
R	1.023	0.969	0.941	0.997	1.006

Table 2.3: Simulation result for sample size 100

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.2	0.7
Mean	0.204	0.318	1.172	-0.205	0.713
RMS	0.145	0.153	0.626	0.122	0.118
SE	0.151	0.162	0.698	0.106	0.105
SD	0.145	0.152	0.602	0.122	0.118
R	0.963	0.936	0.862	1.146	1.120

Table 2.4: Simulation result for sample size 80

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.2	0.7
Mean	0.194	0.305	1.165	-0.210	0.713
RMS	0.163	0.167	0.731	0.131	0.124
SE	0.169	0.182	0.873	0.113	0.112
SD	0.163	0.167	0.712	0.131	0.123
R	0.966	0.919	0.816	1.152	1.100

From the results in Tables 2.2-2.4, we can see that the mean values of the estimates are very close to the true values in all situations, and the root mean squared errors are reasonably small. The values of R are all between 0.8 and 1.2. Therefore, the simulation results indicate that the proposed estimation procedure can provide the users of acceptable parameter estimates and standard errors no matter the sample size is large or not, and that the larger sample size leads to more accurate parameter estimates (smaller RMS) and standard errors.

The simulation results suggest that these estimates and standard errors are reliable and therefore applicable for constructing confidence intervals for non-inferiority test.

2.5 Bootstrap Confidence Interval

In testing non-inferiority, the confidence interval approach rather than hypothesis testing method is widely recommended (Blackwelder, 1998). As stated in Section 2.3, a one-sided $100(1-\alpha)\%$ CI for μ_T is given by $(\hat{\mu}_T - z_\alpha SE(\hat{\mu}_T), +\infty)$, where z_α is the upper α probability point of the standard normal distribution and $SE(\hat{\mu}_T)$ is the standard error of $\hat{\mu}_T$ produced by the bootstrap method.

To evaluate the performance of CIs empirically, a commonly used evaluation standard is the expected coverage probability

(ECP). Specifically, in the context of non-inferiority test, let L be the lower bound of the produced CI and μ_{T0} the true value of the parameter μ_T, we use

$$ECP = \frac{1}{1000} \sum_{j=1}^{1000} I(L \leq \mu_{T0}). \tag{2.6}$$

We expect that the coverage probability is approximately equal to $1 - \alpha$. The approximate CI is conservative if the coverage probability is larger than $1 - \alpha$, while it is liberal if the coverage probability is smaller than $1 - \alpha$. As 95% CIs are constructed in the simulation studies, we expect that the coverage probability of these CIs on the true population values would be approximately 95%.

We have different sample sizes in the simulation: 300, 100 and 80, and the simulation results based on 1000 replications are as follows.

Table 2.5: Evaluation of the performance of CIs

NOBS	Covering True Value Times	ECP
300	953	0.953
100	959	0.959
80	970	0.970

The values of ECP are quite reasonable but are slightly larger than 0.95, especially when the sample size is small. The result indicate that this method produces CIs that are slightly conservative when the sample size is not large enough. However, this problem improves when sample size becomes larger.

2.6 Examples

2.6.1 Example 1

Table 2.6 reports the result of an example with data sampled from British Royal Ordnance factories during 1943-1946. The ordered contingency table summarizes the right eye and left eye unaided distant vision grade in 7477 women employees, aged 30-39 years. Obviously, the measurements of the two eyes are likely correlated and naturally form paired data. This data set is analyzed by many authors, and most relevant to our research is available in the study by Lui & Cumberland (2001), where they drew the conclusion on equivalence of the two eyes by comparing the marginal probabilities of the two variables.

Table 2.6: Right and left eye unaided distance vision for women in Stuart(1955)

Right eye \ Left eye	Highest(1)	Second(2)	Third(3)	Lowest(4)	Total
Highest(1)	1520	266	124	66	1976
Second(2)	234	1512	432	78	2256
Third(3)	117	362	1772	205	2456
Lowest(4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

As an illustration of our method, we analyzed the data set by using the proposed latent variable approach and test non-inferiority by testing the hypothesis (2.5).

Let $\Delta = 0.2$ and we consider the data for the left eye to be the data for treatment group and those for the right eye to be those for the reference group, then the ML estimate of the mean of the treatment group is $\hat{\mu}_T = 0.0326$. The 95% confidence interval is $(-0.022, +\infty)$ with the lower bound -0.022 greater than -0.2 at 0.05 level of significance. We conclude that unaided distance vision of the left eye group is non-inferior to that of the right eye group. Our conclusion is consistent with the result in Lui & Cumberland (2001) from the non-inferiority test perspective.

2.6.2 Example 2

This example is to compare the diagnostic accuracy between the digitized film and the plain film in detecting the malignant lesions for breast cancer. Table 2.7 presents the mammography data in Zhou et al. (2002) consisting of 45 subjects, and they are classified into one of the four categories by using the digitized and plain films in a mammography study. The four categories are normal (coded 1), benign (coded 2), probably benign (coded 3), and suspicious (coded 4).

Table 2.7: The 45 subjects cross-classified by using digitized and plain films

Digitized film \ Plain film	1	2	3	4	Total
1	17	2	2	1	22
2	1	4	2	1	8
3	1	1	4	1	7
4	2	0	3	3	8
Total	21	7	11	6	45

1=normal, 2=benign, 3=probably benign and 4=suspicious

Zhou et al. (2002) implemented Schuirmann's method (1987), which was based on non-parametric estimate of the area under receiver operating characteristic (ROC) to assess the equivalence of the two types of films. Their result suggests that there is no sufficient evidence that the plain film and digitized film have equivalent diagnostic accuracy. The digitized film is a more advanced technology than the plain film. However, the plain film does not need the facility like the workstation, and it can be applied widely. Let digitized film be the reference group and plain film be the treatment group, so we want to test whether the plain film is non-inferior to the digitized film in diagnostic accuracy.

We set $\Delta = 0.2$, and we get 95% one-sided confidence interval as $(-0.164, +\infty)$. The lower bound -0.164 is larger than -0.2. Thus, we may draw the conclusion that the plain film is non-inferior to the digitized film in terms of accuracy.

2.6.3 Example 3

This data set is taken from Ezzet & Whitehead (1993). It is from a two-treatment, two-period crossover trial, which is aiming to compare the instructions of two inhalation devices for delivering the drug salbutamol in 286 asthma patients. Patients were asked to rate the clarity of leaflet instructions accompanying each device, using a 4-point ordinal scale. In Table 2.8, we have summarized the entries for crossover study on treatment A and B in the table. As an illustrative example, we let Treatment B be the reference group and Treatment A be the treatment group.

The objective is to test whether the introduction for inhaler A is non-inferior to B in terms of easy-to-understand language.

Table 2.8: Clarity of inhaler A and B Instructions in AB sequence order

Treatment A (period 1) \ Treatment B(period 2)	1	2	3	4	Total
1	59	35	3	2	99
2	11	27	3	2	41
3	0	0	0	0	0
4	1	1	0	0	2
Total	71	63	5	3	142

1: Easy, 2: Only clear after re-reading, 3: Not very clear, 4: Confusing

In light of no previous comparison conclusion, we illustrate this example with the most prevailing non-inferiority margin $\Delta = 0.2$. In AB order, the confidence interval is $(0.058, +\infty)$ with 0.058 greater than -0.2 at significance level of 0.05. We can conclude that the clarity of introductions of A is non-inferior to B.

Chapter 3

Paired Data with Misclassification

3.1 Introduction

In Chapter 2, it is assumed that the data collected reflect the true state of the respondents. However, respondents in many studies do not report the true state due to various reasons, leading to misclassification of the responses. The use of non-precise measurement instruments may also lead to misclassification.

The objective of this chapter is to develop statistical procedures for testing non-inferiority for data with misclassification. Two approaches have been developed by Yiu & Poon (2008) and Zhang (2007) for analyzing ordinal categorical data with misclassification. In the first approach, misclassification probabilities are assumed to be known or can be calibrated. In the second approach, these probabilities are assumed to be unknown and information on misclassification is obtained by analyzing data that is collected using the method of partially validated data.

The models that are based on underlying normally distributed variables and the estimation of the models for these two approaches are presented in the following two sections. Moreover, we develop procedures to test non-inferiority, and assess the performance of the proposed procedures by simulation studies.

3.2 Model with Known Misclassification Probabilities

Suppose we have a $K\times K$ contingency table with N observations. Let $\boldsymbol{k} = (k(1), k(2))$ and $n_{\boldsymbol{k}}$ the observed frequency in cell $\boldsymbol{k}$, where $k(1), k(2) = 1, 2, \cdots, K$. $\boldsymbol{E}_k$ is used to denote the $K\times K$ matrix with element $\boldsymbol{k}$ equal to 1 and all other elements equal to 0. $\boldsymbol{Z}_j$ is a $K\times K$ matrix with $Z_{\boldsymbol{k},j}$ as its elements in cell $\boldsymbol{k}$, which is used to represent the classification of the j-th subject. $\boldsymbol{X}_j$ is a $K\times K$ matrix with $X_{\boldsymbol{k},j}$ as its elements in cell $\boldsymbol{k}$ and is used to represent the true classification or true state of the j-th subject.

We also have

$$X_{\boldsymbol{k},j} = \begin{cases} 1 & \text{if the } j\text{-th subject truly belongs to the cell } \boldsymbol{k}, \\ 0 & \text{otherwise} \end{cases}$$

$$Z_{\boldsymbol{k},j} = \begin{cases} 1 & \text{if the } j\text{-th subject is classified into the cell } \boldsymbol{k}, \\ 0 & \text{otherwise} \end{cases}$$

$$Y_j = \begin{cases} 1 & \text{if the } j\text{-th subject tells the truth or is correctly classified,} \\ 0 & \text{otherwise} \end{cases}$$

for $j = 1, \ldots, N$.

Let $\boldsymbol{u} = (u(1), u(2))$, where $u(1), u(2) = 1, 2, \cdots, K$, we define the several probabilities that are involved in the model as follows.

(a) $p_{\boldsymbol{u}}$ is the probability that a subject actually belongs to the cell $\boldsymbol{u}$.

$$P(\boldsymbol{X}_j = \boldsymbol{E}_{\boldsymbol{u}}) = p_{\boldsymbol{u}}.$$

These probabilities satisfy the constraint $0 \leq p_{\boldsymbol{u}} \leq 1$ and

$$\sum_{u(1)=1}^{K} \sum_{u(2)=1}^{K} p_{\boldsymbol{u}} = 1.$$

Similar to Chapter 2, we assume that the true classifications are related to some underlying continuous variables and hence $p_{\boldsymbol{u}}$ is of the form (2.2). That is, $p_{\boldsymbol{u}}$ is a function of the unknown parameters in mean, thresholds and the covariance matrix.

(b) $\tau_{\boldsymbol{u},j}$ is the probability of the j-th subject being classified correctly given that the subject actually belongs to the cell $\boldsymbol{u}$. That is

$$P(Y_j = 1 | \boldsymbol{X}_j = \boldsymbol{E}_{\boldsymbol{u}}) = \tau_{\boldsymbol{u},j},$$

where the $\tau_{\boldsymbol{u},j}$'s are supposed to be known and are called the honesty probabilities. On the other hand, the probability of the j-th subject being misclassified given that the subject actually belongs to cell $\boldsymbol{u}$ is given by

$$P(Y_j = 0|\boldsymbol{X}_j = \boldsymbol{E_u}) = 1 - \tau_{\boldsymbol{u},j}.$$

(c) $\gamma_{\boldsymbol{ku}}$ is the probability of a subject being classified into the cell $\boldsymbol{k}$ given that the subject actually belongs to the cell $\boldsymbol{u}$ and is misclassified.

$$P(\boldsymbol{Z}_j = \boldsymbol{E_k}|\boldsymbol{X}_j = \boldsymbol{E_u}, Y_j = 0) = \gamma_{\boldsymbol{ku}}.$$

where

$$0 \le \gamma_{\boldsymbol{ku}} \le 1, \gamma_{\boldsymbol{uu}} = 0,$$

and $\sum_{k(1)=1}^{K} \sum_{k(2)=1}^{K} \gamma_{\boldsymbol{ku}} = 1$.

3.2.1 Maximum likelihood estimation

Suppose that we have a data set of size N. Based on the afore-described model, let $z_{\boldsymbol{k},j}$ be the realization of the random variable $Z_{\boldsymbol{k},j}$. The log-likelihood function is then given by

$$\begin{aligned} l &= \log(\textstyle\prod_{j=1}^{N} \prod_{k(1)=1}^{K} \prod_{k(2)=1}^{K} [Pr(\boldsymbol{Z}_j = \boldsymbol{E_k})]^{z_{\boldsymbol{k},j}}) \\ &= \textstyle\sum_{j=1}^{N} \sum_{k(1)=1}^{K} \sum_{k(2)=1}^{K} z_{\boldsymbol{k},j} \log Pr(\boldsymbol{Z}_j = \boldsymbol{E_k}) \\ &= \textstyle\sum_{j=1}^{N} \sum_{k(1)=1}^{K} \sum_{k(2)=1}^{K} z_{\boldsymbol{k},j} \log(\sum_{u(1)=1}^{K} \sum_{u(2)=1}^{K} \psi_{\boldsymbol{ku},j} p_{\boldsymbol{u}}) \end{aligned} \quad (3.1)$$

where $p_{\boldsymbol{u}}$ is given by (2.2), and

$$\psi_{\boldsymbol{ku},j} = \begin{cases} \tau_{\boldsymbol{k},j} & \text{if } \boldsymbol{u} = \boldsymbol{k} \\ \gamma_{\boldsymbol{ku}}(1 - \tau_{\boldsymbol{u},j}) & \text{otherwise} \end{cases} \tag{3.2}$$

are the elements of the matrix Ψ_j, which is the matrix of transition probabilities with which the j-th subject is transited from the true state to another state with misclassification. It is a two-dimensional matrix of $(K \times K)$ by $(K \times K)$. For the problem with known misclassifications, $\psi_{\boldsymbol{ku},j}$ are known constants, and

$$\textstyle\sum_{\boldsymbol{k}} \psi_{\boldsymbol{ku},j} = 1 \qquad \text{for} \qquad 1 \leq j \leq N.$$

Similar to Chapter 2, the unknown parameter vector is $\boldsymbol{\theta} = (\mu_T, \sigma_{RT}, \sigma_T^2, \boldsymbol{\beta})^T$. From (2.2) and (3.1), the parameters in $\boldsymbol{\theta}$ are involved only in the cell probability $p_{\boldsymbol{u}}$, the ML estimate $\hat{\boldsymbol{\theta}}$ can be obtained by solving the likelihood equation

$$0 = \frac{\partial l}{\partial \theta_k} = \sum_{v(1)=1}^{K} \sum_{v(2)=1}^{K} \frac{\partial l}{\partial p_{\boldsymbol{v}}} \times \frac{\partial p_{\boldsymbol{v}}}{\partial \theta_k},$$

where $k = 1, ..., s$, and s is the number of unknown parameters. Taking into consideration the accessibility by practitioners, we also use Mx to obtain the ML estimates.

When $\hat{\boldsymbol{\theta}}$ becomes available, we can find CI for μ_T to examine non-inferiority by testing the hypothesis in (2.5). The CI is given by $(\hat{\mu}_T - z_\alpha SE(\hat{\mu}_T), +\infty)$, where z_α is the upper α probability point of the standard normal distribution and SE is the bootstrap standard error.

3.2.2 Simulation

To assess the performance of the CIs such constructed, we need to first assess the performance of the SE and then the CI. This is accomplished by simulation studies.

A number of data sets are generated based on a set of known parameter values to examine the performance of the proposed method. We set K equal to 3. The true parameter values are $(\mu_T, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2))^T = (0.2, 0.3, 1, -0.3, 0.6)^T$.

The number of simulation replications was 1000, and we considered three different sample sizes: 1000, 500 and 300. Assuming $\tau_{u,j} = \tau_u$ for all $j = 1, ..., N$, We examined three different misclassification probabilities: $\tau_u = 0.7, 0.8$ and 0.9.

The set of γ_{ku} values that were adopted to analyze the data is presented in Table 3.1. The values were compiled based on the assumption that misclassification only occurred between adjacent cells, and with equal probability.

The γ_{ku} values in combination with the τ_u values produce the transition misclassification probability matrix $\boldsymbol{\Psi}$, where $\boldsymbol{\Psi} = \boldsymbol{\Psi}_j$ for all j as $\tau_{u,j} = \tau_u$ for all j. As a result, we obtained three Ψ matrices that are presented in Tables 3.2 to 3.4. The simulation results are presented in Tables 3.5 to 3.13, with different values of τ_u indicating different misclassification levels. We used the same assessment standards as those in Chapter 2 to evaluate the proposed method.

Table 3.1: γ_{ku} values

		False								
		(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)	(3, 1)	(3, 2)	(3, 3)
True	(1, 1)	0	1/2	0	1/2	0	0	0	0	0
	(1, 2)	1/3	0	1/3	0	1/3	0	0	0	0
	(1, 3)	0	1/2	0	0	0	1/2	0	0	0
	(2, 1)	1/3	0	0	0	1/3	0	1/3	0	0
	(2, 2)	0	1/4	0	1/4	0	1/4	0	1/4	0
	(2, 3)	0	0	1/3	0	1/3	0	0	0	1/3
	(3, 1)	0	0	0	1/2	0	0	0	1/2	0
	(3, 2)	0	0	0	0	1/3	0	1/3	0	1/3
	(3, 3)	0	0	0	0	0	1/2	0	1/2	0

Table 3.2: Transition misclassification probability matrix $\mathbf{\Psi}$ for $\tau_{\boldsymbol{u}} = 0.7$

		False								
		$(1,1)$	$(1,2)$	$(1,3)$	$(2,1)$	$(2,2)$	$(2,3)$	$(3,1)$	$(3,2)$	$(3,3)$
True	$(1,1)$	0.7	0.15	0	0.15	0	0	0	0	0
	$(1,2)$	0.1	0.7	0.1	0	0.1	0	0	0	0
	$(1,3)$	0	0.15	0.7	0	0	0.15	0	0	0
	$(2,1)$	0.1	0	0	0.7	0.1	0	0.1	0	0
	$(2,2)$	0	0.075	0	0.075	0.7	0.075	0	0.075	0
	$(2,3)$	0	0	0.1	0	0.1	0.7	0	0	0.1
	$(3,1)$	0	0	0	0.15	0	0	0.7	0.15	0
	$(3,2)$	0	0	0	0	0.1	0	0.1	0.7	0.1
	$(3,3)$	0	0	0	0	0	0.15	0	0.15	0.7

Table 3.3: Transition misclassification probability matrix $\mathbf{\Psi}$ for $\tau_u = 0.8$

		False								
		$(1,1)$	$(1,2)$	$(1,3)$	$(2,1)$	$(2,2)$	$(2,3)$	$(3,1)$	$(3,2)$	$(3,3)$
True	$(1,1)$	0.8	0.1	0	0.1	0	0	0	0	0
	$(1,2)$	0.067	0.8	0.067	0	0.067	0	0	0	0
	$(1,3)$	0	0.1	0.8	0	0	0.1	0	0	0
	$(2,1)$	0.067	0	0	0.8	0.067	0	0.067	0	0
	$(2,2)$	0	0.05	0	0.05	0.8	0.05	0	0.05	0
	$(2,3)$	0	0	0.067	0	0.067	0.8	0	0	0.067
	$(3,1)$	0	0	0	0.1	0	0	0.8	0.1	0
	$(3,2)$	0	0	0	0	0.067	0	0.067	0.8	0.067
	$(3,3)$	0	0	0	0	0	0.1	0	0.1	0.8

Table 3.4: Transition misclassification probability matrix $\boldsymbol{\Psi}$ for $\tau_u = 0.9$

		False								
		$(1,1)$	$(1,2)$	$(1,3)$	$(2,1)$	$(2,2)$	$(2,3)$	$(3,1)$	$(3,2)$	$(3,3)$
True	$(1,1)$	0.9	0.05	0	0.05	0	0	0	0	0
	$(1,2)$	0.033	0.9	0.033	0	0.033	0	0	0	0
	$(1,3)$	0	0.05	0.9	0	0	0.05	0	0	0
	$(2,1)$	0.033	0	0	0.9	0.033	0	0.033	0	0
	$(2,2)$	0	0.025	0	0.025	0.9	0.025	0	0.025	0
	$(2,3)$	0	0	0.033	0	0.033	0.9	0	0	0.033
	$(3,1)$	0	0	0	0.05	0	0	0.9	0.05	0
	$(3,2)$	0	0	0	0	0.033	0	0.033	0.9	0.033
	$(3,3)$	0	0	0	0	0	0.05	0	0.05	0.9

Table 3.5: Simulation result with $\tau_u = 0.7$, sample size 1000

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.199	0.283	1.017	-0.435	0.874
RMS	0.054	0.057	0.216	0.143	0.281
SE	0.055	0.057	0.222	0.042	0.056
SD	0.054	0.055	0.216	0.045	0.059
R	0.988	0.955	0.973	1.080	1.052

Table 3.6: Simulation result with $\tau_u = 0.8$, sample size 1000

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.198	0.291	1.026	-0.373	0.746
RMS	0.051	0.054	0.212	0.087	0.155
SE	0.053	0.054	0.213	0.045	0.050
SD	0.051	0.053	0.210	0.046	0.052
R	0.966	0.988	0.989	1.024	1.033

Table 3.7: Simulation result with $\tau_u = 0.9$, samplc size 1000

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.200	0.299	1.030	-0.331	0.661
RMS	0.049	0.051	0.193	0.053	0.074
SE	0.050	0.050	0.193	0.043	0.042
SD	0.049	0.051	0.191	0.043	0.042
R	0.971	1.006	0.986	1.002	1.016

Table 3.8: Simulation result with $\tau_u = 0.7$, sample size 500

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.200	0.278	1.034	-0.429	0.866
RMS	0.080	0.081	0.316	0.142	0.278
SE	0.077	0.082	0.320	0.054	0.074
SD	0.080	0.078	0.315	0.060	0.080
R	1.030	0.957	0.983	1.094	1.085

Table 3.9: Simulation result with $\tau_u = 0.8$, sample size 500

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.197	0.287	1.040	-0.371	0.742
RMS	0.074	0.074	0.301	0.096	0.159
SE	0.075	0.077	0.313	0.060	0.069
SD	0.074	0.072	0.298	0.064	0.071
R	0.981	0.940	0.952	1.065	1.030

Table 3.10: Simulation result with $\tau_u = 0.9$, sample size 500

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.199	0.296	1.043	-0.330	0.659
RMS	0.071	0.067	0.274	0.067	0.083
SE	0.072	0.072	0.284	0.059	0.058
SD	0.071	0.067	0.271	0.060	0.059
R	0.987	0.937	0.952	1.015	1.017

Table 3.11: Simulation result with $\tau_u = 0.7$, sample size 300

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.203	0.274	1.025	-0.424	0.861
RMS	0.100	0.109	0.381	0.143	0.277
SE	0.100	0.106	0.415	0.065	0.089
SD	0.100	0.106	0.380	0.072	0.094
R	0.999	1.001	0.917	1.098	1.061

Table 3.12: Simulation result with $\tau_u = 0.8$, sample size 300

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.198	0.284	1.052	-0.373	0.744
RMS	0.098	0.100	0.391	0.108	0.170
SE	0.098	0.100	0.418	0.072	0.086
SD	0.098	0.099	0.388	0.080	0.090
R	1.001	0.989	0.928	1.102	1.046

Table 3.13: Simulation result with $\tau_u = 0.9$, sample size 300

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.200	0.290	1.046	-0.332	0.660
RMS	0.094	0.090	0.341	0.083	0.097
SE	0.094	0.093	0.376	0.073	0.072
SD	0.094	0.090	0.338	0.077	0.076
R	0.998	0.965	0.900	1.055	1.056

From the results presented in the Tables 3.5-3.13, we see that the mean values of the estimates are very close to the true values in all situations, the root mean squared errors and the mean of standard errors are reasonably small. The values of the ratio (R) are all between 0.8 and 1.2. Therefore, the simulation results indicate that the proposed estimate procedure in general can provide reliable parameter estimates and standard errors.

Table 3.14: Non-inferiority CI for different τ_u values

τ_u	NOBS	ECP
0.7	1000	0.959
0.8	1000	0.958
0.9	1000	0.962
0.7	500	0.957
0.8	500	0.962
0.9	500	0.967
0.7	300	0.955
0.8	300	0.960
0.9	300	0.960

Moreover, in Table 3.14, we can see that the ECPs of the CIs are close to 0.95 in all cases, suggesting that the use of the CI to examine non-inferiority will produce reliable results.

We also summarized the simulation results with respect to the different sample sizes (NOBS) and τ_u values in the following graphs, where MIS7, MIS8, and MIS9 stand for the results with $\tau_u = 0.7$, $\tau_u = 0.8$, and $\tau_u = 0.9$ respectively. The five parameters estimates in $\hat{\boldsymbol{\theta}} = (\hat{\mu}_T, \hat{\sigma}_{RT}, \hat{\sigma}_T^2, \hat{\beta}_2, (\hat{\beta}_3 - \hat{\beta}_2))^T$ are labeled from Parameter1, Parameter2 to Parameter5. We can see that for all the unknown parameters, the accuracy of the estimates will be improved if the sample size increases or if τ_u gets bigger, as reflected by the fact that RMS becomes smaller and R gets closer to 1.

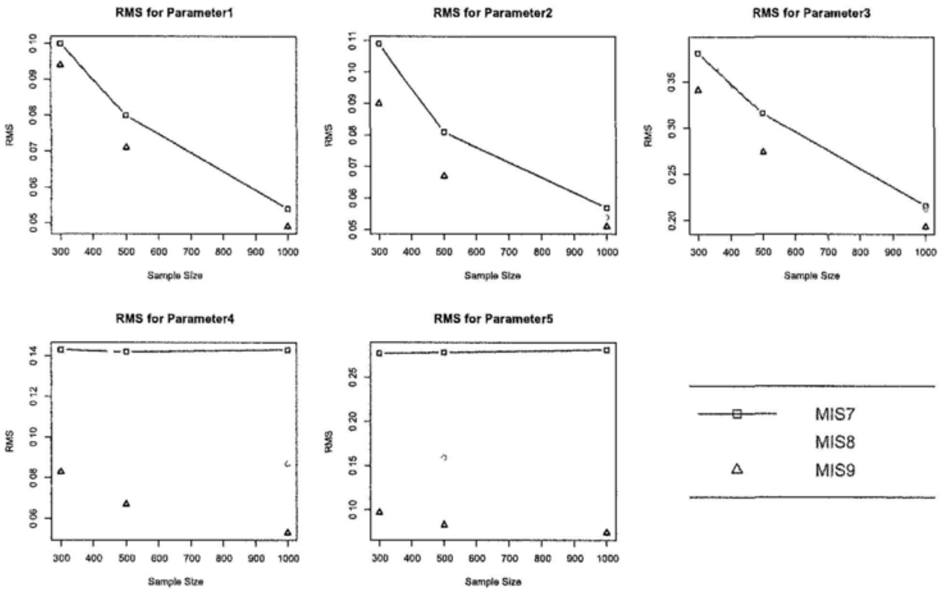

Figure 3 1: RMS known misclassification probabilities

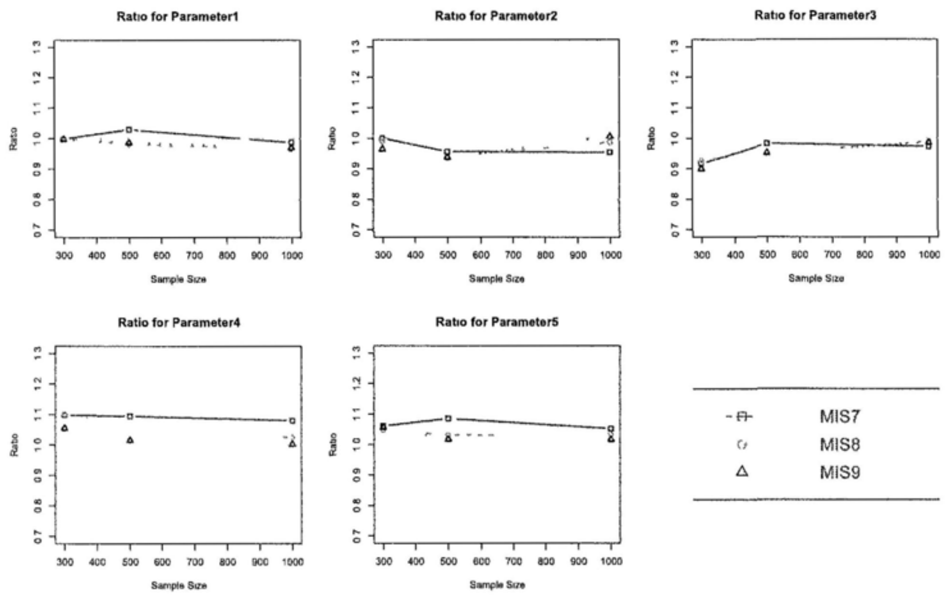

Figure 3.2: SD/SE Ratio, known misclassification probabilities

3.2.3 Example 4

We used the data set in Table 2.6 from Stuart (1955) again, and combined the sparse "lowest" group and the "third" group to obtain a contingency table as presented in Table 3.15. We assume that the honesty probability is known and is $\tau_{\boldsymbol{u},j} = \tau_{\boldsymbol{u}} = 0.8$, and the transition misclassification probability matrix is given in Table 3.3.

Table 3.15: Right and left eye unaided distance vision for women

Right eye \ Left eye	High(1)	Medium(2)	Low(3)	Total
High(1)	1520	266	190	1976
Medium(2)	234	1512	510	2256
Low(3)	153	444	2668	3245
Total	1907	2222	3348	7477

Using the right eye as the reference and the left eye as the treatment, we obtained $\hat{\mu}_T = 0.034$, suggesting that the distance vision of the left eye is lower than that of the right eye. The lower bound of non-inferiority test CI is 0.014, which is larger than -0.2 at 0.05 level of significance. Therefore, we can conclude non-inferiority. That is to say, when the honesty probability is known as 0.8, we can achieve a conclusion that is the same as in the case of no misclassification. (See Section 2.6.1)

3.3 Model for Partially Validated Data

When the misclassification probabilities are not known, we can use partially validated data to seek information in relation to misclassification. For partially validated data, it is assumed that two devices are available to classify the participating subjects. One device is called the true classifier. It can classify participants accurately but usually with higher cost. The other device, called the fallible classifier, is more likely to lead to misclassification but it is usually less expensive. The responses of the participants who are classified by the true classifier reflect the true situation and other responses obtained by the fallible classifier may not reflect the true state of the respondents.

Let $\boldsymbol{k}$ represent the cell classified by the fallible device, and $\boldsymbol{u}$ the one classified by true measurement device. The total sample size of a data set that is obtained from partially validated data is N. Out of which n randomly drawn units are classified by both devices and $n_{\boldsymbol{k}(\boldsymbol{u})}$ is the total number of units for which the true and fallible devices result to the classification to cells $\boldsymbol{u}$ and $\boldsymbol{k}$, respectively. The remaining $n^* = N - n$ units are classified by the fallible device only, and $n^*_{\boldsymbol{k}}$ is the total number of units that have been classified into cell $\boldsymbol{k}$. The data structure is summarized in Table 3.16.

Table 3.16: Paired design structure with partially validated data

Fallible k / True u	(1,1)	(1,2)		(1,K)	(2,1)		(2 K)		(K,K)	Total
(1 1)	$n_{11(11)}$	$n_{12(11)}$		$n_{1K(11)}$	$n_{21(11)}$		$n_{2K(11)}$		$n_{KK(11)}$	$n_{+(11)}$
(1,2)	$n_{11(12)}$	$n_{12(12)}$		$n_{1K(12)}$	$n_{21(12)}$		$n_{2K(12)}$		$n_{KK(12)}$	$n_{+(12)}$
(1,K)	$n_{11(1K)}$	$n_{12(1K)}$		$n_{1K(1K)}$	$n_{21(1K)}$		$n_{2K(1K)}$		$n_{KK(1K)}$	$n_{+(1K)}$
(2,1)	$n_{11(21)}$	$n_{12(21)}$		$n_{1K(21)}$	$n_{21(21)}$		$n_{2K(21)}$		$n_{KK(21)}$	$n_{+(21)}$
(2,K)	$n_{11(2K)}$	$n_{12(2K)}$		$n_{1K(2K)}$	$n_{21(2K)}$		$n_{2K(2K)}$		$n_{KK(2K)}$	$n_{+(2K)}$
(K,K)	$n_{11(KK)}$	$n_{12(KK)}$		$n_{1K(KK)}$	$n_{21(KK)}$		$n_{2K(KK)}$		$n_{KK(KK)}$	$n_{+(KK)}$
Total	$n_{11(+)}$	$n_{12(+)}$		$n_{1K(+)}$	$n_{21(+)}$		$n_{2K(+)}$		$n_{KK(+)}$	n
Fallible only	n^*_{11}	n^*_{12}		n^*_{1K}	n^*_{21}		n^*_{2K}		n^*_{KK}	n^*
Grand total	$N_{11} = n^*_{11} + n_{11(+)}$	$N_{12} = n^*_{12} + n_{12(+)}$		$N_{1K} = n^*_{1K} + n_{1K(+)}$	$N_{21} = n^*_{21} + n_{21(+)}$		$N_{2K} = n^*_{2K} + n_{2K(+)}$		$N_{KK} = n^*_{KK} + n_{KK(+)}$	$N = n + n^*$

Let $p_{\boldsymbol{u}}$ be the probability that a unit actually belongs to cell $\boldsymbol{u}$; $\pi_{\boldsymbol{k}}$ be the probability that a unit is classified into cell $\boldsymbol{k}$ by the fallible device; and $\omega_{\boldsymbol{k}(\boldsymbol{u})}$ be the probability that one is classified into cell $\boldsymbol{k}$ when it actually falls in cell $\boldsymbol{u}$. Using the method suggested in Yiu & Poon (2008) and Zhang (2007), it can be shown that the ML estimates $\hat{p}_{\boldsymbol{u}}$ of $p_{\boldsymbol{u}}$ and $\hat{\omega}_{\boldsymbol{k}(\boldsymbol{u})}$ of $\omega_{\boldsymbol{k}(\boldsymbol{u})}$ are given as follows:

$$\hat{p}_{\boldsymbol{u}} = \sum_{\boldsymbol{k}} \frac{(n^*_{\boldsymbol{k}} + n_{\boldsymbol{k}(+)})n_{\boldsymbol{k}(\boldsymbol{u})}}{Nn_{\boldsymbol{k}(+)}}, \tag{3.3}$$

$$\hat{\omega}_{\boldsymbol{k}(\boldsymbol{u})} = \frac{(n^*_{\boldsymbol{k}} + n_{\boldsymbol{k}(+)})n_{\boldsymbol{k}(\boldsymbol{u})}}{Nn_{\boldsymbol{k}(+)}\hat{p}_{\boldsymbol{u}}}. \tag{3.4}$$

Let p^* be a $(K \times K) \times 1$ vector that stores $p_{\boldsymbol{u}}$'s for $u(1) = 1, \ldots, K; u(2) = 1, \ldots, K$, and $\hat{p}^*$ be its ML estimate with elements that are given by (3.3). Let $\boldsymbol{\theta} = (\mu_T, \sigma_{RT}, \sigma_T^2, \boldsymbol{\beta})^T$ be the unknown parameters vector as before. Then $p^* = p^*(\boldsymbol{\theta})$ and the estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained by the method of Modified Minimum Chi-square (MMC) in which $\hat{\boldsymbol{\theta}}$ minimizes the function

$$G(\boldsymbol{\theta}) = (\hat{p}^* - p^*(\boldsymbol{\theta}))^T \hat{\boldsymbol{\Omega}}^{-1} (\hat{p}^* - p^*(\boldsymbol{\theta})), \tag{3.5}$$

where $\boldsymbol{\Omega}$ is a $(K \times K)$ by $(K \times K)$ matrix. The diagonal elements of $\boldsymbol{\Omega}$ that correspond to $\hat{p}_{\boldsymbol{u}}$ are given by

$$\frac{p_{\boldsymbol{u}}}{n} + p_{\boldsymbol{u}}^2 (\frac{1}{N} - \frac{1}{n}) \sum_{\boldsymbol{k}} \frac{\omega^2_{\boldsymbol{k}(\boldsymbol{u})}}{\pi_{\boldsymbol{k}}},$$

and the off-diagonal elements that correspond to $\hat{p}_{\boldsymbol{u}}$ and $\hat{p}_{\boldsymbol{u}'}$ for $\boldsymbol{u} \neq \boldsymbol{u}'$ are given by

$$\frac{p_u p_{u'}}{N} \sum_{k} \frac{\omega_{k(u)} \omega_{k(u')}}{\pi_k}. \tag{3.6}$$

The matrix $\mathbf{\Omega}$ can be derived analytically and $\hat{\mathbf{\Omega}}$ is obtained by replacing the unknown parameters in $\mathbf{\Omega}$ with their consistent estimates. We used Mx to compile a program to find the estimates of the parameters by minimizing (3.5).

3.3.1 Simulation

In the simulation, we set K equal to 3, and set the true parameter values $(\mu_T, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2))^T = (0.2, 0.3, 1, -0.3, 0.6)^T$. The two thresholds are set to be identical as $\beta_2 = -0.3$ and $\beta_3 = 0.3$ for both dimensions.

To generate the data with misclassification, we make use of the misclassification matrices in Tables 3.2, 3.3 and 3.4 that were compiled based on different τ_u values and the assumption that misclassification only arises between adjacent cells with equal probability.

The simulation results are summarized by using both tables and graphs. To make it clear in the graph, we use NOBS and NRES to represent N and n in the presentation. We simulated three different combinations of data sets with sizes $(N, n) = (NOBS, NRES) = (1000, 600), (500, 300)$ and $(500, 200)$.

The simulation study was completed by using the following steps:

1. Simulate N observations with $\boldsymbol{\theta}$ known to get a set of ordinal categorical data and a contingency table without misclassification.

2. Use the misclassification matrix to generate misclassified data so as to get a set of ordinal categorical data with misclassification.

3. For n randomly drawn observations out of the N observations, the classification results based on steps 1 and 2 were used. For the remaining $N-n$ observations, only the classification results based on step 2 were used.

Based on the data set, we can find the parameter estimates by the Modified Minimum Chi-square (MMC) method (See (3.5)). We repeated steps 1 to 3 for 1000 times, generating 1000 replications for the simulation. We can then get the mean of estimates based on the 1000 sets, and we can also compute RMS and the standard deviation (SD) based on the 1000 estimates.

To assess the confidence interval that is obtained by using the bootstrap standard errors, we used the following steps 4 and 5.

4. For the data set generated in each simulation replication, we draw samples of size n from the n observations 100 times with replacement, and draw 100 samples of size $N-n$ from the remaining $N-n$ observations with replacement.

5. For each of 100 bootstrap samples, find the estimates of the parameters, so we have 100 sets of estimates for the unknown parameters. For each parameter, the standard error is obtained as the standard deviation based on the 100 estimates.

The simulation results are presented in Tables 3.17 to 3.25 as follows.

Table 3.17: Simulation result for NOBS=1000 NRES=600 $\tau_u = 0.7$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.202	0.303	1.021	-0.299	0.595
RMS	0.053	0.056	0.216	0.046	0.043
SE	0.055	0.055	0.215	0.045	0.041
SD	0.053	0.056	0.215	0.046	0.043
R	0.970	1.030	1.000	1.011	1.049

Table 3.18: Simulation result for NOBS=1000 NRES=600 $\tau_u = 0.8$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.202	0.302	1.016	-0.299	0.595
RMS	0.052	0.054	0.202	0.045	0.042
SE	0.053	0.053	0.204	0.044	0.040
SD	0.052	0.054	0.202	0.045	0.042
R	0.981	1.016	0.989	1.023	1.049

Table 3.19: Simulation result for NOBS=1000 NRES=600 $\tau_u = 0.9$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.202	0.303	1.016	-0.299	0.596
RMS	0.048	0.052	0.189	0.043	0.039
SE	0.051	0.050	0.192	0.042	0.038
SD	0.048	0.052	0.188	0.043	0.039
R	0.945	1.032	0.982	1.014	1.037

Table 3.20: Simulation result for NOBS=500 NRES=300 $\tau_u = 0.7$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.205	0.307	1.060	-0.301	0.598
RMS	0.077	0.079	0.323	0.063	0.058
SE	0.082	0.083	0.354	0.064	0.056
SD	0.077	0.079	0.318	0.063	0.058
R	0.937	0.949	0.896	0.992	1.046

Table 3.21: Simulation result for NOBS=500 NRES=300 $\tau_u = 0.8$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.205	0.307	1.056	-0.301	0.598
RMS	0.074	0.076	0.311	0.061	0.056
SE	0.079	0.079	0.329	0.062	0.054
SD	0.074	0.075	0.306	0.061	0.056
R	0.938	0.951	0.929	0.983	1.044

Table 3.22: Simulation result for NOBS=500 NRES=300 $\tau_u = 0.9$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.204	0.306	1.048	-0.301	0.598
RMS	0.072	0.072	0.289	0.059	0.052
SE	0.075	0.074	0.299	0.060	0.051
SD	0.071	0.072	0.285	0.059	0.052
R	0.952	0.967	0.953	0.981	1.025

Table 3.23: Simulation result for NOBS=500 NRES=200 $\tau_u = 0.7$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.204	0.308	1.064	-0.297	0.593
RMS	0.085	0.092	0.378	0.069	0.067
SE	0.094	0.101	0.488	0.070	0.062
SD	0.085	0.092	0.373	0.068	0.066
R	0.898	0.915	0.764	0.974	1.077

Table 3.24: Simulation result for NOBS=500 NRES=200 $\tau_u = 0.8$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.204	0.307	1.060	-0.298	0.594
RMS	0.080	0.086	0.353	0.065	0.062
SE	0.087	0.092	0.425	0.067	0.058
SD	0.080	0.086	0.348	0.064	0.062
R	0.917	0.935	0.818	0.965	1.060

Table 3.25: Simulation result for NOBS=500 NRES=200 $\tau_u = 0.9$

	$\hat{\mu}_T$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$
True value	0.2	0.3	1	-0.3	0.6
Mean	0.201	0.308	1.053	-0.300	0.597
RMS	0.075	0.079	0.321	0.060	0.057
SE	0.079	0.081	0.339	0.062	0.054
SD	0.075	0.079	0.317	0.060	0.057
R	0.948	0.972	0.935	0.967	1.050

Table 3.26: Non-inferiority CI for different τ_u values

τ_u	NOBS	NRES	ECP
0.7	1000	600	0.958
0.8	1000	600	0.959
0.9	1000	600	0.965
0.7	500	300	0.970
0.8	500	300	0.965
0.9	500	300	0.965
0.7	500	200	0.976
0.8	500	200	0.979
0.9	500	200	0.962

From Tables 3.17 to 3.25, we see that the mean values of the estimates are very close to the true values in all situations, and the RMS and the mean of standard errors are reasonably small. The ratios are all between 0.8 and 1.2. Therefore, the simulation results indicate that the proposed estimate procedure in general can provide the users with reliable parameter estimates and standard errors. We can also see from Table 3.26 the ECPs of the CIs are close to 0.95 in all cases, suggesting that the use of the CI to examine non-inferiority are reliable but a little conservative.

We also summarized the simulation results with respect to the different NOBS, NRES and τ_u values in the following Figures 3.3 to 3.6, where MIS7, MIS8, and MIS9 stand for $\tau_u = 0.7$, $\tau_u = 0.8$, and $\tau_u = 0.9$ respectively. We can see from the figures that for all of the unknown parameters, the accuracy of the estimates will be improved if NOBS, NRES or the τ_u value increases, as evident by the fact that the RMS gets smaller and the ratio $R = SD/SE$ gets closer to 1.

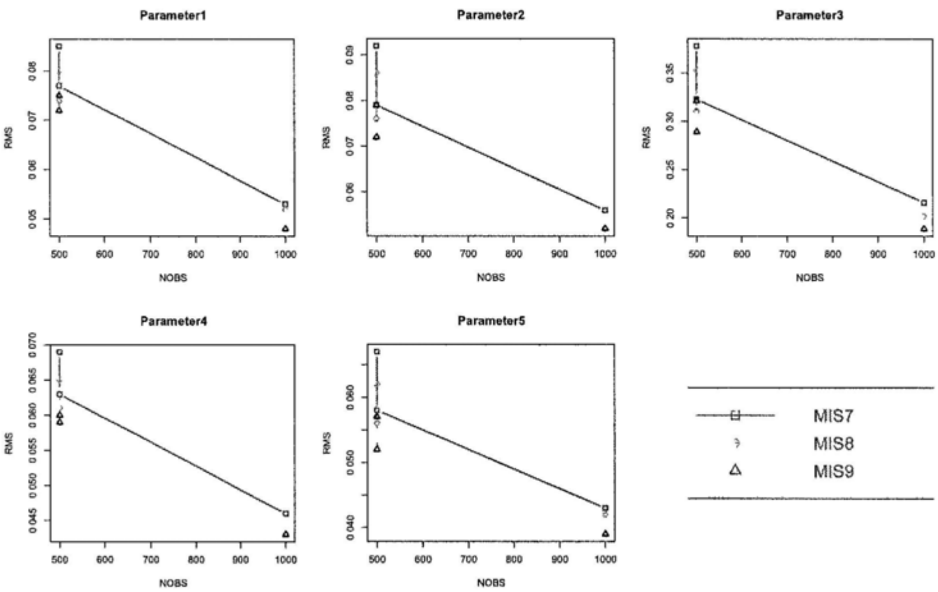

Figure 3.3: RMS NOBS paired design with partially validated data

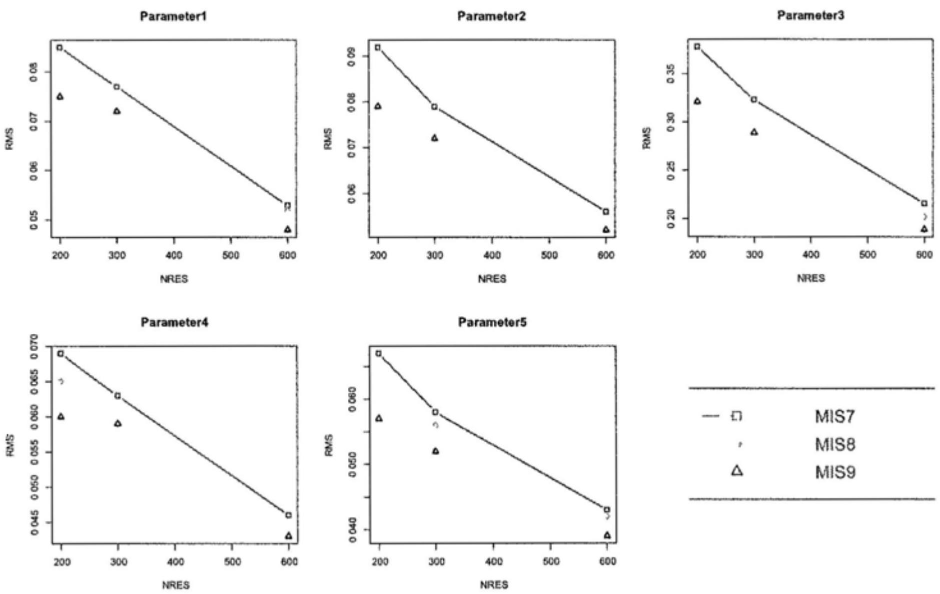

Figure 3.4: RMS NRES paired design with partially validated data

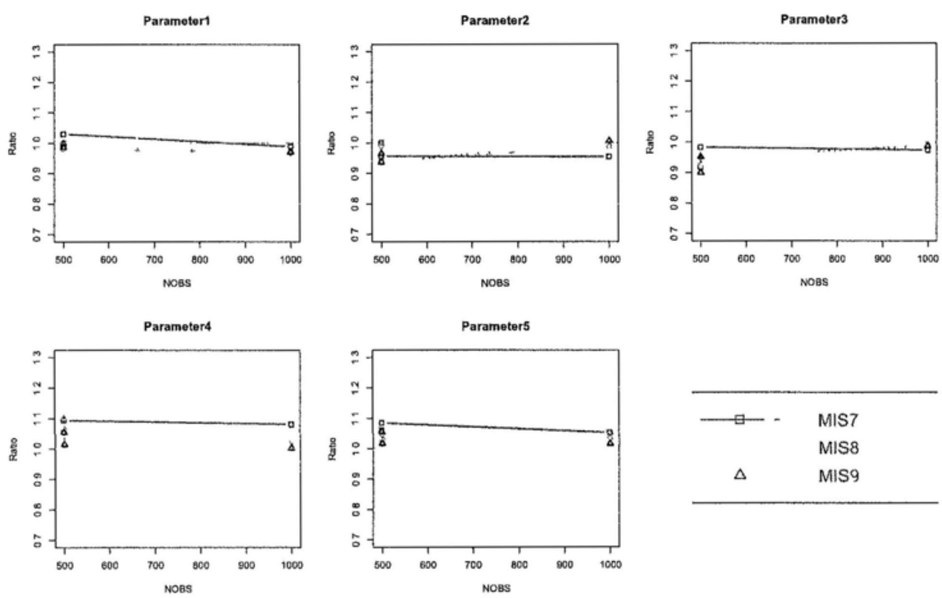

Figure 3.5: Ratio NOBS paired design with partially validated data

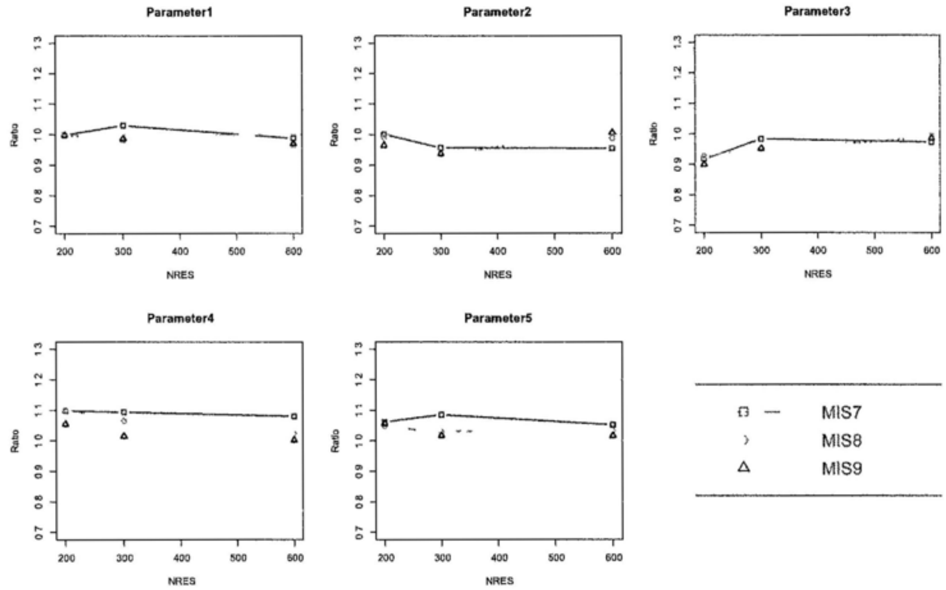

Figure 3 6: SD/SE Ratio NRES paired design with partially validated data

3.3.2 The data set on How Effective the Learning Paths (HELP) are

We analyze a data set on "How Effective the Learning Paths (HELP) are" to illustrate how the proposed partially validated data method can be used in a flexible manner to facilitate a more comprehensive and informative data analysis. The study investigated the effectiveness of four different learning approaches of university students, and the data is obtained through questionnaires.

In addition to attending lectures, university students use different learning approaches to gain a better understanding of key concepts. Four approaches are commonly used, they are (1) discussing with lecturers, (2) discussing with teaching assistants, (3) discussing with peer students, or (4) conducting independent self-study. To compare the effectiveness of the four learning approaches, data were collected from university students. Students were asked to express their opinions on the four approaches by means of rating the degree of agreement on the effectiveness for the four different approaches.

However, it is natural that not all students have used all the four methods in their university study, and hence some can only respond to the questions using their perceptions. For the purpose of evaluating the effectiveness of an approach, students' responses that are based on general perception can be regarded as being obtained from the fallible classifier in the partially validated data. The cost of obtaining perception data is usually lower. On the other hand, students who are frequent users of a specific learning approach can provide more reliable assessment in terms of the effectiveness of the approach, especially if responses were made with reference to recent experiences. Responses such obtained can be regarded as responses being obtained from the true classifier.

Therefore, to achieve a better comparison among the four different learning approaches, data were collected from 871 students for the following three groups of questions. Each group has 4 questions corresponding to the four different approaches.

a) Communication frequencies

In the past 6 months, how often have you used the following learning paths to make better understanding of a concept out of the time you spend in lecture?

a1) Discussing with course lecturers

a2) Discussing with teaching assistants

a3) Discussing with peers

a4) Independent learning

b) To solve difficulties in understanding a concept after class, how much do you agree or disagree with the following statements?

b1) Discussing with course lecturers can make confusing concepts clear

b2) Discussing with teaching assistants can make confusing concepts clear

b3) Peer discussions can make confusing concepts clear

b4) Independent learning can make confusing concepts clear

c) Considering the last time that you conducted the following learning approaches to make better understanding of a concept after class, how much do you agree or disagree with the following statements?

c1) Discussing with course lecturers made confusing concepts clear last time

c2) Discussing with teaching assistants made confusing concepts clear last time

c3) Peer discussions made confusing concepts clear last time

c4) Independent learning made confusing concepts clear last time

For the questions in Group b and Group c, a five-point scale was adopted with "1" representing "highly disagree" and "5" representing "highly agree". In other words, a response of "1" means that the student found the method not helpful at all and "5" very helpful.

There are five response categories for the questions in Group a. They are (i) Never, (ii) 1 to 3 times, (iii) 4 to 6 times, (iv) 7 to 9 times, and (v) more than 9 times. Students whose responses fall in the category "never" will not have responses available for the questions in Group c. In this case, only fallible perception type responses to questions in Group b were available for analysis.

Such method of data collection with the Group a questions in HELP allows us to clearly distinguish the types of fallible perception and more reliable experience data. We will illustrate in the next subsection on how the proposed method for analyzing partially validated data can be used to make use of all the available data to achieve a comparison of the learning approaches.

In this chapter, we will analyze two of the four learning paths by using the paired data model suggested in this chapter. The analysis of more variables will be discussed in Chapter 5.

3.3.3 Example 5

In this section, we present the analysis that was based on the data in relation to "discussing with lecturers"(X_P), and "independent learning"(X_T).

The data is presented in Table 3.27. The objective is to examine whether independent learning is non-inferior to discussing with lecturers. There are 871 (NOBS) effective questionnaires after data cleaning. By using the questions in Group a as the discriminator, we have 679 (NRES) respondents that have been classified by both the true device and the fallible device. As a minor modification of the original data, we collapsed the five-category data into three-category by combining the first three categories to reduce the number of sparse cells.

Table 3.27: Data structure of example 5 in 3.3.3

True \ Fallible	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	Total
(1,1)	35	14	4	11	3	1	4	0	1	73
(1,2)	12	32	2	5	14	2	4	4	0	75
(1,3)	3	6	29	0	5	6	1	0	2	52
(2,1)	12	4	3	34	15	0	8	3	2	81
(2,2)	9	16	3	19	56	8	6	11	2	130
(2,3)	0	2	8	3	10	15	2	6	3	49
(3,1)	2	0	0	9	2	0	34	11	1	59
(3,2)	0	1	0	12	8	3	14	32	8	78
(3,3)	0	1	2	1	10	6	1	14	47	82
Total	73	76	51	94	123	41	74	81	66	679
Fallible only	33	33	19	31	30	7	12	12	15	192
Grand total	106	109	70	125	153	48	86	93	81	871

Fallible device: Group b (lecturer, independent learning)

True device: Group c (lecturer, independent learning)

Based on the proposed method, the estimation of the parameters are given by $\hat{\mu}_T = -0.074$, $\hat{\sigma}_{RT} = 0.1376$, $\hat{\sigma}_T = 0.8088$, $\hat{\beta}_2 = -0.4808$, and $\hat{\beta}_3 - \hat{\beta}_2 = 0.982$, and the corresponding standard errors are 0.050, 0.050, 0.097, 0.037 and 0.038. As $\hat{\mu}_T = -0.074$ is smaller than $\hat{\mu}_R = 0$, it provides evidence that independent learning is less effective than discussing with lecturers.

To test the non-inferiority of the two paths, we use the margin $\Delta = 0.2$. The 95% CI for μ_T is $(-0.156, +\infty)$. The lower bound -0.156 is greater than -0.2. Therefore, we can conclude that independent learning is non-inferior to discussing with lecturers. That is, independent learning approach could be promoted as an extension and enhancement of the proper lecture, since students themselves can crystallize the key concepts as effective as seeking help from lecturers after class.

Chapter 4

Testing Non-inferiority in Block Design

4.1 Introduction

The applications of a non-inferiority trial design that does not contain a placebo arm may be controversial, and the results from such a trial design can be rather difficult to interpret with confidence. To address this issue, one can adopt a three-armed trial design (see Tang & Tang (2004); Pigeot et al. (2003)). In three-armed trials, investigators use both a known effective standard treatment and a placebo as control. This design can clearly distinguish between a treatment that does not work (*i.e.*, the reference treatment is superior to placebo but the new treatment is not) and a study that does not work (*i.e.*, neither the reference nor the new treatment is superior to the placebo), as stated in Tang & Tang (2004).

Koch & Tangen (1999) and Pigeot et al. (2003) studied the problem of non-inferiority testing in a three-armed clinical trial including a placebo. Their methods are suitable to employ when the outcome responses from the placebo, active control, and the new experimental treatments are all independently normally distributed with the same variance. Tang & Tang (2004) and Munk et al. (2007) considered assessment of non-inferiority with binary data. However, little has been achieved in the literature for non-inferiority analysis of ordinal categorical data, and in particular when the responses of the placebo, active control and new treatments are obtained via a block-design. In this section, we develop a latent variable approach to study non-inferiority in a three-armed design.

4.2 Model and Data

Let $(X_P, X_R, X_T)^T$ be the multivariate normally distributed variables of interest with the subscripts P, R, and T stand for the placebo, reference and new treatment groups, respectively. Moreover, let the mean vector of $(X_P, X_R, X_T)^T$ be

$$\boldsymbol{\mu} = (\mu_P, \mu_R, \mu_T)^T,$$

and the covariance matrix be

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_P^2 & \sigma_{PR} & \sigma_{PT} \\ \sigma_{PR} & \sigma_R^2 & \sigma_{RT} \\ \sigma_{PT} & \sigma_{RT} & \sigma_T^2 \end{pmatrix}.$$

We operate on the assumption that X_P, X_R, and X_T cannot be observed, and these underlying continuous variables X_P, X_R and X_T are related to the observed ordinal categorical variables Z_P, Z_R and Z_T. Specifically, the relationship is given by

$$Z_i = k \quad \text{if } \beta_k \leq X_i < \beta_{k+1}, \tag{4.1}$$

for $i = P, R, T$; $k = 1, \ldots, K$, where $\beta_1 = -\infty$, $\beta_{K+1} = \infty$ and $\boldsymbol{\beta} = (\beta_2, \ldots, \beta_K)^T$ is a vector of unknown thresholds. Here, to enable straightforward comparison between μ_T and μ_R, we assume that the thresholds are invariant for X_P, X_R and X_T.

As every observed data point of Z_P, Z_R, and Z_T must fall in a cell of a $K \times K \times K$ contingency table, an ordered contingency table can be easily generated as follows:

Table 4.1: Block design: Response of Placebo Group for $Z_P = m$ ($m = 1, \cdots, K$)

Z_R \ Z_T	1	2	3	$\cdots$	K	Total
1	n_{m11}	n_{m12}	n_{m13}	$\cdots$	n_{m1K}	n_{m1+}
2	n_{m21}	n_{m22}	n_{m23}	$\cdots$	n_{m2K}	n_{m2+}
3	n_{m31}	n_{m32}	n_{m33}	$\cdots$	n_{m3K}	n_{m3+}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
K	n_{mK1}	n_{mK2}	n_{mK3}	$\cdots$	n_{mKK}	n_{mK+}
Total	n_{m+1}	n_{m+2}	n_{m+3}	$\ddots$	n_{m+K}	n_{m++}

Our objective is to test whether the treatment is non-inferior to the reference group, with the effectiveness of the reference group relative to that of the placebo group as a benchmark for internal validation.

Let n_{mij} (where $\sum_{m=1}^{K}\sum_{i=1}^{K}\sum_{j=1}^{K} n_{mij} = n$) be the observed number of respondents that falls in the mij-th cell with the corresponding cell probability p_{mij}, so the data set consists of n three dimensional measurements that are all expressed in an ordinal scale.

We use the frequencies in the $K \times K \times K$ contingency table to first estimate the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and then conduct non-inferiority test.

Since the location and dispersion of the underlying continuous variables are not defined, we fix the population mean and the variance of the placebo group. That is, we fix $\mu_P = 0$ and $\sigma_P^2 = 1$, and we have

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ \mu_R \\ \mu_T \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma_{PR} & \sigma_{PT} \\ \sigma_{PR} & \sigma_R^2 & \sigma_{RT} \\ \sigma_{PT} & \sigma_{RT} & \sigma_T^2 \end{pmatrix}. \quad (4.2)$$

The means and variances of the other two underlying variables are then compared in a relative sense. Let $\boldsymbol{\theta} = (\mu_R, \mu_T, \sigma_{PR}, \sigma_R^2, \sigma_{PT}, \sigma_{RT}, \sigma_T^2, \boldsymbol{\beta})^T$ be the unknown parameter vector of the model. The parameters μ_R, μ_T, σ_R^2, and σ_T^2 are treated as the means and variances of X_R and X_T respectively, which can be interpreted in a relative sense to those of X_P.

Let p_{mij} be the probability of an observation that falls in the (m, i, j)th cell, we have

$$\begin{aligned} p_{mij} &= \Pr(Z_P = m, Z_R = i, Z_T = j) \\ &= \Phi_3(\beta_{m+1}, \beta_{i+1}, \beta_{j+1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) - \Phi_3(\beta_m, \beta_{i+1}, \beta_{j+1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\quad -\Phi_3(\beta_{m+1}, \beta_i, \beta_{j+1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) - \Phi_3(\beta_{m+1}, \beta_{i+1}, \beta_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\quad +\Phi_3(\beta_m, \beta_i, \beta_{j+1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \Phi_3(\beta_{m+1}, \beta_i, \beta_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\quad +\Phi_3(\beta_m, \beta_{i+1}, \beta_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}) - \Phi_3(\beta_m, \beta_i, \beta_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (4.3) \end{aligned}$$

where $\Phi_3(\beta_m, \beta_i, \beta_j; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the three-dimensional multivariate distribution function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, evaluating at β_m, β_i and β_j, respectively. The maximum likelihood estimate of $\boldsymbol{\theta}$ can be obtained by maximizing the log likelihood function

$$\sum_{m=1}^{K}\sum_{i=1}^{K}\sum_{j=1}^{K} n_{mij} \log p_{mij}(\boldsymbol{\theta}). \tag{4.4}$$

4.3 Non-inferiority Test

Once the MLE $\hat{\mu}_T$ of μ_T is available, non-inferiority can be examined by testing hypothesis on the difference of the means of the various treatments. Let us first recap the testing problem in terms of the difference of the population means. Similar to the discussion in the previous chapter, let the positive quantity Δ stand for the minimum difference to be detected, the hypothesis of non-inferiority is

$$\begin{aligned} H_0 &: \mu_T - \mu_R \leq -\Delta \\ H_1 &: \mu_T - \mu_R > -\Delta \end{aligned} \tag{4.5}$$

and non-inferiority can be established if H_0 is rejected. Following Tang & Tang (2004), in three-armed design, $-\Delta$ can be chosen as a fraction f of the difference of the population means μ_R and μ_P, i.e., $-\Delta = f(\mu_R - \mu_P)$. Thus, the non-inferiority test becomes

$$
\begin{aligned}
H_0 &: \mu_T - \mu_R \leq f(\mu_R - \mu_P) \\
H_1 &: \mu_T - \mu_R > f(\mu_R - \mu_P).
\end{aligned} \tag{4.6}
$$

To choose an appropriate value of f, the paper of Committee for Proprietary Medicinal Product (CPMP) and Tang & Tang (2004) suggested that $f = -\frac{1}{2}$ or $f = -\frac{1}{3}$ are acceptable. They represent a tolerable amount of reduction in efficacy. Besides, the commonly shared level for clinically unimportant is 20 percent, which is equivalent to $f = -\frac{1}{5}$.

Assuming that $\mu_R - \mu_P > 0$ and let $\pi = 1 + f$, then we will have the non-inferiority hypothesis test as

$$
\begin{aligned}
H_0 &: \tfrac{\mu_T - \mu_P}{\mu_R - \mu_P} \leq \pi \\
H_1 &: \tfrac{\mu_T - \mu_P}{\mu_R - \mu_P} > \pi.
\end{aligned} \tag{4.7}
$$

As a result, the non-inferiority test can be expressed by

$$
\begin{aligned}
H_0 &: \mu_T - \pi\mu_R - (1 - \pi)\mu_P \leq 0 \\
H_1 &: \mu_T - \pi\mu_R - (1 - \pi)\mu_P > 0.
\end{aligned} \tag{4.8}
$$

In this context, we may draw conclusions based on $\psi^* = \mu_T - \pi\mu_R - (1-\pi)\mu_P$. That is, if $\hat{\psi}^* - z_\alpha SE(\hat{\psi}^*) > 0$, we can reject the null hypothesis and establish non-inferiority of the new treatment to the standard treatment.

Since we have $\mu_P = 0$, the non-inferiority test can be simplified as

$$
\begin{aligned}
&H_0 : \mu_T - \pi\mu_R \leq 0 \\
&H_1 : \mu_T - \pi\mu_R > 0.
\end{aligned}
\tag{4.9}
$$

In this case, we also have $\psi^* = \mu_T - \pi\mu_R$. That is, if $\hat{\psi}^* - z_\alpha SE(\hat{\psi}^*) > 0$, we can reject the null hypothesis and conclude non-inferiority.

4.4 Simulation

To examine the performance of the proposed method, a number of data sets that are based on a set of known parameter values were generated. We set $K = 3$ and the thresholds for the three dimensions to be identical, given by $\beta_2 = -0.2$ and $\beta_3 = 0.5$. For each simulation, a data set was generated with the true values of $\boldsymbol{\theta} = (\mu_R, \mu_T, \sigma_{PR}, \sigma_R^2, \sigma_{PT}, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2))^T$ equal to $(0.2, 0.4, 0.3, 1, 0.3, 0.3, 1, -0.2, 0.7)^T$. In non-inferiority test with three-armed design, we also pay much attention to ψ^* since it is the benchmark of decision making in the hypothesis test. Given $\boldsymbol{\theta}$ and $f = -\frac{1}{5}$, the true value of $\psi^* = 0.24$.

The number of simulation replications is 1000. We consider three different sample sizes: 1000, 500 and 300.

We still use the mean, RMS, and the Ratio $R = SD/SE$ that are compiled based on the 1000 replications to assess the performance of the proposed method. We have summarized the simulation results in the following three tables.

Table 4.2: Block design simulation result for sample size 1000

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.200	0.403	0.302	1.030	0.304	0.308	1.033	-0.201	0.703	0.243
RMS	0.044	0.049	0.045	0.156	0.044	0.059	0.160	0.039	0.038	0.045
SE	0.045	0.049	0.045	0.157	0.045	0.058	0.159	0.040	0.038	0.045
SD	0.045	0.050	0.046	0.160	0.045	0.060	0.164	0.039	0.038	0.045
R	0.989	1.013	1.015	1.018	0.989	1.035	1.030	0.980	1.013	1.00

Table 4.3: Block design simulation result for sample size 500

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.200	0.404	0.304	1.037	0.299	0.308	1.038	-0.200	0.700	0.243
RMS	0.064	0.071	0.064	0.227	0.063	0.082	0.225	0.058	0.053	0.062
SE	0.064	0.070	0.064	0.226	0.064	0.083	0.228	0.056	0.054	0.063
SD	0.064	0.072	0.064	0.227	0.063	0.082	0.225	0.058	0.053	0.062
R	1.001	1.031	1.011	1.007	0.982	0.983	0.985	1.019	0.985	0.984

Table 4.4: Block design simulation result for sample size 300

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.204	0.406	0.304	1.076	0.308	0.314	1.065	-0.197	0.700	0.242
RMS	0.081	0.088	0.084	0.297	0.083	0.111	0.301	0.071	0.071	0.080
SE	0.084	0.090	0.084	0.309	0.085	0.112	0.307	0.073	0.069	0.083
SD	0.082	0.089	0.086	0.312	0.085	0.115	0.315	0.071	0.071	0.081
R	0.976	0.981	1.021	1.010	1.000	1.026	1.026	0.973	1.027	0.975

From the results in Tables 4.2-4.4, we can see that the mean values of the estimates are very close to the true values in all situations, and the root mean squared errors and the mean of standard errors are reasonably small. The ratios (R) are all between 0.8 and 1.2. Therefore, the simulation results indicate that the proposed estimate procedure in general can provide the users with acceptable parameter estimates and standard errors.

The simulation results for ψ^* are also satisfactory, suggesting statistical inferences based on the estimate $\hat{\psi}^*$ of ψ^* will likely provide reliable results. We further explore this point in the next section.

4.5 Bootstrap Confidence Interval

Similar to the discussion in Section 2.3, a CI for ψ^*, which is given by $(\hat{\psi}^* - z_\alpha SE(\hat{\psi}^*), +\infty)$, can be used to test non-inferiority in three-armed block design, where z_α is the upper α probability point of the normal distribution function and SE is the standard error produced by the bootstrap method. If $\hat{\psi}^* - z_\alpha SE(\hat{\psi}^*) > 0$, we can conclude non-inferiority.

To assess the performance of the CIs such constructed, we need to assess the performance of the SE and the CIs, and it is accomplished by simulation studies.

As before, to evaluate the performance of CIs empirically, we use expected coverage probability (ECP), measuring the expected proportion of the times that the CIs contain the true value of the parameter.

Three sample sizes have been studied: 1000, 500, and 300. In all of the simulation studies, 1000 replications have been used.

Table 4.5: Evaluation of the Non-inferiority Test CIs

NOBS	Covering True Value Times	Expected Coverage Probability
1000	958	0.958
500	959	0.959
300	969	0.969

The CI for non-inferiority test is in general a bit conservative because the coverage probability is slightly larger than the expected level.

4.6 Example 6

This data set first appeared in Kenward & Jones (1991), and it was originally designed as the three-period three-treatment cross-over trial on patients with primary dysmenorrheal. There are 86 subjects available for the trial, which were randomly assigned to receive all of the three treatments in six possible orders ABC; ACB; BAC; BCA; CAB and CBA, where $A = Placebo$; $B = Low\ dose\ analgesic\ as\ the\ treatment\ arm$; $C = High\ dose\ an algesic\ as\ the\ reference\ arm.$ At the end of each treatment, each subject rated the amount of relief obtained in an ordinal scale: $none = 1$; $moderate = 2$; and $complete = 3$. Here, we assume that the carry-over effect can be ignorable in the study to conduct the non-inferiority test in a block design. That is, we combined outcomes (1,2,3) in ABC order and (1,3,2) in ACB order. After this modification, the data set can be used as an illustration as follows:

Table 4.6: Three-armed trial on 86 patients, $X_P = 1$

X_R \ X_T	1	2	3	Total
1	6	4	5	15
2	3	13	10	26
3	0	8	14	22
Total	9	25	29	63

Table 4.7: Three-armed trial on 86 patients, $X_P = 2$

X_R \ X_T	1	2	3	Total
1	1	3	2	6
2	2	3	1	6
3	2	1	2	5
Total	5	7	5	17

Table 4.8: Three-armed trial on 86 paticnts, $X_P = 3$

X_R \ X_T	1	2	3	Total
1	2	0	2	4
2	0	0	0	0
3	1	1	0	2
Total	3	1	2	6

We are concerned with whether low dose is non-inferior to high-dose level in this example. The MLE of the unknown parameters $\mu_R, \mu_T, \sigma_{PR}, \sigma_R^2, \sigma_{PT}, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2), \psi^*$ are 1.307, 1.105, -0.254, 0.657, -0.206, 0.215, 0.779, 0.621, 0.856, 0.060 and $SE(\hat{\psi}^*) = 0.123$. The 95% one-sided confidence interval for ψ^* is $(-0.142, +\infty)$, it shows that the null hypothesis of non-inferiority can not be rejected at $\alpha = 0.05$, since the lower bound is less than 0. In other words, "Low dose analgesic" is not non-inferior to the "High dose analgesic".

Chapter 5

Block Data with Misclassification

5.1 Introduction

In Chapter 3, we have proposed the method to test non-inferiority for the data with misclassification in a pair design. In this chapter, we will consider the test of non-inferiority in the context of three-armed design. Test procedure for situations with known misclassification probabilities and with partially validated data will be developed. The performance of the proposed method will be assessed through simulation, and the applications of the proposed method will be demonstrated by two real data examples.

5.2 Model with Known Misclassification Probabilities in Three-armed Trial

Suppose we have a $K \times K \times K$ contingency table with N observations. Let $\boldsymbol{k} = (k(1), k(2), k(3))$ and $n_{\boldsymbol{k}}$ the observed frequency in cell $\boldsymbol{k}$, where $k(1), k(2), k(3) = 1, 2, \cdots, K$. $\boldsymbol{E}_{\boldsymbol{k}}$ is used to denote the $K \times K \times K$ array with element $\boldsymbol{k}$ equals to 1 and all other elements equal to 0. $\boldsymbol{Z}_j$ is a $K \times K \times K$ array with $Z_{\boldsymbol{k},j}$ as its elements in cell $\boldsymbol{k}$ and it is used to represent the classification of the j-th subject. $\boldsymbol{X}_j$ is a $K \times K \times K$ array with $X_{\boldsymbol{k},j}$ as its elements in cell $\boldsymbol{k}$, and it is used to represent the true state of the j-th subject.

We also have

$$X_{\boldsymbol{k},j} = \begin{cases} 1 & \text{if the } j\text{-th subject truly belongs to the cell } \boldsymbol{k}, \\ 0 & \text{otherwise} \end{cases}$$

$$Z_{\boldsymbol{k},j} = \begin{cases} 1 & \text{if the } j\text{-th subject is classified into the cell } \boldsymbol{k}, \\ 0 & \text{otherwise} \end{cases}$$

$$Y_j = \begin{cases} 1 & \text{if the } j\text{-th subject tells the truth or is correctly classified,} \\ 0 & \text{otherwise} \end{cases}$$

for $j = 1, \ldots, N$.

Let $\boldsymbol{u} = (u(1), u(2), u(3))$, where $u(1), u(2), u(3) = 1, 2, \cdots, K$. Generalizing the model in Section 3.2, we define the several probabilities that are involved in the model as follows.

(a) $p_{\boldsymbol{u}}$ is the probability that a subject actually belongs to the cell $\boldsymbol{u}$.

$$P(\boldsymbol{X}_j = \boldsymbol{E}_{\boldsymbol{u}}) = p_{\boldsymbol{u}}.$$

These probabilities satisfy the constraint $0 \leq p_{\boldsymbol{u}} \leq 1$ and

$$\sum_{u(1)=1}^{K} \sum_{u(2)=1}^{K} \sum_{u(3)=1}^{K} p_{\boldsymbol{u}} = 1.$$

Similar to Chapter 4, we assume that the true classifications are related to some underlying continuous variables and hence $p_{\boldsymbol{u}}$ is of the form (4.3). That is, $p_{\boldsymbol{u}}$ is a function of the unknown parameters mean, thresholds and the covariance matrix.

(b) $\tau_{\boldsymbol{u},j}$ is the probability of the j-th subject being classified correctly given that the subject actually belongs to the cell $\boldsymbol{u}$. That is

$$P(Y_j = 1|\boldsymbol{X}_j = \boldsymbol{E}_{\boldsymbol{u}}) = \tau_{\boldsymbol{u},j},$$

where the $\tau_{\boldsymbol{u},j}$'s are supposed to be known and are called the honesty probabilities. On the other hand, the probability of the j-th subject being misclassified given that the subject actually belongs to cell $\boldsymbol{u}$ is given by

$$P(Y_j = 0|\boldsymbol{X}_j = \boldsymbol{E}_{\boldsymbol{u}}) = 1 - \tau_{\boldsymbol{u},j}.$$

(c) $\gamma_{\boldsymbol{k}\boldsymbol{u}}$ is the probability of a subject being classified into the cell $\boldsymbol{k}$ given that the subject actually belongs to the cell $\boldsymbol{u}$ and is misclassified.

$$P(\boldsymbol{Z}_\jmath = E_{\boldsymbol{k}}|\boldsymbol{X}_\jmath = \boldsymbol{E}_{\boldsymbol{u}}, Y_\jmath = 0) = \gamma_{\boldsymbol{ku}},$$

where

$$0 \le \gamma_{\boldsymbol{ku}} \le 1, \gamma_{\boldsymbol{uu}} = 0,$$

and $\sum_{k(1)=1}^{K}\sum_{k(2)=1}^{K}\sum_{k(3)=1}^{K}\gamma_{\boldsymbol{ku}} = 1$, for all $\boldsymbol{u}$.

5.2.1 Maximum likelihood estimation

Suppose that we have a data set of size N. Based on the afore-described model, let $z_{\boldsymbol{k},\jmath}$ be the realization of the random variable $Z_{\boldsymbol{k},\jmath}$. The log-likelihood function is then given by

$$\begin{aligned} l &= \log(\textstyle\prod_{\jmath=1}^{N}\prod_{k(1)=1}^{K}\prod_{k(2)=1}^{K}\prod_{k(3)=1}^{K}[Pr(\boldsymbol{Z}_\jmath = \boldsymbol{E}_{\boldsymbol{k}})]^{z_{\boldsymbol{k},\jmath}}) \\ &= \textstyle\sum_{\jmath=1}^{N}\sum_{k(1)=1}^{K}\sum_{k(2)=1}^{K}\sum_{k(3)=1}^{K} z_{\boldsymbol{k},\jmath}\log Pr(\boldsymbol{Z}_\jmath = \boldsymbol{E}_{\boldsymbol{k}}) \\ &= \textstyle\sum_{\jmath=1}^{N}\sum_{k(1)=1}^{K}\sum_{k(2)=1}^{K}\sum_{k(3)=1}^{K} z_{\boldsymbol{k},\jmath} \\ &\quad \log(\textstyle\sum_{u(1)=1}^{K}\sum_{u(2)=1}^{K}\sum_{u(3)=1}^{K}\psi_{\boldsymbol{ku},\jmath}p_{\boldsymbol{u}}), \end{aligned} \tag{5.1}$$

and

$$\psi_{\boldsymbol{ku},\jmath} = \begin{cases} \tau_{\boldsymbol{k},\jmath} & \text{if } \boldsymbol{u} = \boldsymbol{k} \\ \gamma_{\boldsymbol{ku}}(1-\tau_{\boldsymbol{u},\jmath}) & \text{otherwise} \end{cases} \tag{5.2}$$

are the elements of the matrix $\boldsymbol{\Psi}_\jmath$, which is the matrix of transition probabilities with which the j-th subject is transited from the true state to another state with misclassification. It is a two-dimensional matrix of $(K \times K \times K)$ by $(K \times K \times K)$, and

$\sum_k \psi_{\boldsymbol{ku},j} = 1 \qquad \text{for} \qquad 1 \leq j \leq N.$

For the problem with known misclassifications, $\psi_{\boldsymbol{ku},j}$ values are known constants. Similar to Chapter 4, the unknown parameter vector is $\boldsymbol{\theta} = (\mu_R, \mu_T, \sigma_{PR}, \sigma_R^2, \sigma_{PT}, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2))^T$. Let θ_k be the kth element in $\boldsymbol{\theta}$. As parameters in $\boldsymbol{\theta}$ are only involved in the cell probability $p_{\boldsymbol{u}}$ in (5.1), the ML estimate $\hat{\boldsymbol{\theta}}$ can be obtained by solving:

$$0 = \frac{\partial l}{\partial \theta_k} = \sum_{v(1)=1}^{K} \sum_{v(2)=1}^{K} \sum_{v(3)=1}^{K} \frac{\partial l}{\partial p_{\boldsymbol{v}}} \times \frac{\partial p_{\boldsymbol{v}}}{\partial \theta_k}, \tag{5.3}$$

where $k = 1, ..., s$, and s is the number of unknown parameters or the dimension of $\boldsymbol{\theta}$. Taking into consideration the accessibility by practitioners, we also use Mx to obtain the ML estimates $\hat{\boldsymbol{\theta}}$. The standard error estimates are acquired through bootstrap method. We will test non-inferiority in three-armed design in a way as described in Section 4.3. That is, we find the CI for ψ^* that is given by $(\hat{\psi}^* - z_\alpha SE(\hat{\psi}^*), +\infty)$, where z_α is the upper α probability point of the standard normal distribution function. We employ the CI to test non-inferiority, and the non-inferiority can be concluded at the significant level α if $\hat{\psi}^* - z_\alpha SE(\hat{\psi}^*) > 0$.

5.2.2 Simulation

We used a simulation study to assess the performance of the proposed method for testing non-inferiority with the presence of misclassified data.

A number of data sets are generated based on a set of known parameter values. We set K equal to 3. The true values for the parameters are $\boldsymbol{\theta} = (\mu_R, \mu_T, \sigma_{PR}, \sigma_R^2, \sigma_{PT}, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2))^T = (0.2, 0.4, 0.3, 1, 0.3,\ 0.3, 1, -0.2, 0.7)^T$. Moreover, we also examined the performance of ψ^* since it is the benchmark of decision making in our hypothesis testing. Given $\boldsymbol{\theta}$ and $f = -\frac{1}{5}$, the true value of ψ^* is 0.24.

The number of simulation replications was 1000 and we considered three different sample sizes: 2000, 1000 and 800. Assuming $\tau_{\boldsymbol{u},j} = \tau_{\boldsymbol{u}}$ for all $j = 1, ..., N$, three different misclassification probabilities $\tau_{\boldsymbol{u}} = 0.7, 0.8$ and 0.9 were used in the simulation.

The set of $\gamma_{\boldsymbol{k}\boldsymbol{u}}$ values that were adopted to analyze the data is presented in Table 5.1. The values were compiled based on the assumption that misclassification only occurred between adjacent cells, with equal probability.

The $\gamma_{\boldsymbol{k}\boldsymbol{u}}$ values combined with the $\tau_{\boldsymbol{u}}$ values produce the transition misclassification probability matrix $\boldsymbol{\Psi}$, where $\boldsymbol{\Psi} = \boldsymbol{\Psi}_j$ for all j as $\tau_{\boldsymbol{u},j} = \tau_{\boldsymbol{u}}$ for all j. For the different $\tau_{\boldsymbol{u}} = 0.7$, 0.8, and 0.9, the resulting $\boldsymbol{\Psi}$ matrices are reported in Tables 5.2 to 5.4. The simulation results are presented in Tables 5.5 to 5.13, with different values of $\tau_{\boldsymbol{u}}$ indicating different misclassification levels. We used the same assessment standards as those in Chapter 2 to evaluate the proposed method for three-armed design with misclassification.

Table 5.1: γ_{ku} values

k/u	111	112	113	121	122	123	131	132	133	211	212	213	221	222	223	231	232	233	311	312	313	321	322	323	331	332	333
111	0	1/4	0	1/4	0	0	0	0	0	1/4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
112	1/3	0	1/3	0	1/5	0	0	0	0	0	1/5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
113	0	1/4	0	0	0	1/4	0	0	0	0	0	1/4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121	1/3	0	0	0	1/5	0	1/3	0	0	0	0	0	1/5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	0	1/4	0	1/4	0	1/4	0	1/4	0	0	0	0	0	1/6	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	1/3	0	1/5	0	0	0	1/3	0	0	0	0	0	1/5	0	0	0	0	0	0	0	0	0	0	0	0
131	0	0	0	1/4	0	0	0	1/4	0	0	0	0	0	0	0	1/4	0	0	0	0	0	0	0	0	0	0	0
132	0	0	0	0	1/5	0	1/3	0	1/3	0	0	0	0	0	0	0	1/5	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	0	1/4	0	1/4	0	0	0	0	0	0	0	0	0	1/4	0	0	0	0	0	0	0	0	0
211	1/3	0	0	0	0	0	0	0	0	0	1/5	0	1/5	0	0	0	0	0	1/3	0	0	0	0	0	0	0	0
212	0	1/4	0	0	0	0	0	0	0	1/4	0	1/4	0	1/6	0	0	0	0	0	1/4	0	0	0	0	0	0	0
213	0	0	1/3	0	0	0	0	0	0	0	1/5	0	0	0	1/5	0	0	0	0	0	1/3	0	0	0	0	0	0
221	0	0	0	1/4	0	0	0	0	0	1/4	0	0	0	1/6	0	1/4	0	0	0	0	0	1/4	0	0	0	0	0
222	0	0	0	0	1/5	0	0	0	0	0	1/5	0	1/5	0	1/5	0	1/5	0	0	0	0	0	1/5	0	0	0	0
223	0	0	0	0	0	1/4	0	0	0	0	0	1/4	0	1/6	0	0	0	1/4	0	0	0	0	0	1/4	0	0	0
231	0	0	0	0	0	0	1/3	0	0	0	0	0	1/5	0	0	0	1/5	0	0	0	0	0	0	0	1/3	0	0
232	0	0	0	0	0	0	0	1/4	0	0	0	0	0	1/6	0	1/4	0	1/4	0	0	0	0	0	0	0	1/4	0
233	0	0	0	0	0	0	0	0	1/3	0	0	0	0	0	1/5	0	1/5	0	0	0	0	0	0	0	0	0	1/3
311	0	0	0	0	0	0	0	0	0	1/4	0	0	0	0	0	0	0	0	0	1/4	0	1/4	0	0	0	0	0
312	0	0	0	0	0	0	0	0	0	0	1/5	0	0	0	0	0	0	0	1/3	0	1/3	0	1/5	0	0	0	0
313	0	0	0	0	0	0	0	0	0	0	0	1/4	0	0	0	0	0	0	0	1/4	0	0	0	1/4	0	0	0
321	0	0	0	0	0	0	0	0	0	0	0	0	1/5	0	0	0	0	0	1/3	0	0	0	1/5	0	1/3	0	0
322	0	0	0	0	0	0	0	0	0	0	0	0	0	1/6	0	0	0	0	0	1/4	0	1/4	0	1/4	0	1/4	0
323	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1/5	0	0	0	0	0	1/3	0	1/5	0	0	0	1/3
331	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1/4	0	0	0	0	0	1/4	0	0	0	1/4	0
332	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1/5	0	0	0	0	0	1/5	0	1/3	0	1/3
333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1/4	0	0	0	0	0	1/4	0	1/4	0

Table 5.2: Transition misclassification probability matrix Ψ for $\tau_{\boldsymbol{u}} = 0.7$

k/u	111	112	113	121	122	123	131	132	133	211	212	213	221	222	223	231	232	233	311	312	313	321	322	323	331	332	333
111	.7	.075	0	.075	0	0	0	0	0	.075	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
112	.1	.7	.1	0	.06	0	0	0	0	0	.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
113	0	.075	.7	0	0	.075	0	0	0	0	0	.075	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121	.1	0	0	.7	.06	0	.1	0	0	0	0	0	.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	0	.075	0	.075	.7	.075	0	.075	0	0	0	0	0	.05	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	.1	0	.06	.7	0	0	.1	0	0	0	0	0	.06	0	0	0	0	0	0	0	0	0	0	0	0
131	0	0	0	.075	0	0	.7	.075	0	0	0	0	0	0	0	.075	0	0	0	0	0	0	0	0	0	0	0
132	0	0	0	0	.06	0	.1	.7	.1	0	0	0	0	0	0	0	.06	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	0	.075	0	.075	.7	0	0	0	0	0	0	0	0	.075	0	0	0	0	0	0	0	0	0
211	.1	0	0	0	0	0	0	0	0	.7	.06	0	.06	0	0	0	0	0	.1	0	0	0	0	0	0	0	0
212	0	.075	0	0	0	0	0	0	0	.075	.7	.075	0	.05	0	0	0	0	0	.075	0	0	0	0	0	0	0
213	0	0	.1	0	0	0	0	0	0	0	.06	.7	0	0	.06	0	0	0	0	0	.1	0	0	0	0	0	0
221	0	0	0	.075	0	0	0	0	0	.075	0	0	.7	.05	0	.075	0	0	0	0	0	.075	0	0	0	0	0
222	0	0	0	0	.06	0	0	0	0	0	.06	0	.06	.7	.06	0	.06	0	0	0	0	0	.06	0	0	0	0
223	0	0	0	0	0	.075	0	0	0	0	0	.075	0	.05	.7	0	0	.075	0	0	0	0	0	.075	0	0	0
231	0	0	0	0	0	0	.1	0	0	0	0	0	.06	0	0	.7	.06	0	0	0	0	0	0	0	.1	0	0
232	0	0	0	0	0	0	0	.075	0	0	0	0	0	.05	0	.075	.7	.075	0	0	0	0	0	0	0	.075	0
233	0	0	0	0	0	0	0	0	.1	0	0	0	0	0	.06	0	.06	.7	0	0	0	0	0	0	0	0	.1
311	0	0	0	0	0	0	0	0	0	.075	0	0	0	0	0	0	0	0	.7	.075	0	.075	0	0	0	0	0
312	0	0	0	0	0	0	0	0	0	0	.06	0	0	0	0	0	0	0	.1	7	.1	0	.06	0	0	0	0
313	0	0	0	0	0	0	0	0	0	0	0	.075	0	0	0	0	0	0	0	.075	.7	0	0	.075	0	0	0
321	0	0	0	0	0	0	0	0	0	0	0	0	.06	0	0	0	0	0	.1	0	0	.7	06	0	.1	0	0
322	0	0	0	0	0	0	0	0	0	0	0	0	0	.05	0	0	0	0	0	.075	0	.075	.7	.075	0	.075	0
323	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.06	0	0	0	0	0	.1	0	.06	.7	0	0	.1
331	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.075	0	0	0	0	0	.075	0	0	.7	.075	0
332	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.06	0	0	0	0	0	.06	0	.1	.7	.1
333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.075	0	0	0	0	0	.075	0	.075	.7

Table 5.3: Transition misclassification probability matrix $\mathbf{\Psi}$ for $\tau_{\boldsymbol{u}} = 0.8$

k/u	111	112	113	121	122	123	131	132	133	211	212	213	221	222	223	231	232	233	311	312	313	321	322	323	331	332	333
111	8	05	0	05	0	0	0	0	0	05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
112	067	8	067	0	04	0	0	0	0	0	04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
113	0	05	8	0	0	05	0	0	0	0	0	05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121	067	0	0	8	04	0	067	0	0	0	0	0	04	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	0	05	0	05	8	05	0	05	0	0	0	0	0	033	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	067	0	04	8	0	0	067	0	0	0	0	0	04	0	0	0	0	0	0	0	0	0	0	0	0
131	0	0	0	05	0	0	8	05	0	0	0	0	0	0	0	05	0	0	0	0	0	0	0	0	0	0	0
132	0	0	0	0	04	0	067	8	067	0	0	0	0	0	0	0	04	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	0	05	0	05	8	0	0	0	0	0	0	0	0	05	0	0	0	0	0	0	0	0	0
211	067	0	0	0	0	0	0	0	0	8	04	0	04	0	0	0	0	0	067	0	0	0	0	0	0	0	0
212	0	05	0	0	0	0	0	0	0	05	8	05	0	033	0	0	0	0	0	05	0	0	0	0	0	0	0
213	0	0	067	0	0	0	0	0	0	0	04	8	0	0	04	0	0	0	0	0	067	0	0	0	0	0	0
221	0	0	0	05	0	0	0	0	0	05	0	0	8	033	0	05	0	0	0	0	0	05	0	0	0	0	0
222	0	0	0	0	0 04	0	0	0	0	0	04	0	04	8	04	0	04	0	0	0	0	0	04	0	0	0	0
223	0	0	0	0	0	05	0	0	0	0	0	05	0	033	8	0	0	05	0	0	0	0	0	05	0	0	0
231	0	0	0	0	0	0	067	0	0	0	0	0	04	0	0	8	04	0	0	0	0	0	0	0	067	0	0
232	0	0	0	0	0	0	0	05	0	0	0	0	0	033	0	05	8	05	0	0	0	0	0	0	0	05	0
233	0	0	0	0	0	0	0	0	067	0	0	0	0	0	04	0	04	8	0	0	0	0	0	0	0	0	067
311	0	0	0	0	0	0	0	0	0	05	0	0	0	0	0	0	0	0	8	05	0	05	0	0	0	0	0
312	0	0	0	0	0	0	0	0	0	0	04	0	0	0	0	0	0	0	067	8	067	0	04	0	0	0	0
313	0	0	0	0	0	0	0	0	0	0	0	05	0	0	0	0	0	0	0	05	8	0	0	05	0	0	0
321	0	0	0	0	0	0	0	0	0	0	0	0	04	0	0	0	0	0	067	0	0	8	04	0	067	0	0
322	0	0	0	0	0	0	0	0	0	0	0	0	0	033	0	0	0	0	0	05	0	05	8	05	0	05	0
323	0	0	0	0	0	0	0	0	0	0	0	0	0	0	04	0	0	0	0	0	067	0	04	8	0	0	067
331	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	05	0	0	0	0	0	05	0	0	8	05	0
332	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	04	0	0	0	0	0	04	0	067	8	067
333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	05	0	0	0	0	0	05	0	05	8

Table 5.4: Transition misclassification probability matrix $\mathbf{\Psi}$ for $\tau_{\boldsymbol{u}} = 0.9$

k/u	111	112	113	121	122	123	131	132	133	211	212	213	221	222	223	231	232	233	311	312	313	321	322	323	331	332	333
111	9	025	0	025	0	0	0	0	0	025	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
112	033	9	033	0	02	0	0	0	0	0	02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
113	0	025	9	0	0	025	0	0	0	0	0	025	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121	033	0	0	9	02	0	033	0	0	0	0	0	02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	0	025	0	025	9	025	0	025	0	0	0	0	0	017	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	033	0	02	9	0	0	033	0	0	0	0	0	02	0	0	0	0	0	0	0	0	0	0	0	0
131	0	0	0	025	0	0	9	025	0	0	0	0	0	0	0	025	0	0	0	0	0	0	0	0	0	0	0
132	0	0	0	0	02	0	033	9	033	0	0	0	0	0	0	0	02	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	0	025	0	025	9	0	0	0	0	0	0	0	0	025	0	0	0	0	0	0	0	0	0
211	033	0	0	0	0	0	0	0	0	9	02	0	02	0	0	0	0	0	033	0	0	0	0	0	0	0	0
212	0	025	0	0	0	0	0	0	0	025	9	025	0	017	0	0	0	0	0	025	0	0	0	0	0	0	0
213	0	0	033	0	0	0	0	0	0	0	02	9	0	0	02	0	0	0	0	0	033	0	0	0	0	0	0
221	0	0	0	025	0	0	0	0	0	025	0	0	9	017	0	025	0	0	0	0	0	025	0	0	0	0	0
222	0	0	0	0	02	0	0	0	0	0	02	0	02	9	02	0	02	0	0	0	0	0	02	0	0	0	0
223	0	0	0	0	0	025	0	0	0	0	0	025	0	017	9	0	0	025	0	0	0	0	0	025	0	0	0
231	0	0	0	0	0	0	033	0	0	0	0	0	02	0	0	9	02	0	0	0	0	0	0	0	033	0	0
232	0	0	0	0	0	0	0	025	0	0	0	0	0	017	0	025	9	025	0	0	0	0	0	0	0	025	0
233	0	0	0	0	0	0	0	0	033	0	0	0	0	0	02	0	02	9	0	0	0	0	0	0	0	0	033
311	0	0	0	0	0	0	0	0	0	025	0	0	0	0	0	0	0	0	9	025	0	025	0	0	0	0	0
312	0	0	0	0	0	0	0	0	0	0	02	0	0	0	0	0	0	0	033	9	033	0	02	0	0	0	0
313	0	0	0	0	0	0	0	0	0	0	0	025	0	0	0	0	0	0	0	025	9	0	0	025	0	0	0
321	0	0	0	0	0	0	0	0	0	0	0	0	02	0	0	0	0	0	033	0	0	9	02	0	033	0	0
322	0	0	0	0	0	0	0	0	0	0	0	0	0	017	0	0	0	0	0	025	0	025	9	025	0	025	0
323	0	0	0	0	0	0	0	0	0	0	0	0	0	0	02	0	0	0	0	0	033	0	02	9	0	0	033
331	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	025	0	0	0	0	0	025	0	0	9	025	0
332	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	02	0	0	0	0	0	02	0	033	9	033
333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	025	0	0	0	0	0	025	0	025	9

Table 5.5: Simulation data with $\tau_{\boldsymbol{u}} = 0.7$ known for sample size 2000

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	σ_{RT}	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.199	0.400	0.294	1.009	0.292	0.292	1.009	-0.251	0.801	0.241
RMS	0.035	0.038	0.036	0.132	0.038	0.047	0.137	0.060	0.107	0.034
SE	0.035	0.039	0.037	0.135	0.037	0.048	0.137	0.032	0.037	0.035
SD	0.035	0.038	0.036	0.132	0.037	0.047	0.137	0.032	0.036	0.034
R	0.998	0.996	0.974	0.981	0.998	0.983	1.001	0.988	0.982	0.972

Table 5.6: Simulation data with $\tau_{\boldsymbol{u}} = 0.8$ known for sample size 2000

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.201	0.397	0.296	1.009	0.295	0.296	1.003	-0.230	0.758	0.237
RMS	0.034	0.037	0.035	0.129	0.036	0.045	0.128	0.042	0.067	0.033
SE	0.034	0.037	0.035	0.125	0.035	0.045	0.125	0.031	0.033	0.033
SD	0.034	0.037	0.034	0.129	0.035	0.045	0.128	0.030	0.033	0.033
R	1.001	0.997	0.983	1.032	1.005	1.009	1.017	0.985	1.001	0.979

Table 5.7: Simulation data with $\tau_{\boldsymbol{u}} = 0.9$ known for sample size 2000

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.199	0.400	0.298	1.002	0.297	0.296	1.000	-0.213	0.724	0.241
RMS	0.033	0.035	0.032	0.114	0.033	0.041	0.116	0.032	0.038	0.032
SE	0.033	0.036	0.033	0.116	0.033	0.042	0.117	0.029	0.029	0.032
SD	0.033	0.035	0.031	0.114	0.033	0.041	0.116	0.029	0.030	0.032
R	1.017	0.980	0.955	0.977	1.000	0.981	0.991	1.003	1.007	0.974

Table 5.8: Simulation data with $\tau_{\boldsymbol{u}} = 0.7$ known for sample size 1000

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.197	0.400	0.290	1.010	0.293	0.294	1.009	-0.250	0.800	0.242
RMS	0.050	0.054	0.053	0.181	0.053	0.065	0.181	0.068	0.111	0.049
SE	0.050	0.055	0.053	0.194	0.053	0.069	0.195	0.045	0.052	0.050
SD	0.050	0.054	0.052	0.181	0.052	0.065	0.181	0.046	0.049	0.049
R	1.017	0.998	0.987	0.933	0.984	0.941	0.927	1.015	0.943	0.989

Table 5.9: Simulation data with $\tau_u = 0.8$ known for sample size 1000

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.198	0.399	0.294	1.012	0.296	0.296	1.013	-0.230	0.759	0.241
RMS	0.049	0.052	0.050	0.169	0.049	0.061	0.170	0.053	0.074	0.048
SE	0.048	0.053	0.050	0.180	0.050	0.065	0.182	0.043	0.046	0.048
SD	0.049	0.052	0.049	0.169	0.049	0.061	0.169	0.044	0.044	0.048
R	1.027	0.999	0.996	0.939	0.979	0.943	0.931	1.019	0.947	0.992

Table 5.10: Simulation data with $\tau_u = 0.9$ known for sample size 1000

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.199	0.400	0.296	1.010	0.299	0.299	1.012	-0.213	0.726	0.241
RMS	0.047	0.051	0.047	0.160	0.048	0.058	0.157	0.044	0.046	0.046
SE	0.046	0.050	0.047	0.167	0.048	0.061	0.169	0.041	0.042	0.046
SD	0.047	0.051	0.047	0.159	0.048	0.058	0.156	0.042	0.039	0.046
R	1.019	1.003	0.996	0.956	1.015	0.953	0.925	1.019	0.931	0.998

Table 5.11: Simulation data with $\tau_u = 0.7$ known for sample size 800

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.196	0.396	0.292	1.021	0.293	0.299	1.016	-0.252	0.800	0.240
RMS	0.054	0.061	0.060	0.210	0.059	0.074	0.215	0.073	0.115	0.056
SE	0.055	0.061	0.059	0.222	0.060	0.078	0.221	0.051	0.058	0.056
SD	0.054	0.061	0.060	0.209	0.059	0.074	0.215	0.051	0.058	0.056
R	0.979	1.000	1.003	0.945	0.990	0.944	0.973	1.010	1.013	1.006

Table 5.12: Simulation data with $\tau_u = 0.8$ known for sample size 800

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.195	0.397	0.295	1.019	0.296	0.301	1.018	-0.231	0.759	0.241
RMS	0.053	0.060	0.057	0.194	0.056	0.069	0.201	0.058	0.079	0.055
SE	0.054	0.059	0.056	0.204	0.056	0.073	0.206	0.048	0.052	0.054
SD	0.053	0.060	0.057	0.194	0.056	0.069	0.201	0.049	0.052	0.055
R	0.981	1.014	1.015	0.947	0.992	0.945	0.976	1.007	1.009	1.018

Table 5.13: Simulation data with $\tau_u = 0.9$ known for sample size 800

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.197	0.397	0.298	1.020	0.298	0.305	1.016	-0.214	0.726	0.240
RMS	0.052	0.057	0.054	0.185	0.053	0.067	0.188	0.049	0.053	0.052
SE	0.052	0.057	0.053	0.190	0.053	0.069	0.191	0.046	0.047	0.052
SD	0.052	0.057	0.054	0.184	0.053	0.067	0.188	0.047	0.047	0.052
R	0.997	1.012	1.026	0.967	0.991	0.962	0.986	1.023	1.005	1.006

Table 5.14: Non-inferiority test with τ_u known

τ_u	NOBS	ECP
0.7	2000	0.964
0.8	2000	0.969
0.9	2000	0.958
0.7	1000	0.962
0.8	1000	0.958
0.9	1000	0.967
0.7	800	0.959
0.8	800	0.957
0.9	800	0.957

Moreover, in Table 5.14, we have summarized the ECPs of the confidence intervals of ψ^*. The values are close to but slightly larger than 0.95 in all cases, suggesting that the use of the CIs to examine non-inferiority will produce reliable results.

We also summarized the simulation results with respect to the different sample sizes (NOBS) and τ_u values in the Figures 5.2 and 5.3, where the three different curves stand for the results with $\tau_u = 0.7$ (MIS7), $\tau_u = 0.8$ (MIS8), and $\tau_u = 0.9$ (MIS9), respectively (see Figure 5.1). The ten parameters in $\mu_R, \mu_T, \sigma_{PR}, \sigma_R^2, \sigma_{PT}, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2), \psi^*$ are labeled from Parameter1, Parameter2 to Parameter10 (See Figure 5.1). Generally, for all the parameters, the values of RMS decrease to 0 and the values of the ratio SD/SE get closer to 1 as sample size increases, indicating that: (i) the parameter estimates are accurate; and (ii) the estimated standard errors are reliable.

$Parameter1 = \hat{\mu}_R$, $Parameter2 = \hat{\mu}_T$, $Parameter3 = \hat{\sigma}_{PR}$, $Parameter4 = \hat{\sigma}_R^2$, $Parameter5 = \hat{\sigma}_{PT}$, $Parameter6 = \hat{\sigma}_{RT}$, $Parameter7 = \hat{\sigma}_T^2$, $Parameter8 = \hat{\beta}_2$, $Parameter9 = \hat{\beta}_3 - \hat{\beta}_2$, and $Parameter10 = \hat{\psi}^*$

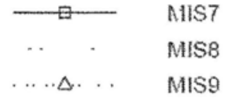

Figure 5.1: Graph legend

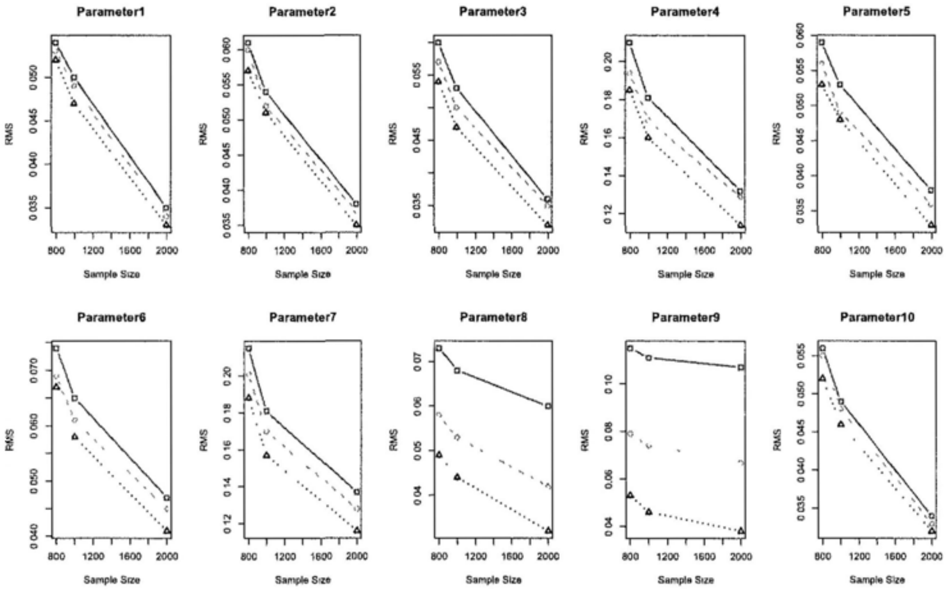

Figure 5.2: RMS, known misclassification rate

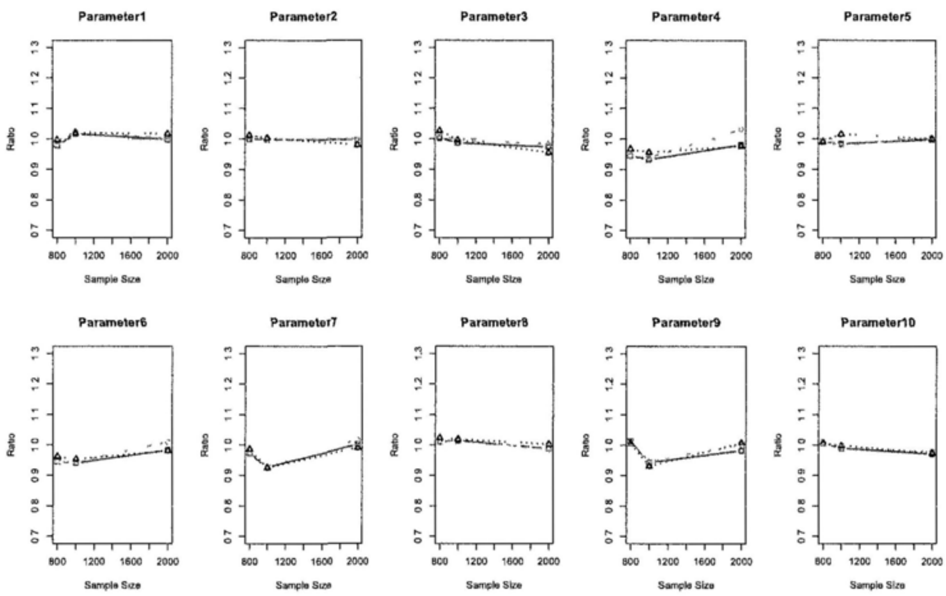

Figure 5.3: SD/SE ratio, known misclassification rate

5.3 Example 7

We analyzed the data set in Example 6 with the data given in Tables 4.6, 4.7 and 4.8. As an illustration, we assume that the transition matrix is known and is the one in Table 5.3 with honesty rate 0.8.

With the proposed method, we can obtained $\hat{\mu}_R = 1.517$, $\hat{\mu}_T = 1.260$, $\hat{\sigma}_{PR} = -0.353$, $\hat{\sigma}^2_R = 0.96$, $\hat{\sigma}_{PT} = -0.282$, $\hat{\sigma}_{RT} = 0.344$, $\hat{\sigma}^2_T = 1.152$, $\hat{\beta}_2 = 0.636$, $\hat{\beta}_3 - \hat{\beta}_2 = 1.1$, and $\hat{\psi}^* = 0.046$. The 95% CI for ψ^* is $(-0.245, +\infty)$. Since the lower bound of CI -0.245 is a negative value, the non-inferiority can not be concluded, which is consistent with the conclusion of Example 6. That is, even if there exists misclassifications in the data set, we are confident that the conclusion will be the same.

5.4 Block Design with Partially Validated Data

5.4.1 Model for partially validated data

When the misclassification probabilities are not known, we can use partially validated data to seek information in relation to misclassification. Similar as before, partially validated data is assumed to have obtained by using two devices to classify the responses. The true classifier can classify units correctly but usually with higher cost. The fallible classifier may lead to misclassification but it is less expensive.

Let $\boldsymbol{k}$ represent the cell classified by the fallible device, and $\boldsymbol{u}$ the one for true measurement device. The total sample size of a data set that is obtained from partially validated data is N. Out of which n randomly drawn units are classified by both devices and $n_{\boldsymbol{k}(\boldsymbol{u})}$ is the total number of units for which the true and fallible devices result to the classification to cells $\boldsymbol{u}$ and $\boldsymbol{k}$, respectively. The remaining $n^* = N - n$ units are classified by the fallible device only, and $n^*_{\boldsymbol{k}}$ is the total number of units that have been classified into cell $\boldsymbol{k}$. The data structure is summarized in Table 5.15.

Table 5.15: Block design data structure in partially validated data method

True	Fallible $(h\ k\ l)$											Grand
$(u\ v, t)$	$(1,1,1)$		$(1,1,K)$		$(1,K\ 1)$		$(1\ K,K)$		$(K\ K\ 1)$		$(K\ K\ K)$	Total
$(1,1,1)$	$n_{111(111)}$		$n_{11K(111)}$		$n_{1K1(111)}$		$n_{1KK(111)}$		$n_{KK1(111)}$		$n_{KKK(111)}$	$n_{+(111)}$
$(1,1,K)$	$n_{111(11K)}$		$n_{11K(11K)}$		$n_{1K1(11K)}$		$n_{1KK(11K)}$		$n_{KK1(11K)}$		$n_{KKK(11K)}$	$n_{+(11K)}$
$(1,K,1)$	$n_{111(1K1)}$		$n_{11K(1K1)}$		$n_{1K1(1K1)}$		$n_{1KK(1K1)}$		$n_{KK1(1K1)}$		$n_{KKK(1K1)}$	$n_{+(1K1)}$
$(1,K,K)$	$n_{111(1KK)}$		$n_{11K(1KK)}$		$n_{1K1(1KK)}$		$n_{1KK(1KK)}$		$n_{KK1(1KK)}$		$n_{KKK(1KK)}$	$n_{+(1KK)}$
$(K,K,1)$	$n_{111(KK1)}$		$n_{11K(KK1)}$		$n_{1K1(KK1)}$		$n_{1KK(KK1)}$		$n_{KK1(KK1)}$		$n_{KKK(KK1)}$	$n_{+(KK1)}$
(K,K,K)	$n_{111(KKK)}$		$n_{11K(KKK)}$		$n_{1K1(KKK)}$		$n_{1KK(KKK)}$		$n_{KK1(KKK)}$		$n_{KKK(KKK)}$	$n_{+(KKK)}$
Total	$n_{111(+)}$		$n_{11K(+)}$		$n_{1K1(+)}$		$n_{1KK(+)}$		$n_{KK1(+)}$		$n_{KKK(+)}$	n
	n^*_{111}		n^*_{11K}		n^*_{1K1}		n^*_{1KK}		n^*_{KK1}		n^*_{KKK}	n^*
Grand Total	$N_{111} = n_{111(+)} + n^*_{111}$		$N_{11K} = n_{11K(+)} + n^*_{11K}$		$N_{1K1} = n_{1K1(+)} + n^*_{1K1}$		$N_{1KK} = n_{1KK(+)} + n^*_{1KK}$		$N_{KK1} = n_{KK1(+)} + n^*_{KK1}$		$N_{KKK} = n_{KKK(+)} + n^*_{KKK}$	$N = n + n^*$

Let $p_{\boldsymbol{u}}$ be the probability that a unit actually belongs to cell $\boldsymbol{u}$; $\pi_{\boldsymbol{k}}$ be the probability that a unit is classified into cell $\boldsymbol{k}$ by the fallible device; and $\omega_{\boldsymbol{k}(\boldsymbol{u})}$ be the probability that one is classified into cell $\boldsymbol{k}$ when it actually belongs to cell $\boldsymbol{u}$. As before, it can be shown that the ML estimates $\hat{p}_{\boldsymbol{u}}$ of $p_{\boldsymbol{u}}$ and $\hat{\omega}_{\boldsymbol{k}(\boldsymbol{u})}$ of $\omega_{\boldsymbol{k}(\boldsymbol{u})}$ are given as follows:

$$\hat{p}_{\boldsymbol{u}} = \sum_{\boldsymbol{k}} \frac{(n^*_{\boldsymbol{k}} + n_{\boldsymbol{k}(+)})n_{\boldsymbol{k}(\boldsymbol{u})}}{Nn_{\boldsymbol{k}(+)}}, \tag{5.4}$$

$$\hat{\omega}_{\boldsymbol{k}(\boldsymbol{u})} = \frac{(n^*_{\boldsymbol{k}} + n_{\boldsymbol{k}(+)})n_{\boldsymbol{k}(\boldsymbol{u})}}{Nn_{\boldsymbol{k}(+)}\hat{p}_{\boldsymbol{u}}}. \tag{5.5}$$

Let p^* be a $(K \times K \times K) \times 1$ vector that stores $p_{\boldsymbol{u}}$'s for $u(1) = 1, \ldots, K; u(2) = 1, \ldots, K; u(3) = 1, \ldots, K$, and $\hat{p}^*$ be its ML estimate with elements that are given by (5.4).

Let $\boldsymbol{\theta} = (\mu_R, \mu_T, \sigma_{PR}, \sigma^2_R, \sigma_{PT}, \sigma_{RT}, \sigma^2_T, \boldsymbol{\beta})^T$ be the unknown parameters vector as before, then $p^* = p^*(\boldsymbol{\theta})$. The estimate $\hat{\boldsymbol{\theta}}$ can be obtained by the method of Modified Minimum Chi-square (MMC) with $\hat{\boldsymbol{\theta}}$ minimizing the function

$$G(\boldsymbol{\theta}) = (\hat{p}^* - p^*(\boldsymbol{\theta}))^T \hat{\boldsymbol{\Omega}}^{-1} (\hat{p}^* - p^*(\boldsymbol{\theta})), \tag{5.6}$$

where $\boldsymbol{\Omega}$ is a $(K \times K \times K)$ by $(K \times K \times K)$ matrix. The diagonal elements of $\boldsymbol{\Omega}$ that correspond to $\hat{p}_{\boldsymbol{u}}$ are given by

$$\frac{p_{\boldsymbol{u}}}{n} + p^2_{\boldsymbol{u}}(\frac{1}{N} - \frac{1}{n}) \sum_{\boldsymbol{k}} \frac{\omega^2_{\boldsymbol{k}(\boldsymbol{u})}}{\pi_{\boldsymbol{k}}},$$

and the off-diagonal elements that correspond to $\hat{p}_{\boldsymbol{u}}$ and $\hat{p}_{\boldsymbol{u}'}$ for

$\boldsymbol{u} \neq \boldsymbol{u}'$ are given by

$$\frac{p_{\boldsymbol{u}} p_{\boldsymbol{u}'}}{N} \sum_{\boldsymbol{k}} \frac{\omega_{\boldsymbol{k}(\boldsymbol{u})} \omega_{\boldsymbol{k}(\boldsymbol{u}')}}{\pi_{\boldsymbol{k}}}. \tag{5.7}$$

The matrix $\boldsymbol{\Omega}$ can be derived analytically by generalizing the method in Yiu & Poon (2008) and Zhang (2007), and $\hat{\boldsymbol{\Omega}}$ is obtained by replacing the unknown parameters in $\boldsymbol{\Omega}$ with their consistent estimates. We construct program in the freely available software Mx to get the estimates of the unknown parameters.

5.4.2 Simulation

To examine the performance of the proposed method by simulation, a number of data sets that are based on a set of known parameter values were generated.

We set K equal to 3 and set the true parameter values $\boldsymbol{\theta} = (\mu_R, \mu_T, \sigma_{PR}, \sigma_R^2, \sigma_{PT}, \sigma_{RT}, \sigma_T^2, \beta_2, (\beta_3 - \beta_2))^T$ equal to $(0.2, 0.4, 0.3, 1, 0.3, 0.3, 1, -0.2, 0.7)^T$. The two thresholds are set to be identical as $\beta_2 = -0.2$ and $\beta_3 = 0.5$ for all of the three dimensions. In particular, we examined the performance of ψ^* since it is the benchmark of decision making in our hypothesis testing. Given $\boldsymbol{\theta}$ and $f = -\frac{1}{5}$, the true value of ψ^* is 0.24.

To generate the data with misclassification we make use of the misclassification matrices in Tables 5.2, 5.3 and 5.4 that were compiled based on different $\tau_{\boldsymbol{u}}$ values and the assumption that misclassification only arises in the adjacent cells with equal probability.

The simulation results are summarized by using both tables and graphs. To make it clear in the graph, we use NOBS and NRES to replace N and n in the presentation. We simulated three different combinations of data sets with sizes $(N, n) = (NOBS, NRES) = (2000, 1200), (1500, 900)$ and $(1000, 600)$. The simulation steps are the same as those in Section 3.3.1.

As before, there are three different misclassification probabilities used in the simulation. They are $\tau_{\boldsymbol{u}} = 0.7$, $\tau_{\boldsymbol{u}} = 0.8$ and $\tau_{\boldsymbol{u}} = 0.9$. The simulation results are presented in Tables 5.16 to 5.24 as follows.

Table 5.16: Simulation result for NOBS=2000 NRES=1200 $\tau_{\boldsymbol{u}} = 0.7$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.205	0.409	0.305	1.011	0.307	0.306	1.004	-0.194	0.697	0.246
RMS	0.036	0.040	0.037	0.137	0.038	0.047	0.129	0.032	0.033	0.036
SE	0.037	0.041	0.038	0.136	0.039	0.049	0.136	0.032	0.033	0.037
SD	0.035	0.039	0.036	0.136	0.038	0.047	0.129	0.032	0.032	0.036
R	0.958	0.956	0.964	1.002	0.977	0.953	0.949	0.981	0.988	0.970

Table 5.17: Simulation result for NOBS=2000 NRES=1200 $\tau_u = 0.8$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.205	0.409	0.305	1.010	0.306	0.305	1.004	-0.195	0.697	0.245
RMS	0.035	0.039	0.035	0.130	0.037	0.045	0.126	0.031	0.032	0.036
SE	0.036	0.039	0.036	0.130	0.037	0.047	0.130	0.031	0.031	0.036
SD	0.035	0.038	0.035	0.130	0.036	0.045	0.126	0.031	0.032	0.035
R	0.979	0.972	0.960	1.000	0.985	0.957	0.969	0.986	1.005	0.996

Table 5.18: Simulation result for NOBS=2000 NRES=1200 $\tau_u = 0.9$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}_R^2$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}_T^2$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.205	0.408	0.305	1.010	0.306	0.305	1.005	-0.196	0.698	0.244
RMS	0.034	0.037	0.033	0.123	0.035	0.043	0.117	0.030	0.030	0.034
SE	0.034	0.038	0.035	0.123	0.035	0.045	0.123	0.030	0.030	0.034
SD	0.033	0.037	0.033	0.122	0.034	0.043	0.117	0.030	0.030	0.034
R	0.979	0.972	0.951	0.996	0.977	0.955	0.954	0.988	0.996	0.993

Table 5.19: Simulation result for NOBS=1500 NRES=900 $\tau_u = 0.7$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.204	0.411	0.306	1.012	0.308	0.311	1.016	-0.196	0.697	0.248
RMS	0.041	0.047	0.043	0.149	0.044	0.057	0.155	0.037	0.037	0.043
SE	0.044	0.049	0.045	0.163	0.046	0.060	0.166	0.038	0.039	0.044
SD	0.041	0.046	0.042	0.149	0.044	0.056	0.154	0.036	0.037	0.042
R	0.935	0.922	0.942	0.915	0.942	0.941	0.930	0.944	0.956	0.943

Table 5.20: Simulation result for NOBS=1500 NRES=900 $\tau_u = 0.8$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.203	0.409	0.305	1.013	0.307	0.310	1.013	-0.197	0.697	0.246
RMS	0.040	0.045	0.041	0.144	0.043	0.054	0.145	0.035	0.036	0.041
SE	0.042	0.047	0.043	0.155	0.044	0.057	0.157	0.037	0.037	0.043
SD	0.040	0.044	0.041	0.143	0.043	0.054	0.145	0.035	0.035	0.041
R	0.946	0.939	0.942	0.925	0.968	0.944	0.921	0.950	0.954	0.963

Table 5.21: Simulation result for NOBS=1500 NRES=900 $\tau_{\boldsymbol{u}} = 0.9$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.203	0.408	0.305	1.011	0.306	0.309	1.014	-0.198	0.698	0.246
RMS	0.039	0.043	0.040	0.139	0.041	0.052	0.140	0.034	0.034	0.040
SE	0.041	0.045	0.041	0.146	0.042	0.054	0.148	0.036	0.035	0.041
SD	0.039	0.042	0.039	0.139	0.041	0.052	0.139	0.034	0.034	0.039
R	0.952	0.944	0.960	0.950	0.979	0.961	0.939	0.951	0.955	0.964

Table 5.22: Simulation result for NOBS=1000 NRES=600 $\tau_{\boldsymbol{u}} = 0.7$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.209	0.417	0.310	1.019	0.315	0.315	1.022	-0.190	0.694	0.249
RMS	0.054	0.060	0.053	0.196	0.057	0.072	0.203	0.047	0.049	0.054
SE	0.060	0.072	0.061	0.217	0.063	0.081	0.223	0.051	0.051	0.062
SD	0.053	0.058	0.052	0.195	0.055	0.071	0.202	0.045	0.049	0.053
R	0.885	0.803	0.856	0.897	0.867	0.871	0.907	0.899	0.948	0.851

Table 5.23: Simulation result for NOBS=1000 NRES=600 $\tau_u = 0.8$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.208	0.414	0.309	1.018	0.314	0.314	1.018	-0.191	0.695	0.247
RMS	0.052	0.057	0.050	0.187	0.054	0.069	0.188	0.044	0.047	0.051
SE	0.057	0.068	0.058	0.205	0.060	0.077	0.209	0.048	0.049	0.059
SD	0.051	0.056	0.050	0.186	0.052	0.068	0.188	0.043	0.046	0.051
R	0.897	0.815	0.856	0.907	0.867	0.888	0.898	0.894	0.948	0.865

Table 5.24: Simulation result for NOBS=1000 NRES=600 $\tau_u = 0.9$

	$\hat{\mu}_R$	$\hat{\mu}_T$	$\hat{\sigma}_{PR}$	$\hat{\sigma}^2_R$	$\hat{\sigma}_{PT}$	$\hat{\sigma}_{RT}$	$\hat{\sigma}^2_T$	$\hat{\beta}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\psi}^*$
True value	0.2	0.4	0.3	1	0.3	0.3	1	-0.2	0.7	0.24
Mean	0.207	0.411	0.308	1.020	0.312	0.314	1.018	-0.193	0.696	0.246
RMS	0.050	0.054	0.048	0.179	0.052	0.066	0.181	0.042	0.044	0.048
SE	0.054	0.064	0.055	0.191	0.056	0.072	0.193	0.046	0.046	0.055
SD	0.049	0.053	0.048	0.178	0.050	0.065	0.180	0.042	0.044	0.048
R	0.911	0.838	0.873	0.933	0.901	0.908	0.933	0.899	0.964	0.878

Table 5.25: Non-inferiority test for different τ_u values

τ_u	NOBS	NRES	ECP
0.7	2000	1200	0.942
0.8	2000	1200	0.946
0.9	2000	1200	0.944
0.7	1500	900	0.95
0.8	1500	900	0.955
0.9	1500	900	0.948
0.7	1000	600	0.961
0.8	1000	600	0.967
0.9	1000	600	0.961

From Tables 5.16 to 5.24, we see that the mean values of the estimates are very close to the true values in all situations, the RMS and the mean of standard errors are reasonably small. The values of the ratio R are all between 0.8 and 1.2. Therefore, the simulation results indicate that the proposed estimate procedure in general can provide the users with reliable parameter estimates and standard errors.

We summarize the simulation results with respect to the different NOBS, NRES and τ_u values in the following graphs, in which MIS7, MIS8, and MIS9 stand for the $\tau_u = 0.7$, $\tau_u = 0.8$, and $\tau_u = 0.9$ respectively. As before, the legend of the following graph is given in Figure 5.1. We can see for all of the unknown parameters, the accuracy of the estimates will be improved if the NOBS, NRES or the τ_u value increases, as evident by the fact that the RMS becomes smaller and the ratio $R = SD/SE$ becomes closer to 1.

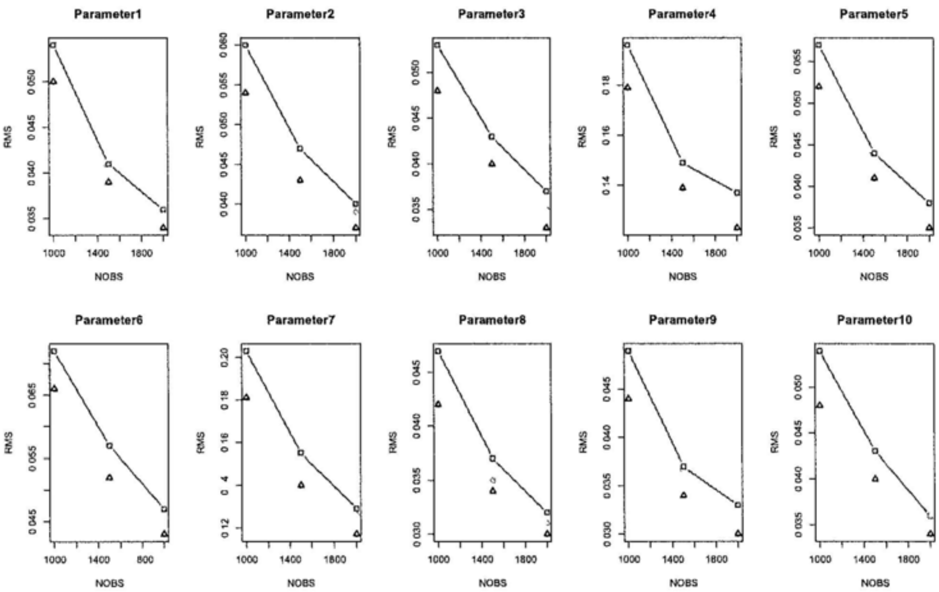

Figure 5 4 RMS-NOBS block design with partially validated data

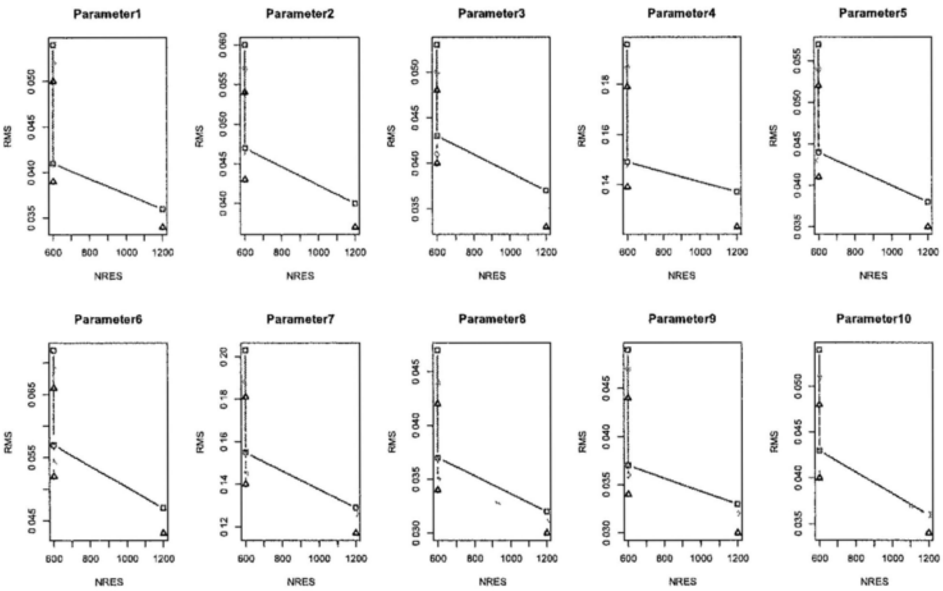

Figure 5.5: RMS-NRES block design with partially validated data

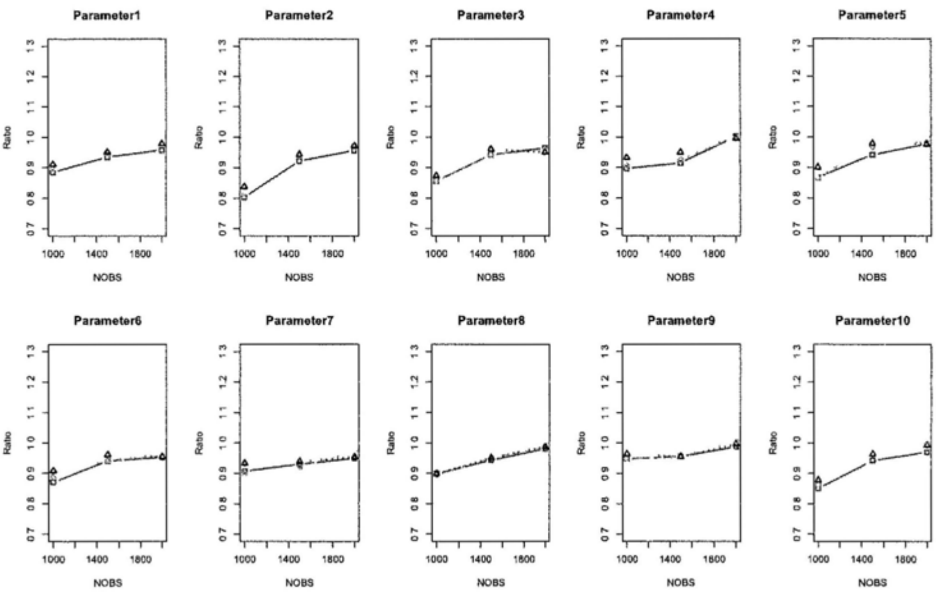

Figure 5 6: Ratio-NOBS block design with partially validated data

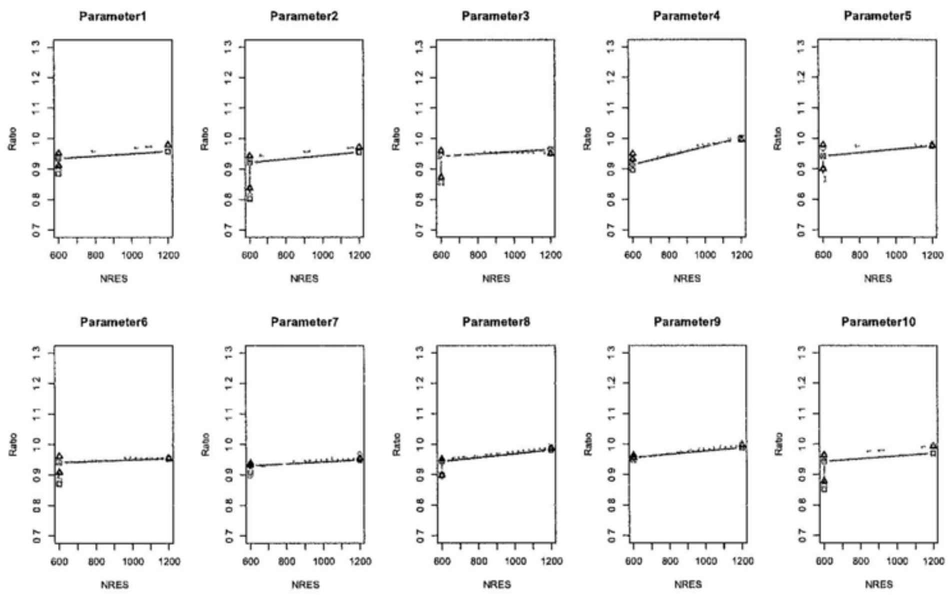

Figure 5.7: Ratio-NRES block design with partially validated data

5.4.3 Example 8

In this section, we illustrate how the proposed method can be used to analyze real data. We used the HELP data set with 871 (NOBS) effective respondents. By using the responses to the questions in Group a as the discriminator, there are 679 (NRES) subjects who have responses from both the true device and the fallible device. As independent learning is the most intuitive and convenient path out of the four choices, it is chosen as the placebo group (X_P). We are interested in detecting whether the peer discussion (X_T) is non-inferior to the discussion with course lecturers (X_R).

Based on the previous analysis in Chapter 3, we have established that independent learning is non-inferior to discussing with course lecturers. The estimate $\hat{\mu}_R$ is greater than $\hat{\mu}_P$ and it is a common sense that the lecturers can provide more professional explanation in concept clarification than students, so it suggests that the prerequisite $\mu_R - \mu_P > 0$ in three-armed design is likely to be true. Under these specifications with $f = -\frac{1}{5}$, the HELP data set was analyzed by using the three-armed design model with partially validated data.

The estimates of the parameters are given by $\hat{\mu}_R = 0.138$, $\hat{\mu}_T = -0.207$, $\hat{\sigma}_{PR} = 0.196$, $\hat{\sigma}_R^2 = 1.312$, $\hat{\sigma}_{PT} = 0.420$, $\hat{\sigma}_{RT} = 0.156$, $\hat{\sigma}_T^2 = 0.9415$, $\hat{\beta}_2 = -0.423$, $\hat{\beta}_3 - \hat{\beta}_2 = 1.124$, and $\hat{\psi}^* = -0.317$. The 95% CI for ψ^* is $(-0.476, +\infty)$. The negative lower bound of the CI means that discussing with peers is not non-inferior to discussing with teachers.

Chapter 6

Conclusion

We have developed statistical procedures for conducting non-inferiority tests in two-armed and three-armed designs with ordinal categorical outcomes. We used a latent variable normal model and used the maximum likelihood approach to get the parameter estimates and bootstrap method to obtain the standard errors of the parameter estimates. The proposed procedure can be widely used not only in clinical and pharmaceutical fields (see Metzler & Haung (1983) and Tang & Poon (2007)), but also in many other fields, such as sensory and consumer field in Bi (2005).

As misclassification is frequently encountered in contingency tables, we have also developed non-inferiority testing procedures for ordinal categorical data with misclassifications. Two methods have been developed to analyze data sets with misclassification. The first assumes that the probabilities of misclassification are known, and the other assumes that information on misclassification is available in a data set that is obtained from partially validated data.

It is also worth noting that many research studies use a closely related surrogate variable to replace a primary variable of interest when the primary variable is difficult to measure, as explained in Poon & Wang (2010). We can consider the classification based on a surrogate variable to be fallible and the classification based on the primary variable of interest to be true, and the procedure that has been developed for analyzing partially validated data can then be applied. We have illustrated the application of the proposed method by using the HELP data set in Examples 5 and 8.

In HELP data set, it is worthy of note that many students had not discussed with teaching assistant in the past six months, so we have not used the relevant variables (a2, b2 and c2) in any of our analysis. We have attached the questionnaire in Appendix.

In many research studies, we may have different surrogate variables with different cost requirements. Besides, we have already illustrated that the higher honest rates may lead to more reliable results. The selection of appropriate surrogate variables from a cost-effectiveness perspective is an interesting topic for further research. Another interesting research topic is to explore the optimal allocation in the design of collecting partially validated data. That is, determining how many respondents should be classified by the two classifiers so as to achieve optimal result in non-inferiority testing, given the different costs of the two classifiers.

In effect, the power and sample size determination is also a hot topic in three-armed design. Pigeot et al.(2003) have calculated the sample size that is required to achieve a given power assuming normality and homogeneity of variances in the three independent treatments when the data are observed in a continuous scale. The sample size requirement in the context of ordinal categorical data is an interesting topic for further research. In this thesis, we use Wald-type CIs to test non-inferiority, and other methods can also be applied to develop CIs, such that CIs based on score-test. The comparison of the performance of the Wald-type intervals and the score-type intervals is an interesting topic in future studies.

The nature of non-inferiority test and equivalence test are the same. Following the notation that we have been used, testing of equivalence involves the testing of the following hypothesis:

$$\begin{aligned} &H_0: \mu_T - \mu_R \leq -\Delta \;\text{ or }\; \mu_T - \mu_R \geq \Delta \\ &H_1: -\Delta < \mu_T - \mu_R < \Delta. \end{aligned} \tag{6.1}$$

Following the argument of Blackwelder (1998), equivalence can be concluded when the $100(1-2\alpha)\%$ confidence interval of $\mu_T - \mu_R$ entirely falls within the predefined endpoints $(-\Delta, \Delta)$ given significance level α. The method proposed in this thesis can be employed to conduct equivalence tests with minor modification.

Appendix

Questionnaire on How Effective the Learning Paths are

a) Communication frequencies

In the past 6 months, how often have you used the following learning paths to make better understanding of a concept out of the time you spend in lecture?

Statements	Times				
	Never	1-3	4-6	7-9	> 9
a1) Discussing with course lecturers	1	2	3	4	5
a2) Discussing with teaching assistants	1	2	3	4	5
a3) Discussing with peers	1	2	3	4	5
a4) Independent learning	1	2	3	4	5

b) To solve difficulties in understanding a concept after class, how much do you agree or disagree with the following statements?

Statements	Highly Disagree → Highly Agree				
b1) Discussing with course lecturers can make confusing concepts clear	1	2	3	4	5
b2) Discussing with teaching assistants can make confusing concepts clear	1	2	3	4	5
b3) Peer discussions can make confusing concepts clear	1	2	3	4	5
b4) Independent learning can make confusing concepts clear	1	2	3	4	5

c) Considering the last time that you conducted the following learning approaches to make better understanding of a concept after class, how much do you agree or disagree with the following statements? Please tick your answer.

Statements	Highly Disagree → Highly Agree					NOT Applicable
c1) Discussing with course lecturers made confusing concepts clear last time	1	2	3	4	5	NA
c2) Discussing with teaching assistants made confusing concepts clear last time	1	2	3	4	5	NA
c3) Peer discussions made confusing concepts clear last time	1	2	3	4	5	NA
c4) Independent learning made confusing concepts clear last time	1	2	3	4	5	NA

Reference

Bi, J. (2005). Similarity testing in sensory and consumer research. *Food Quality and Preference*, 16, 139-149.

Blackwelder, W.C. (1998). *Encyclopedia of biostatistics.* 2, 1367-1372.

D'Agostino, R., Massaro, J., & Sullivan, L. (2003). Non-inferiority trails: design concepts and issues - the encounters of academic consultants in statistics. *Statistics in Medicine*, 22, 169-186.

Ezzet, F., & Whitehead, J. (1993). A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine.*10, 901-907.

Hung, H.M.J., Wang, S.J., & Robert, O. (2007). Issues with statistical risks for testing methods in noninferiority trial without a placebo arm. *Journal of Biopharmaceutical Statistics*, 17, 201-213.

Kenward, M.G., & Jones, B. (1991). The analysis of categorical data from cross-over trials using a latent variable model. *Statistics in Medicine*, 10, 1607-1619.

Koch G.G., & Tangen C.M. (1999). Nonparametric analysis of covariance and its role in noninferiority clinical trials. *Drug Information Journal*, 33, 1145-1159.

Lui, K.J., & Cumberland, W.G. (2001). A Test Procedure of Equivalence in Ordinal Data with Matched-Pairs, *Biometrical Journal*, 43, 977-983.

Liu, J., Hsueh, H., Hsieh, E., & Chen, J. (2002). Tests for equivalence or non-inferiority for paired binary data. *Statistics in Medicine*, 21, 231-245.

Metzler, C.M., & Haung, D.C. (1983). Statistical methods for bioavailability and bioequivalence. *Clinical Research and Practices & Drug Regulatory Affairs*, 1, 109-132.

Munk, A., Mielke, M., Freitag, G., & Skipka G. (2007). Testing noninferiority in three-armed clinical trials based on the likelihood ratio statistics, *Canadian Journal of Statistics*, 35, 413C431.

Neale, M.C., Boker, S.M., Xie, G., & Maes, H.H. (1999). Mx: *Statistical Modeling*. Box 710 MCV, Richmond, VA23298: Department of Psychiatry. http://www.vcu.edu/mx/

Ng, T.H. (2001). Choice of delta in equivalence testing. *Drug Information Journal*, 35, 1517-1527.

Pigeot, I., Schafer, J., Rohmel, J., & Hauschke, D. (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, 22, 883-899.

Poon, W.Y. (2004). A latent normal distribution model for analysing ordinal responses with applications in meta-analysis, *Statistics in Medicine*, 23, 2155-2172.

Poon, W.Y., & Wang, H.B. (2010). Analysis of ordinal categorical data with misclassification. *British Journal of Mathematical and Statistical Psychology*, 63, 17-42.

Stuart, A. (1955). A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification. *Biometrika*, 42, 412-416.

Schuirmann, D.J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15, 657-680.

Tang, M.L., Ling, M.H., Ling, L., & Tian, G. (2010). Confidence intervals for a difference between proportions based on paired data, *Statistics in Medicine*, 29, 86-96.

Tang, M.L., & Poon W.Y. (2007). Statistical inference for equivalence trials with ordinal responses: A latent normal distribution approach, *Computational Statistics & Data Analysis*, 51, 5918-5926.

Tang, M.L., & Tang, N.S. (2004). Tests of noninferiority via rate difference for three-arm clinical trails with placebo. *Journal* *of Biopharmaceutical Statistics*, 14, 337-347.

Yiu C.F., & Poon, W.Y. (2008). Estimating the polychoric correlation from misclassified data. *British Journal of Mathematical and Statistical Psychology*, 61, 49-74.

Zhang, X.M. (2007). Analysis of Multivariate Ordinal Categorical Variables with Misclassified Data (Master's thesis). The Chinese University of Hong Kong, Hong Kong.

Zhou, X.H., Obuchowski, N.A., & McClish, D.K. (2002). Statistical Methods in Diagnostic Medicine. Wiley: New York.