METHODS FOR META-ANALYSIS OF DIAGNOSTIC TEST ACCURACY STUDIES

CONTRIBUTIONS TO STATISTICAL METHODS FOR META-ANALYSIS OF DIAGNOSTIC TEST ACCURACY STUDIES

BY

ZELALEM FIRISA NEGERI

A THESIS

SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2019) McMaster University

(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE:	Contributions to statistical methods for meta-analysis of diagnostic test accuracy studies
AUTHOR:	Zelalem Firisa Negeri M.Sc., (Statistics) — McMaster University, Canada M.Sc., (Statistics) — Addis Ababa University, Ethiopia B.Sc., (Statistics) — Addis Ababa University, Ethiopia
SUPERVISOR:	Professor Joseph Beyene
NUMBER OF PAGES:	xxv, 165

I dedicate this thesis to the memory of:

my dad, Firisa Negeri, and

my brother, Amayu Firisa Negeri.

Lay Abstract

Diagnostic tests vary from the noninvasive rapid strep test used to identify whether a patient has a bacterial sore throat to the much complex and invasive biopsy test used to examine the presence, cause, and extent of a severe condition, say cancer. Meta-analysis is a widely used statistical method that synthesizes evidence from several studies. In this thesis, we develop novel statistical methods extending the traditional methods for meta-analysis of diagnostic test accuracy studies. Our proposed methods address the issue of modelling asymmetrical data, identifying outlier studies, and optimally accommodating these outlying studies in a meta-analysis of diagnostic test accuracy studies. Using both real-life and simulated datasets, we show that our proposed methods perform better than conventional methods in a wide range of scenarios.

Abstract

Meta-analysis is a popular statistical method that synthesizes evidence from multiple studies. Conventionally, both the hierarchical and bivariate models for meta-analysis of diagnostic test accuracy (DTA) studies assume that the random-effects follow the bivariate normal distribution. However, this assumption is restrictive, and inferences could be misleading when it is violated. On the other hand, subjective methods such as inspection of forest plots are used to identify outlying studies in a meta-analysis of DTA studies. Moreover, inferences made using the well-established bivariate random-effects models, when outlying or influential studies are present, may lead to misleading conclusions. Thus, the aim of this thesis is to address these issues by introducing alternative and robust statistical methods. First, we extend the current bivariate linear mixed model (LMM) by assuming a flexible bivariate skew-normal distribution for the random-effects. The marginal distribution of the proposed model is analytically derived so that parameter estimation can be performed using standard likelihood methods. Overall, the proposed model performs better in terms of confidence interval width of the overall sensitivity and specificity, and with regards to bias and root mean squared error of the between-study (co)variances

than the traditional bivariate LMM. Second, we propose objective methods based on solid statistical reasoning for identifying outlying and/or influential studies in a meta-analysis of DTA studies. The performances of the proposed methods are evaluated using a simulation study. The proposed methods outperform and avoid the subjectivity of the currently used ad hoc approaches. Finally, we develop a new robust bivariate random-effects model which accommodates outlying and influential observations and leads to a robust statistical inference by down-weighting the effect of outlying and influential studies. The proposed model produces robust point estimates of sensitivity and specificity compared to the standard models, and also generates a similar point and interval estimates of sensitivity and specificity as the standard models in the absence of outlying or influential studies.

Contents

Acknowledgements

Above all, I praise the creator of heaven and earth – the Almighty God, the source of my hope, strength and perseverance – for being with me at all times and for all the blessings in my life.

I am very grateful to my supervisor, Professor Joseph Beyene, for his unreserved guidance and mentoring throughout this Thesis. Your hands-off style of supervision allowed me to grow both personally and professionally and led me to the many discoveries all through my graduate studies. I am very thankful for that. It has been a pleasure working with you, and I am sure that our collaboration will continue for many years.

Special thanks to my Ph.D. supervisory committee members Dr. Roman Viveros, and Dr. Gregory Pond for meticulously reading my thesis and providing me with their insightful and constructive feedback.

I am indebted to the Department of Mathematics and Statistics, and the school of graduate studies, McMaster University, for giving me the resources to make my dream a reality and providing me with financial support throughout my graduate studies. Moreover, I want to thank the staff at the Department of Mathematics and

Statistics of McMaster University: July, Diana, Paula, Sheree and Kenneth for their unreserved assistance during my studies.

I want to express my deepest appreciation to the family and friends I have had met in Canada for their invaluable support and encouragement while pursuing my studies: Obbo Zelalem, Ayelech and their family; Obbo Desalegn, Zewditu and their family; Abdissa, Pam, Shawn and Pastor Loretta; Ray, Melody and their family. These people have been my mentors and teachers outside of academics. I am lucky to have been surrounded by the many supportive classmates and friends at McMaster: Ashley, Taddele, Regina, Joycelyn, Mateen, Binod, Satish, Sayantee, Alexander, Lorena, Neela, Kavitha, William, Bethi and others.

Finally, yet most importantly, this work would not have been realized without the unconditional love, encouragement and understanding of my mommy Alemitu, my wife Hawi, and my sisters Kena, Mignot, Yirgalem and Rahel.

List of Abbreviations

AIC	Akaike information criteria
AUDIT	Alcohol use disorder identification test
AUDIT-C	Alcohol use disorder identification test-consumption
BBN	Bivariate binomial-normal
BIC	Bayesian information criteria
BMSOM	Bivariate mean-shift outlier model
BNL	Bivariate normal-Laplace
BNN	Bivariate normal-normal
BNSN	Bivariate normal-skew-normal
BSN	Bivariate skew-normal
cdf	Cumulative distribution function
CI	Confidence interval
CP	Coverage probability

DTA	Diagnostic test accuracy
EM	Expectation-maximization
EUS	Endoscopic ultrasonography
FN	False negatives
FP	False positives
GLMM	Generalized linear mixed-effects model
HSROC	Hierarchical summary receiver operating characteristic
IPD	Individual patient data
LMM	Linear mixed-effects model
LRT	Likelihood ratio test
MA	Meta-analysis
MC	Monte Carlo
MCEM	Monte Carlo expectation-maximization
ML	Maximum likelihood
MLE	Maximum likelihood estimates
MMSE	Mini-mental state examination
MH	Metropolis-Hastings
NEE	Naked-eye examination

NL	Normal-Laplace
pdf	Probability density function
RE	Random-effects
RMSE	Root mean squared error
Se	Sensitivity
SN	Skew-normal
Sp	Specificity
SROC	Summary receiver operating characteristic
TN	True negatives
TP	True positives
US	Ultrasonography

Declaration of Academic Achievement

Throughout this thesis, I was the principal investigator for developing the research topics, reviewing the literature, designing and developing statistical methods under the supervision of Dr. Joseph Beyene. Additionally, I was responsible for running the study every day, analyzing the data and assembling the results. I also prepared research reports and manuscripts for each project and was responsible for disseminating my research findings in a variety of forms.

Preface

Three manuscripts, which were prepared between September 2015 and July 2019, and are either published or under review in different peer-reviewed journals, constitute this sandwich thesis. The three independent but coherent articles are outlined in Chapters 2, 3, and 4 of the thesis, respectively. The introduction of the thesis given in Chapter 1 and the summary and conclusion of the thesis in Chapter 5 provide the overarching context of the thesis and a precise summary of the contributions, respectively. My supervisor, Dr. Joseph Beyene has co-authored all three articles constituting this thesis. We describe how I, Zelalem Firisa Negeri, contributed to each of the three manuscripts and to the entire thesis in the following paragraphs.

In Chapter 2, we present manuscript 1, which was submitted to and is currently under review for the journal *Biometrical*. This article proposed a novel bivariate random-effects model by introducing a new random-effects distribution for meta-analysis of diagnostic test accuracy (DTA) studies. I am the first author of the paper and was responsible for developing the research question and study idea, doing an extensive literature review, developing and deriving the statistical method, running both the simulation and real data analyses and writing up of the manuscript. Both

Dr. Joseph Beyene and I were responsible for the approval and submission of the final version of the manuscript to the journal.

Chapter 3 is the second article, which has been published in the journal *Statistical Methods in Medical Research*. This manuscript proposed several new and robust statistical methods for identifying outlying and/or influential studies in meta-analysis of DTA studies. I am the first author of the article and was responsible for developing the research question and study idea, doing an extensive literature review, deriving and developing the statistical methods, running both the simulation and real data analyses, writing up of the manuscript, and drafting responses to the anonymous reviewers. Both Dr. Joseph Beyene and I were responsible for the approval and submission of the final version of the manuscript to the journal.

The third manuscript, presented in Chapter 4, has been submitted to and is currently under review for the journal *Statistical Methods in Medical Research*. This manuscript proposed a novel and robust bivariate random-effects model for dealing with outlying and/or influential observations in a meta-analysis of DTA studies. I am the first author of the manuscript and was responsible for developing the research question and study idea, doing an extensive literature review, deriving and developing the statistical methods, running both the simulation and real data analyses, and writing up of the manuscript. Both Dr. Joseph Beyene and I were responsible for the approval and submission of the final version of the manuscript to the journal.

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Meta-analysis, a statistical procedure for combining evidence from several multiple studies that aim to address the same research question (Borenstein et al., 2009), is an active area of research that dates back to the early 20th century when Pearson (1904) combined correlation coefficients to study the effectiveness of vaccination in reducing typhoid fever and mortality among the British army (Shannon, 2016; Chalmers and Hedges, 2002). Moreover, Hedges and Olkin (1985) discussed the momentum meta-analysis had gained in the 1930s when it was used to synthesize data from agricultural research. Some of the earlier works in the 1930s include the synthesis of p-values by Fisher (1932) and Pearson (1933), and the aggregation of effect sizes with the objective of quantifying the effect of agricultural treatments by Cochran (1937). However, the popularity of meta-analysis has increased in the past three decades

as it has been widely used in several fields of study including medicine, psychology, education, ecology and pharmacy (Borenstein et al., 2009).

Conducting meta-analysis by synthesizing evidence from several related studies has many statistical as well as clinical advantages over the study-specific summary reported by the individual studies. Whereas the original studies are able to summarize the property of an effect size separately for each study using hypothesis testing and p-values, the results are not generalizable to the population beyond that studied by each study due to the variation in each study because of the characteristics of their respective study-population, difference in the quality of research laboratories and expertise of personnel, geographical variation, and the difference in time at which the studies are conducted. Moreover, due to the relatively small sample size they employ, the individual studies may suffer from lack of statistical power to detect small effect sizes and hence may incorrectly declare those effect sizes statistically insignificant. In contrast, when consistent studies are combined and average effect size and its corresponding statistical significance at the population level are reported in a meta-analysis, the results are generalizable to the general population from which each of the separate studies is taken from. Additionally, both the variation in results due to within- and between-study variability are taken into account, quantified and explained if necessary. Furthermore, due to the combined evidence and increment in sample size, the statistical power of a meta-analysis to detect small effect sizes increases too (Borenstein et al., 2009; Pepe, 2010). As a result of the above-mentioned reasons, meta-analysis is considered an important statistical procedure and plays a vital role as an evidence-based decision-making tool in medical and many other fields

of study.

In medical sciences, meta-analysis can be applied to synthesize effect sizes from two or more interventions using the conventional meta-analysis of intervention studies or network meta-analysis of intervention studies to study the efficacy of new treatments or interventions. For example, pharmaceutical companies might be interested in determining the effectiveness of a new drug against a placebo by conducting a conventional meta-analysis of intervention studies or by carrying out network meta-analysis if the goal is to compare the efficacy of the new drug against more than one competing drugs. Furthermore, meta-analysis can be used to combine test characteristics from several studies to quantify and summarize the overall diagnostic ability of a diagnostic test of interest. For instance, evidence from several studies can be synthesized using meta-analysis to assess the diagnostic ability of a newly discovered diagnostic test compared to the gold standard test or another reference test before the new diagnostic test is deemed usable to diagnose patients.

A meta-analysis of diagnostic test accuracy (DTA) studies is a relatively new area of research compared to the conventional meta-analysis of interventions (Schwarzer et al., 2015). Since the publication of the first papers on the topic in the 1990s (Willis and Quigley, 2011), researchers have been actively working on the methodological advancement of methods for the meta-analysis of DTA studies. Among the models, the hierarchical (Rutter and Gatsonis, 2001) and bivariate (Reitsma et al., 2005; Chu and Cole, 2006) random-effects models are recommended to combine diagnostic test characteristics, primarily test sensitivity and test specificity, across studies. Whereas other test characteristics such as the diagnostic odds ratio, predictive values and

likelihood ratios could also be meta-analyzed to quantify the accuracy of diagnostic tests, test sensitivity and specificity are generally combined since they preserve the binary nature of the data (Schwarzer et al., 2015) and also have clinically appealing interpretations. Test sensitivity (Se) quantifies the ability of a diagnostic test to correctly identify subjects with a condition (e.g. a disease) among subjects who truly have the condition; whereas, test specificity (Sp) measures the ability of a diagnostic test to correctly identify individuals without the condition among those who truly do not have the condition of interest.

Although the hierarchical and bivariate random-effects models were shown to be equivalent when there are no covariates (Harbord et al., 2007), the bivariate random-effects models are most widely used in the literature due to their ease of interpretation and availability in standard statistical software. The bivariate random-effects models themselves differ by design as the Reitsma et al. (2005) model is based on the linear mixed-effects model (LMM) approach while that of Chu and Cole (2006) follows the generalized linear mixed-effects model (GLMM) approach. What distinguishes the two models is the assumption they make about how the within-study variation in sensitivity and specificity is modelled. Specifically, the LMM-based model assumes the bivariate normal distribution after logit-transforming the study-specific sensitivity and specificity to describe the within-study variation, while the GLMM-based model accommodates the within-study variation by modelling the observed positive test results using two independent binomial distributions. However, both the LMM-based and GLMM-based models commonly assume the bivariate normal distribution to describe the unexplained variation beyond that predicted by

the within-study variation. Therefore, based on their distributional assumption, we name the LMM-based and GLMM-based models as bivariate normal-normal (BNN) and bivariate binomial-normal (BBN) model, respectively, throughout this thesis. The mathematical description of the two standard models, the BNN and BBN, is given in detail in Chapter 2, Subsection 2.4.1 – 2.4.2 and in Chapter 4, Subsection 4.4.1 – 4.4.2.

Although the standard bivariate random-effects models are well-defined and widely used in practice, we point out some of the gaps that the methods have in the following paragraphs. First, whereas the assumption of bivariate normal distribution for explaining the within-study variation in sensitivity and specificity can be supported by invoking the Central Limit Theorem due to the large within-study sample sizes, it is harder to rationalize the same assumption for describing the random-effects since the number of studies in a meta-analysis are generally small, unless the same distribution is assumed for ease of computation and mathematical tractability. Accordingly, unless this restrictive distributional assumption for the random-effects is met, the inference using those standard methods may lead to misleading conclusions when the distribution of the random-effects deviates from the traditionally assumed bivariate normal distribution (Lee and Thompson, 2008; Baker and Jackson, 2008; Beath, 2014; Baker and Jackson, 2016). To this date, although researchers have attempted to relax the restrictive distributional assumption for the random-effects in the context of the conventional meta-analysis of interventions, no study pursued the same objective in the context of meta-analysis of DTA studies.

Second, the current approaches for detecting outlying and influential studies in

meta-analysis of DTA studies are based on ad hoc approaches such as investigation of different plots including the forest plot, bivariate box plot, and scatter plots of observed sensitivity and specificity (Devillé et al., 2002; Doria et al., 2006; Petignat et al., 2007; Singal et al., 2009; Zhou et al., 2016; Pormohammad et al., 2017; de Jesus et al., 2009; Kriston et al., 2008). Although these plots are a good starting point, they are subjective and have some drawbacks. Since the current tools are completely based on the observed pairs of sensitivity and specificity, they do not take into account the inevitable within-study variation and between-study heterogeneity which defines the weights that would, in turn, define whether a study is outlying or not in a given meta-analysis. Moreover, the current methods are not able to tell whether an outlying study is actually influential and therefore would impact the meta-analysis results. To our knowledge, there are no objective methods in the meta-analysis of DTA studies literature to overcome the limitations of the current approaches of identifying outlying and influential studies, although there are attempts in the conventional meta-analysis of interventions (Viechtbauer and Cheung, 2010; Gumedze and Jackson, 2011; Hedges and Olkin, 1985).

Third, after outlying and influential studies are identified in a meta-analysis of DTA studies, the question of how to deal with such studies becomes important. Applying the current standard methods, the BNN and BBN models, in the presence of outlying and influential studies may result in wrong inferences and misleading conclusions since both methods assume the bivariate normal distribution for the

random-effects. On the other hand, we might lose important information if we exclude the outlying and influential studies from the meta-analysis. Thus, robust statistical methods that would accommodate outlying and influential studies and result in robust inference would be needed. Whereas such methods have been proposed in the context of the conventional meta-analysis of intervention studies (Baker and Jackson, 2008; Lee and Thompson, 2008; Viechtbauer and Cheung, 2010; Gumedze and Jackson, 2011; Beath, 2014; Baker and Jackson, 2016), to the best of our knowledge, we did not find any in the situation of meta-analysis of DTA studies.

1.2 Objectives of the Thesis

Based on the above-mentioned three problems with the current standard methods for meta-analysis of DTA studies, this thesis aims at addressing the following three objectives:

1. Developing a flexible bivariate random-effects model and examining whether meta-analysis of DTA studies inferences are affected or not in the presence of skewness in DTA datasets.

2. Developing statistical methods based on objective measures for identifying outlying and/or influential studies in a meta-analysis of DTA studies and examining the impact of those outlying and influential studies.

3. Developing a robust bivariate random-effects model to optimally accommodate outlying and influential studies in a meta-analysis of diagnostic test accuracy studies.

To achieve the first objective, we propose an alternative random-effects distribution – the bivariate skew-normal distribution and propose a new random-effects model – the bivariate normal-skew-normal (BNSN) model as discussed in detail in Subsection 2.4.3 of Chapter 2.

As discussed in Subsection 3.4 of Chapter 3, we accomplished the second objective by proposing two statistical approaches that are based on objective measures namely, the residual-based approach and the likelihood ratio test-based approach for identifying outlying studies, and by proposing two indexes, the SIGMARATIO and VARCOVRATIO for identifying influential studies.

In Chapter 4, Subsection 4.4.3, we achieve the third objective by proposing the bivariate Laplace distribution for the random-effects and thus developing a new bivariate normal-Laplace (BNL) random-effects model that accommodates and down-weights the impact of outlying and influential studies in a meta-analysis of DTA studies.

1.3 Scope of the Thesis

This thesis was prepared in a 'Sandwich Thesis' style where three independent but related manuscripts, which focused on the topic of meta-analysis of DTA studies, are sandwiched between an introductory chapter and a concluding chapter.

In Chapter 2, we propose a new bivariate random-effects model by suggesting a bivariate skew-normal distribution for modelling the random-effects in a meta-analysis of DTA studies. Similar to the BNN model, we assume the bivariate normal

distribution for describing the within-study variation in sensitivity and specificity. Therefore, we name our proposed model the bivariate normal-skew-normal (BNSN) model. The marginal distribution of the proposed model is analytically derived, and thus the BNSN model has a closed-form likelihood function, which is maximized to obtain the maximum likelihood estimates (MLE) of the parameters of the BNSN model. To study whether the presence of skewness affects the meta-analysis results and conclusions, an extensive simulation study was carried out, and a real-life data example is also used to illustrate the methods. Since the estimation of the skewness parameter is not robust to the choice of starting values, a parametric bootstrap was employed to test the significance of the skewness parameter of the proposed model. The overall sensitivity and specificity, the between-study variances and covariance estimates returned by the proposed model and the two standard models, the BNN and the BBN, are assessed to examine the effect of introducing skewness in a meta-analysis of DTA studies.

Chapter 3 is concerned with the objective of proposing solid statistical methods for identifying outlying and influential studies in a meta-analysis of DTA studies. The proposed methods are motivated by two published meta-analyses that consist of varying numbers of potentially outlying studies. The proposed methods for identifying outlying studies are the residual-based approach, which is based on the standardized residual obtained from the BNN model, and, the likelihood ratio test-based approach, which is based on the likelihood ratio test derived from a newly proposed bivariate random-effects mean-shift outlier model (BMSOM). A ratio of the between-study covariance matrix when each study is removed from the meta-analysis to when

all the studies are included in the meta-analysis (SIGMARATIO), and the ratio of the overall variance-covariance matrix when each study is removed from the meta-analysis to when all the studies are included in the meta-analysis (VARCOVRATIO) are used to detect whether the outlying studies are influential or not. A simulation study and real data are used to validate and illustrate the proposed methods.

As a follow-up to the findings in Chapter 3, we ask the question of 'how to deal with outlying and influential studies in a meta-analysis of DTA studies' the focus of Chapter 4. As such, we introduce a new bivariate random-effects model by assuming the bivariate Laplace distribution for the random-effects and the bivariate normal-distribution for modelling the within-study variation in sensitivity and specificity. We call our new model the bivariate normal-Laplace (BNL) random-effects model. Although the marginal model of our proposed BNL model was derived analytically, it does not have a closed-form density function and likelihood function. Therefore, numerical optimization procedures, specifically the Monte Carlo expectation-maximization (MCEM) algorithm is used to obtain the MLE of the parameters of our proposed model. Both simulation study and real data are used to appraise the performance of the proposed method. The simulation study is designed such that outlying and influential studies are introduced both in Se and Sp, only in Se, only in Sp and neither in Se nor in Sp to demonstrate how robust the proposed model would be in these scenarios.

It is worth noting that throughout the thesis we aim at comparing the performance of the proposed methods and the standard methods using the overall Se and Sp estimates, their respective confidence intervals and the estimated between-study

covariance matrix. Additionally, we would like to note that this thesis does not cover possible extensions of the proposed models such as the development of the summary receiver operating characteristic (SROC) curve (Moses et al., 1993; Walter, 2002; Dukic and Gatsonis, 2003; Gatsonis and Paliwal, 2006) or the meta-regression (Bashore et al., 1989; Berlin and Antman, 1992; Van Houwelingen et al., 2002).

Chapter 2

Skew-normal random-effects model for meta-analysis of diagnostic test accuracy (DTA) studies

2.1 Abstract

Hierarchical models are recommended for meta-analyzing diagnostic test accuracy (DTA) studies. The bivariate random-effects model is currently widely used to synthesize a pair of test sensitivity and specificity using logit transformation across studies. This model assumes a bivariate normal distribution for the random-effects. However, this assumption is restrictive and can be violated. When the assumption fails, inferences could be misleading. In this paper, we extend the current bivariate random-effects model by assuming a flexible bivariate skew-normal distribution for

the random effects in order to robustly model logit sensitivities and logit specificities. The marginal distribution of the proposed model is analytically derived so that parameter estimation can be performed using standard likelihood methods. The method of weighted-average is adopted to estimate the overall logit-transformed sensitivity and specificity. An extensive simulation study is carried out to investigate the performance of the proposed model compared to other standard models. Overall, the proposed model performs better in terms of confidence interval width of the average logit-transformed sensitivity and specificity compared to the standard bivariate linear mixed model and bivariate generalized linear mixed model. Simulations also show that the proposed model performs better than the well-established bivariate linear mixed model in terms of bias and comparable with regards to the RMSE of the between-study (co)variances. The proposed method is also illustrated using previously published meta-analysis study.

Keywords: meta-analysis; diagnostic test accuracy; skew-normal random-effects; sensitivity; specificity

2.2 Introduction

Meta-analysis (MA) is a method for statistical synthesis of evidence from several studies that aim to address the same research question (Borenstein et al., 2009; Kovalchik, 2013). Meta-analysis has been widely used to synthesize diagnostic test

accuracy (DTA) studies and make inference about the accuracy of one or more diagnostic tests. To model test sensitivity (Se) and specificity (Sp) of diagnostic tests, researchers suggest a random-effects model that incorporates the potential correlation between Se and Sp and the between-study variation.

Reitsma et al. (2005) developed a bivariate random-effects model to jointly meta-analyze Se and Sp by assuming that the random effects follow a bivariate normal distribution. This method also models the study-specific variability in Se and Sp by an approximate (asymptotic) bivariate normal distribution. We refer to this method as the bivariate normal-normal (BNN) model. Since the BNN model is easy to understand and simple to apply, it has been widely applied in the literature and was implemented in the `mada` package of the 'R' programming language (Doebler, 2017).

Chu and Cole (2006) proposed a generalized bivariate random-effects model, in a generalized linear mixed-model framework, to meta-analyze a pair of Se and Sp from each study. Chu and Cole's bivariate random-effects model has the advantage of modelling the observed within-study variability exactly using the binomial distribution instead of approximating it with the normal distribution. The model is then completed by assuming a bivariate normal distribution for the random effects. We call this method the bivariate binomial-normal (BBN) model. The BBN model avoids the addition of the ad hoc continuity correction to any DTA data cell containing zero counts (Chu and Cole, 2006). The BBN method has recently been implemented in the `Metatron` package of the 'R' programming language (Huang, 2018).

Using simulation studies and real data, Hamza et al. (2008) demonstrated that the BBN model is superior to that of BNN method. They reported that the BNN

model produces biased estimates of the pooled accuracy measures, Se and Sp, and coverage probabilities below the nominal 95%, particularly when there are few studies in the meta-analysis and when the true Se and Sp are closer to one (Hamza et al., 2008).

Despite the above-mentioned advantages over the BNN model, the BBN model has not been widely applied in the literature (Hamza et al., 2008), possibly because of the need to numerically approximate the likelihood function. That is, similar to all GLMM models, the BBN model does not have a closed-form likelihood function since the marginal model is obtained by integrating out random-effects and the integral has no closed-form solution. This necessitates approximation of the likelihood function which can be computationally intensive.

Both the BNN and BBN assume that the random-effects follow a bivariate normal distribution. Since the within-study sample sizes in a meta-analysis are typically large, the Central Limit Theorem could justify the approximate normality assumption to model the within-study variability. However, the assumption of normal random-effects can be unrealistic for meta-analysis models (Lee and Thompson, 2008; Baker and Jackson, 2008; Beath, 2014; Baker and Jackson, 2016).

Furthermore, a normal approximation to the proportions (Se and Sp) deteriorates, even when logit transformation is employed, for large values of Se and Sp since the distribution of the logit-transformed Se and Sp are skewed for values of Se and Sp closer to 1 (Doebler et al., 2012). Doebler et al. (2012) proposed a parametric transformation known as t_α family of transformations, which includes the logit transformation as a special case, to enhance the rationale of the distributional assumption

of the BNN model. Although the t_α transformation could yield plausible empirical results, it has a number of practical and theoretical shortcomings. First, when $\alpha = 0$ or $\alpha = 2$, the resulting summary receiver operating characteristic curve (SROC) takes values outside of its range—the unit square $[0, 1]^2$ (Doebler et al., 2012). Secondly, since the t_α transformation requires the estimation of two more parameters than the logit transformation, over-parameterization occurs with few studies in the MA.

In the conventional univariate meta-analysis, there have been studies that acknowledge the need for more flexible random-effects distribution. Lee and Thompson (2008) proposed the t, skew-t and skew-normal distributions. Baker and Jackson (2008) recommended several heavy-tailed distributions including the t and Beta for the random effects to accommodate outliers in meta-analysis. Similarly, a finite mixture of normal distributions was introduced by Beath (2014), to accommodate outliers by down-weighting their effects in meta-analysis. Most recently, Baker and Jackson (2016) developed new skewed marginal distributions which are mathematically tractable and avoid the need for numerical integration of the likelihood functions in their earlier work.

To the best of our knowledge, there are no current studies that have examined flexible random effect distributions in a meta-analysis of DTA studies, although as discussed above, the distribution of the logit-transformed accuracy measures is highly skewed for large values of Se and Sp. Therefore, this paper aims to address this gap by proposing a bivariate skew-normal (BSN) distribution, which includes the bivariate normal distribution as a special case, as a distribution of the random effects

in a meta-analysis of DTA studies. We model the within-study variation using the (approximate) bivariate normal distribution as in the BNN model. Therefore, we see our new method as an extension of the BNN model, and thus we call it the bivariate normal-skew-normal (BNSN) model. We also propose a weighted-average estimator of the pooled (overall) logit(Se) and logit(Sp) and its corresponding standard error for our new method.

The remainder of this chapter is organized as follows. In Section 2.3, we motivate our method using published meta-analysis data. The three bivariate random-effects models, the BNN, BBN, and BNSN are discussed in Section 2.4. We provide the simulation, and real data results in Section 2.5, and 2.6, respectively. We finally present a summary and discussion in Section 2.7.

2.3 Motivating example

In this section, we describe a published meta-analysis dataset to motivate the methods. This particular dataset is chosen to demonstrate the performance of the methods when a model's distributional assumption is not met. We examined the bivariate normality of the logit transformed Se and Sp using the Henze-Zirkler's test (Henze and Zirkler, 1990) as it is recommended by Mecklin and Mundfrom (2005) as a powerful method compared to other methods. The method has been implemented in the R package `MVN` (Korkmaz et al. (2014)). We also present the normal Q-Q plot of the logit-transformed Se and Sp to motivate the bivariate approach.

2.3.1 The endoscopic ultrasonography (EUS) data

The *EUS* dataset (Table 2.1), which is based on the *EUS* test for the preoperative diagnosis of primary gastric cancer, consists of 44 studies and is obtained from the Cochrane review of Mocellin and Pasquali (2015). According to Henze-Zirkler's multivariate normality test, logit-transformed Se and Sp of the *EUS* data are not distributed as bivariate normal (p-value < 0.001). Shapiro-Wilk's univariate normality test suggests that logit(Sp) is responsible for the rejection of joint normality (p-value < 0.001) which is also apparent from the Q-Q plot presented in Figure 2.1 — which reveals that the distribution of logit(Sp) has a departure from normal distribution and logit(Se) is approximately normally distributed.

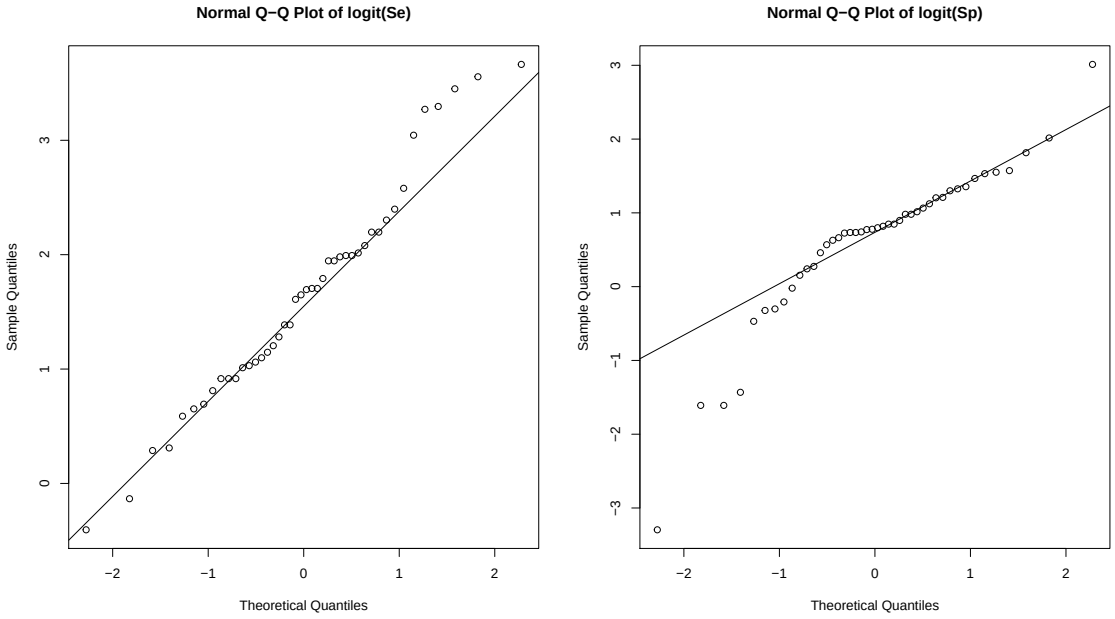

Figure 2.1: Q-Q plot of logit Se (left) and logit Sp (right) for the *EUS* dataset.

Table 2.1: The *EUS* dataset of Mocellin and Pasquali (2015)

Study number	Author	Year	TP	FP	FN	TN
1	Ahn	2009	63	5	2	1
2	Akahoshi	1998	36	5	4	1
3	Ang	2006	20	12	6	19
4	Arocena	2006	4	3	2	8
5	Barbour	2007	85	25	27	69
6	Bentrem	2007	71	27	37	83
7	Bhandari	2004	27	6	1	14
8	Blackshaw	2008	9	4	1	30
9	Bohle	2011	18	9	5	30
10	Botet	1991	10	8	1	31
..	....	..	..	..	..	..
38	Tsendsuren	2006	17	14	0	10
39	Tseng	2000	30	10	5	29
40	Wang	1998	33	17	12	57
41	Willis	2000	52	17	10	37
42	Xi	2003	14	6	5	7
43	Zheng	2011	45	49	20	48
44	Ziegler	1993	44	18	6	40

However, according to the statistical tests and Q-Q plot presented in Figure 2.1, the inference based on the bivariate normal marginal model of BNN model, which is a consequence of the assumption of bivariate normal random-effects, may not be reasonable and could be misleading. In Section 2.6, we present the meta-analysis results when the three methods discussed in this paper are applied to the *EUS* dataset.

2.4 Methods

In this section, we present three bivariate random-effects models that can be used to synthesize a pair of test sensitivity (Se) and test specificity (Sp) obtained from a set of k independent DTA studies. Let $\mathbf{Y}_i = (y_{1i}, y_{2i})^T = (\text{logit}(\widehat{Se}_i), \text{logit}(\widehat{Sp}_i))^T$ be a vector of observed logit sensitivities and logit specificities for each study, i=1,...,k. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote a location and scale parameters of a bivariate distribution, respectively.

2.4.1 The bivariate normal-normal (BNN) model

The BNN model can be hierarchically defined as follows. In the first level, the BNN model assumes that the observed vector of response $\mathbf{Y}_i$ is modelled according to

$$\mathbf{Y}_i|\mathbf{b}_i \sim N_2(\boldsymbol{\mu} + \mathbf{b}_i, \boldsymbol{\Psi}_i), i = 1, ..., k \tag{2.1}$$

where $\boldsymbol{\mu}$ is the fixed-effect denoting overall mean logit(Se) and logit(Sp), $\mathbf{b}_i$ is the random-effects denoting the study-specific true logit(Se) and logit(Sp) and

$$\boldsymbol{\Psi}_i = \begin{pmatrix} \psi_{1i}^2 & 0 \\ 0 & \psi_{2i}^2 \end{pmatrix}, i = 1, ..., k,$$

represents the study-specific within-study covariance matrix of $\mathbf{Y}_i$ and is typically assumed to be known.

In the second level, the BNN model assumes that the random-effects, $\mathbf{b}_i$, varies

across studies and models this variability as

$$\mathbf{b}_i \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}), i = 1, ..., k, \tag{2.2}$$

where $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ is the between-study covariance matrix. Note that when there are no random-effects, model (2.1) reduces to the usual fixed-effects meta-analysis model.

As in any hierarchical model, maximum likelihood estimation of parameters of the BNN model is also based on the marginal model. The marginal model obtained from (2.1) and (2.2) by integrating out the random-effects is

$$\mathbf{Y}_i \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i), i = 1, ..., k, \tag{2.3}$$

where $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} + \boldsymbol{\Psi}_i$.

The likelihood and log-likelihood function of the BNN model follows from (2.3) and is given in (2.4) and (2.5), respectively.

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{y}) = \prod_{i=1}^{k} \frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu})^T {\boldsymbol{\Sigma}_i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\} \tag{2.4}$$

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{y}) = -k\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k} \log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{k}(\mathbf{y}_i - \boldsymbol{\mu})^T {\boldsymbol{\Sigma}_i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) \tag{2.5}$$

From (2.5) it follows that the maximum likelihood (ML) estimator of $\boldsymbol{\Sigma}$ does not have a closed-form solution, therefore, ML estimators of $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ are obtained

iteratively using numerical methods.

Once the parameters are estimated, inference regarding the parameter of interest, $\boldsymbol{\mu}$, which stands for the population average, is made by first deriving its variance (or standard error). From (2.5), the solution to the score function with respect to $\boldsymbol{\mu}$ is

$$\boldsymbol{\mu} = \left(\sum_{i=1}^{k} \mathbf{W}_i\right)^{-1} \sum_{i=1}^{k} \mathbf{W}_i \mathbf{Y}_i \tag{2.6}$$

$$\text{where} \quad \mathbf{W}_i = (\boldsymbol{\Sigma} + \boldsymbol{\Psi}_i)^{-1} = \boldsymbol{\Sigma}_i^{-1}. \tag{2.7}$$

Therefore, the estimated variance-covariance matrix of the pooled logit(Se) and logit(Sp) estimator, $\boldsymbol{\mu}$, can easily be obtained from (2.6) as

$$\widehat{\text{var}(\boldsymbol{\mu})} = \left(\sum_{i=1}^{k} \widehat{\mathbf{W}}_i\right)^{-1} = \left(\sum_{i=1}^{k} \widehat{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1}. \tag{2.8}$$

where $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i$ and $\widehat{\boldsymbol{\Sigma}}$ is the maximum likelihood estimate of $\boldsymbol{\Sigma}$ obtained by maximizing (2.5). Finally, the desired $(1-\alpha)100\%$ Wald-type confidence interval for the unknown pooled logit(Se), (μ_1), and pooled logit(Sp), (μ_2), can be obtained by combining results (2.6)–(2.8).

2.4.2 The bivariate binomial-normal (BBN) model

The BBN model assumes the binomial distribution for modelling the within-study variability and the bivariate normal distribution for modelling the between-study

variability. That is, the BBN assumes

$$\begin{aligned} TP_i|b_{1i} &\sim \text{Binomial}(n_{1i}, Se_i); y_{1i} = \mu_1 + b_{1i}; \\ TN_i|b_{2i} &\sim \text{Binomial}(n_{2i}, Sp_i); y_{2i} = \mu_2 + b_{2i}; \\ \mathbf{b}_i &\sim N_2(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned}$$

where TP_i and TN_i are the respective number of true positives and true negatives for i-th study; n_{1i} and n_{2i} are the total number of diseased and non-diseased individuals, respectively, for i-th study; $y_{1i}, y_{2i}, b_{1i}, b_{2i}, \mu_1, \mu_2$ are such that $\mathbf{y}_i = (y_{1i}, y_{2i})^T, \mathbf{b}_i = (b_{1i}, b_{2i})^T$, and $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$.

The marginal likelihood function of the BBN model, given in (2.9), has no closed-form expression since the integral does not have a closed-form solution. The BBN can be fitted using the `SAS PROC NLMIXED` algorithm or using the R package `Metatron` (Huang, 2018) where in both programs the adaptive Gaussian quadrature algorithm (Chu and Cole, 2006; Paul et al., 2010) is used to approximate the likelihood numerically. We note that both the BNN and BBN models require five parameters, $\mu_1, \mu_2, \sigma_1^2, \sigma_{12}$, and σ_2^2 to be estimated.

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{y}) = \int_{\mathbb{R}^2} \prod_{i=1}^{k} f_{\mathbf{y}_i|\mathbf{b}_i}(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\mu}) f_{\mathbf{b}_i}(\mathbf{b}_i|\boldsymbol{\Sigma}) d\mathbf{b}_i \tag{2.9}$$

2.4.3 The proposed model

In this section, we discuss the proposed bivariate random-effects model. Before discussing our model, we present a review of the multivariate skew-normal distribution.

2.4.3.1 Review of the multivariate skew-normal distribution

There have been several versions of the multivariate skew-normal (SN) distribution (Azzalini and Valle, 1996; Arnold et al., 2002; Arellano-Valle et al., 2005) discussed in the literature. For its theoretical convenience and mathematical tractability, we discuss the multivariate SN of Arellano-Valle et al. (2005). A p–dimensional random vector $\mathbf{X}$ is said to have a multivariate SN distribution with a location vector $\boldsymbol{\mu} \in \mathbb{R}^p$, a $p \times p$ positive-definite dispersion matrix $\boldsymbol{\Sigma}$ and skewness vector $\boldsymbol{\alpha} \in \mathbb{R}^p$, denoted $\mathbf{X} \sim \text{SN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\alpha})$, if its probability density function (pdf) is given by

$$f(\mathbf{x}) = 2\phi_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\alpha}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu})) \tag{2.10}$$

where $\phi_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the pdf of the p–dimensional normal distribution and $\Phi_1(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution. It is easy to see that when the skewness parameter, $\boldsymbol{\alpha} = \mathbf{0}$ the pdf of $\mathbf{X}$ in (2.10) reduces to the pdf of the p–dimensional normal distribution, $\phi_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Arellano-Valle et al. (2005) denotes the standardized p–dimensional SN distribution by $\mathbf{W} \sim \text{SN}_p(\boldsymbol{\alpha})$, which is obtained from (2.10) when $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$. A useful stochastic representation is presented in Proposition 1 of Arellano-Valle et al. (2005). That is, if $T_0 \sim$N(0,1) and independently $\mathbf{T}_1 \sim \text{N}_p(\mathbf{0}, \mathbf{I}_p)$, then

$$\boldsymbol{\delta}|T_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^T)^{\frac{1}{2}}\mathbf{T}_1 \stackrel{d}{=} \mathbf{W} \tag{2.11}$$

$$\text{where } \boldsymbol{\delta} = \frac{\boldsymbol{\alpha}}{\sqrt{1 + \boldsymbol{\alpha}^T\boldsymbol{\alpha}}} \tag{2.12}$$

where $\stackrel{d}{=}$ is read as "has the same distribution as".

Finally, (2.10) can be obtained from (2.11)–(2.12) as shown in Corollary 1 of Arellano-Valle et al. (2005) using a linear transformation. That is, if $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{W}$, where $\mathbf{W}$ is as defined in (2.11), then $\mathbf{X} \sim \text{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$. Furthermore, the mean and variance of $\mathbf{X}$ is

$$\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta} \tag{2.13}$$

$$\text{Var}(\mathbf{X}) = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{\frac{1}{2}} \tag{2.14}$$

We note that the mean vector and the covariance matrix given in (2.13) and (2.14) will reduce to the mean vector and covariance matrix of the multivariate normal random vector when the skewness parameter, $\boldsymbol{\alpha}$, is set equal to $\mathbf{0}$. Moreover, the location parameter can be re-parameterized so that $\boldsymbol{\mu}$ becomes the mean of the multivariate SN distribution. That is, if $\mathbf{X} \sim \text{SN}_p(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$, then

$$\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$$

$$\text{Var}(\mathbf{X}) = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{\frac{1}{2}}$$

where $\boldsymbol{\eta} = \boldsymbol{\mu} - \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}$.

2.4.3.2 The bivariate normal-skew-normal (BNSN) model

In our proposed BNSN model, we generalize the BNN model by relaxing the restrictive distributional assumption for the random-effects. Particularly, we assume that

the within-study variability can be modelled using the bivariate normal distribution as in the BNN model (2.1), but the between-study variability should be modelled using the bivariate skew-normal distribution for the reasons discussed in Section 2.2. Therefore, we define our BNSN model hierarchically as

$$\mathbf{Y}_i|\mathbf{b}_i \sim N_2(\boldsymbol{\eta} + \mathbf{b}_i, \boldsymbol{\Psi}_i), i = 1, ..., k; \tag{2.15}$$

$$\mathbf{b}_i \sim SN_2(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}), i = 1, ..., k. \tag{2.16}$$

where $\mathbf{Y}_i$, $\mathbf{b}_i$, $\boldsymbol{\eta}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}_i$ are as defined previously.

Since our BNSN model is purely parametric, we need to find the marginal model to obtain the ML estimates of the parameters and to make inference using the ML estimates. The marginal model obtained from the hierarchical representation of the BNSN model (2.15)–(2.16), by integrating out the random-effects is derived in the **Appendix I** and given as

$$\mathbf{Y}_i \sim SN_2(\boldsymbol{\eta}, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i) \tag{2.17}$$

where

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} + \boldsymbol{\Psi}_i;\ \ \boldsymbol{\omega}_i = \frac{\boldsymbol{\Sigma}_i^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}}{\sqrt{1 + \boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}}};\ \ \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\alpha};\ \text{ and } \boldsymbol{\Gamma}_i = \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Psi}_i^{-1}.$$

We note that our BNSN random-effects model has some key features that could make it practically more applicable. Firstly, our BNSN model reduces to the standard BNN model when the skewness parameter is set to zero. That is, when $\boldsymbol{\alpha} = \mathbf{0}$;

$\mathbf{Y}_i \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$. Secondly, since the BNN model is nested within the BNSN model, the significance of the skewness parameter can be assessed using the well-known likelihood ratio test (LRT) (Azzalini and Capitanio, 1999).

2.4.3.3 Maximum likelihood estimation of the parameters of the BNSN model

As given in (2.17), the marginal distribution of our BNSN random-effects model is

$$\mathbf{Y}_i \sim SN_2(\boldsymbol{\eta}, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i),$$

and has a pdf,

$$f_{\mathbf{Y}_i}(\mathbf{y}_i) = 2\phi_2(\mathbf{y}_i|\boldsymbol{\eta}, \boldsymbol{\Sigma}_i)\Phi_1(\boldsymbol{\omega}_i^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\eta}))$$

Thus, the likelihood function of the observed data, $\mathbf{y}_i = (\mathbf{y}_{1i}, \mathbf{y}_{2i})^T$, is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}) = \prod_{i=1}^{k} \frac{2}{2\pi|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\eta})^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\eta})\}\Phi_1(\boldsymbol{\omega}_i^T\boldsymbol{\Sigma}_i^{-\frac{1}{2}}(\mathbf{y}_i - \boldsymbol{\eta})),$$

and the log-likelihood function is

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}) = -k\log(\pi) - \frac{1}{2}\sum_{i=1}^{k}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{k}(\mathbf{y}_i - \boldsymbol{\eta})^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\eta}) + \sum_{i=1}^{k}\log(\Phi_1(\boldsymbol{\omega}_i^T\boldsymbol{\Sigma}_i^{-\frac{1}{2}}(\mathbf{y}_i - \boldsymbol{\eta}))) \quad (2.18)$$

Because the log-likelihood function (2.18) is well-behaved, it is suitable for obtaining ML estimates of the parameters using numerical methods – since there is no closed-form solution to the score equations. In this paper, we employed the Nelder-Mead algorithm (Nelder and Mead, 1965) which has been implemented in the *optim()* function of the R software. Maximum likelihood estimates of the five parameters of the BNN model and a two-dimensional zero vector of the skewness parameter is used as starting values for obtaining the corresponding maximum likelihood estimates of the proposed BNSN model. If the estimated between-study covariance matrix of the BNN model is not positive definite, a $2\times$ 2 identity matrix is used as a starting value instead.

2.4.3.4 The pooled diagnostic accuracy estimator

It is evident that meta-analysis models are special cases of the general linear model but with heterogeneous error variances which are further assumed to be known and fixed (Viechtbauer et al., 2010). Therefore, a two step approach, where the between-study (co)variance is first estimated using one of the estimation methods (Veroniki et al., 2016), and then the pooled effect measure is estimated by the weighted least squares where the weights are inverses of sums of the between-and within-study variances, is adopted in meta-analytic models (Borenstein et al., 2009; Viechtbauer et al., 2010).

Therefore, following the two-step estimation approach, we first obtain MLEs of the between-study covariance matrix, $\boldsymbol{\Omega} = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}\boldsymbol{\delta}^{T}\boldsymbol{\Sigma}^{\frac{1}{2}}$, using the method discussed in Section 2.4.3.3. Then the unbiased weighted least squares estimator of the

pooled logit(Se) and logit(Sp) can be obtained as

$$\boldsymbol{\mu} = \left(\sum_{i=1}^{k} \mathbf{W}_i^* \right)^{-1} \sum_{i=1}^{k} \mathbf{W}_i^* \mathbf{Y}_i \tag{2.19}$$

where $\mathbf{W}_i^* = (\boldsymbol{\Omega} + \boldsymbol{\Psi}_i)^{-1} = \boldsymbol{\Omega}_i^{-1}$. Once the between-study covariance matrix is estimated using maximum likelihood method, estimate of the pooled logit(Se) and logit(Sp) is obtained as

$$\widehat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{k} \widehat{\mathbf{W}}_i^* \right)^{-1} \sum_{i=1}^{k} \widehat{\mathbf{W}}_i^* \mathbf{y}_i \tag{2.20}$$

where $\widehat{\mathbf{W}}_i^* = (\widehat{\boldsymbol{\Omega}} + \boldsymbol{\Psi}_i)^{-1} = \widehat{\boldsymbol{\Omega}}_i^{-1}$.

The covariance matrix of the pooled effect size estimator can easily be derived from (2.19) assuming the weights as known, although they are estimated in practice. Thus, the estimate of the variance-covariance matrix which we use to construct confidence intervals (CIs) for the unknown pooled effect size is

$$\text{vcov} = \text{var}(\widehat{\boldsymbol{\mu}}) = \left(\sum_{i=1}^{k} \widehat{\mathbf{W}}_i^* \right)^{-1} = \left(\sum_{i=1}^{k} \widehat{\boldsymbol{\Omega}}_i^{-1} \right)^{-1} \tag{2.21}$$

The (1-α)100% Wald-type CI for the unknown population effect size, $\boldsymbol{\mu}$, can be obtained by combining results (2.19)–(2.21). We note that when the skewness parameter, $\boldsymbol{\alpha} = \mathbf{0}$, the BNSN model's between-study covariance matrix, $\boldsymbol{\Omega}$, weights, $(\boldsymbol{\Omega}_i^{-1})$, and the estimated variance covariance matrix, $\left(\sum_{i=1}^{k} \widehat{\boldsymbol{\Omega}}_i^{-1} \right)^{-1}$, reduces to their corresponding BNN model's between-study covariance matrix, $\boldsymbol{\Sigma}$, weights, $\boldsymbol{\Sigma}_i^{-1}$,

and estimated variance covariance matrix $\left(\sum_{i=1}^{k} \boldsymbol{\Sigma}_i^{-1}\right)^{-1}$, respectively. Finally, the goodness-of-fit of the models can be assessed using the AIC (Akaike, 1974) or the likelihood ratio test (LRT).

2.5 Simulation study

2.5.1 Simulation design

An extensive simulation study is performed with the aim of comparing the models in terms of several empirical performance measures: bias, root mean squared error (RMSE) and the 95% coverage probability (CP). A total of 576 (when BNSN is the true model) and 144 (when BBN is the true model) scenarios are considered by varying the model parameters: $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ and data characteristics: number of studies (k) and number of patients in each health group (n). We analyzed 165 meta-analyses of DTA studies published in the Cochrane database in the years 2011-2015 to inform our simulation study. Based on the empirical study, the following ranges of parameter were used as the true parameters in the simulations.

1. The overall Se and Sp pair (Se, Sp): (0.75,0.70), (0.80,0.85) and (0.90,0.95).
2. The between-study variance pair (σ_1^2, σ_2^2): (0.72, 1.17) and (1.06, 1.40).
3. The between-study covariance (σ_{12}): -0.18, -0.59 and 0.35.
4. The between-study skewness parameter pair (α_1, α_2): (0.56,0.37), (-0.56,-0.37), (1.55, 2.06) and (-1.55, -2.06).

5. The number of patients in each health group (n): 40 and 100.

6. The number of studies in the meta-analyses (k): 10, 20, 50 and 100.

We generated data (study-specific pair of logit(Se) and logit(Sp)) from the bivariate normal distribution (True model: BBN model) and bivariate skew-normal distribution (True model: BNSN model) using the above-listed parameters and compared the performance of the three models discussed in the sequel. The following procedure has been used to generate MA of DTA data for our simulation study.

1. When the true model is the BNSN model, the study-specific logit(Se) and logit(Sp) pair are simulated from the bivariate skew-normal distribution by specifying its location ($\boldsymbol{\eta}$), dispersion ($\boldsymbol{\Sigma}$) and skewness ($\boldsymbol{\alpha}$) parameter using the above-mentioned true parameters. When the true model is the BBN model, the study-specific logit(Se) and logit(Sp) pair are simulated from the bivariate normal distribution by specifying its mean ($\boldsymbol{\mu}$) and covariance ($\boldsymbol{\Sigma}$) parameter using the above-mentioned true parameters.

2. The true study-specific Se and Sp are obtained by back-transforming the above-generated study-specific logit(Se) and logit(Sp) to the original unit scale [0,1].

3. The study-specific frequencies: TP, FN, TN and FP are generated from the binomial distribution using the proportions simulated in step 2 and the sample sizes (n) given in the previous list.

For each scenario, the simulation is repeated 1000 times and the empirical average bias, RMSE and CP are computed for the population average estimator ($\boldsymbol{\mu}$), and the

average bias and RMSE are computed for the estimator of the nuisance parameters σ_1^2, σ_2^2 and σ_{12}.

2.5.2 Simulation results

In this section, we present the performance of the proposed model compared to the two standard models when assessed in terms of bias, RMSE and CP.

Figure 2.2 shows the bias of the overall logit-transformed Se and Sp estimator, μ_1 and μ_2, for the three models when $\sigma_1^2 = 0.72$, $\sigma_2^2 = 1.17$, $\sigma_{12} = -0.59$, $(\alpha_1, \alpha_2) =$ (-1.55,-2.06) and BNSN is the true model. In most of the scenarios, regardless of the sample size and the number of studies, the BBN model marginally performed better than the BNN and BNSN model for small Se and Sp, and outperformed the BNN and BNSN model for large Se and Sp. The BNSN model yielded marginally smaller biases than the BNN model, regardless of the parameters varied in the simulation. The bias of the models consistently decreases as the sample size and number of studies increases. When data are generated from the bivariate normal distribution (Figure 2.3), the proposed model produced marginally less biased estimates of μ_1 and μ_2 to the BNN model, regardless of the parameters varied in the simulation. The BBN model outperformed the BNN and BNSN model regardless of the scenarios considered in the simulation study. The BNN and BNSN methods perform poorly when the true Se and Sp are large because both models, as the maximum likelihood method is employed, underestimate the covariance matrix, which leads to smaller inverse of sums of the weights (first expression in the right-hand side of (2.20)), which in turn results in underestimation of $\boldsymbol{\mu}$.

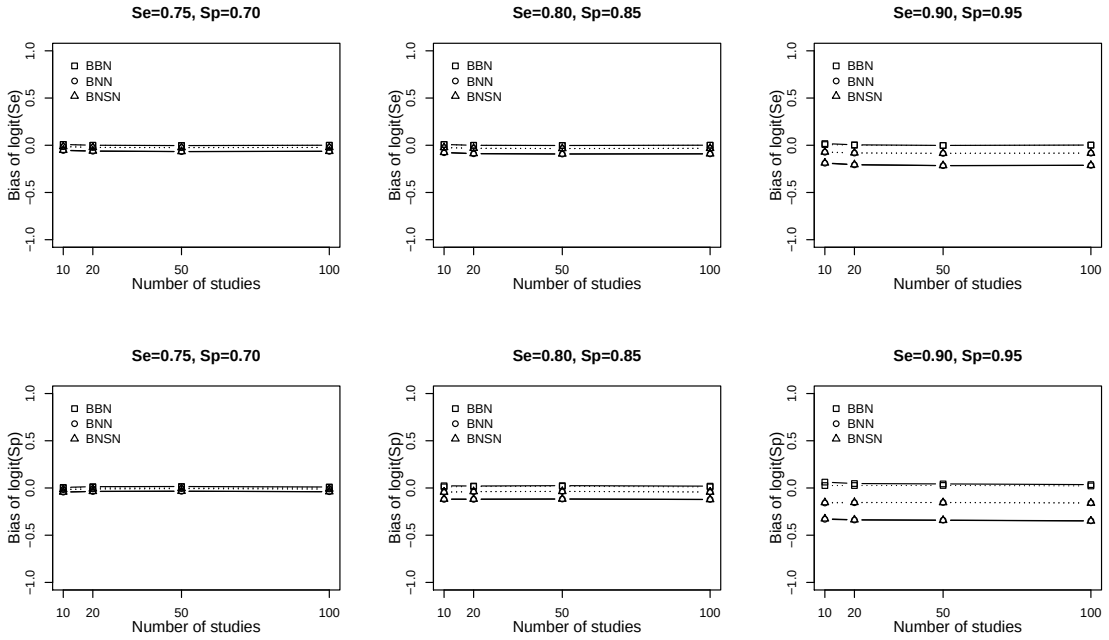

Figure 2.2: Bias of μ_1 and μ_2 for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,-0.59), (α_1,α_2) = (-1.55,-2.06) and BNSN is the true model.

In terms of RMSE of μ_1 and μ_2 estimator (Figure 2.4), the proposed BNSN model performs similarly to the standard models, regardless of the parameters considered in this simulation study. The proposed model yielded identical RMSE in μ_1 and μ_2 to the standard models when data are generated from the bivariate normal distribution (Figure 2.5), in spite of the number of studies varied in the simulation. The BBN model produced better RMSE than the BNN and BNSN model when the sample size is small, and Se and Sp are large and data are generated assuming the BBN model.

With regards to the 95% empirical CP (Figure 2.6), the three models performed

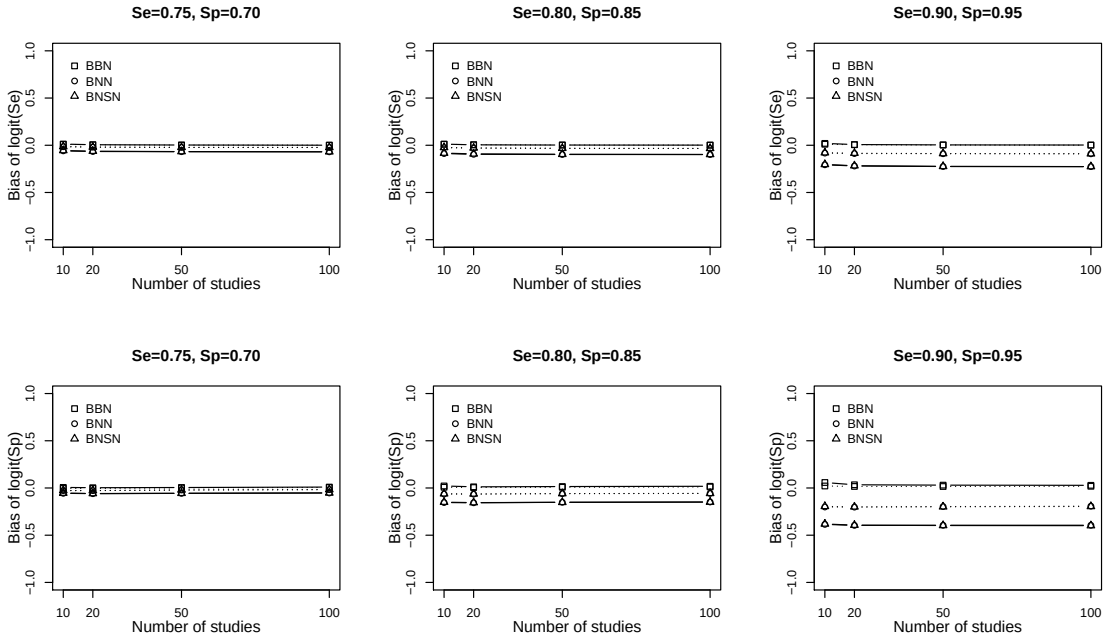

Figure 2.3: Bias of μ_1 and μ_2 for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,-0.59) and BBN is the true model.

similarly for small Se and Sp. The CP of the BNN and BNSN method diminished as Se and Sp increases, particularly when the sample size is small and Se and Sp are closer to 100%. As explained above, the between-study covariance matrix ($\mathbf{\Sigma}$) is underestimated by the BNN and BNSN model resulting in smaller overall covariance matrix (2.21) and standard errors, which will in turn yield narrower confidence intervals and thus smaller CPs. For this scenario, the CP gets worse as the number of studies increases because (2.21) becomes smaller as k increases. Overall, the BBN model yielded robust CP and the BNSN model resulted in marginally better

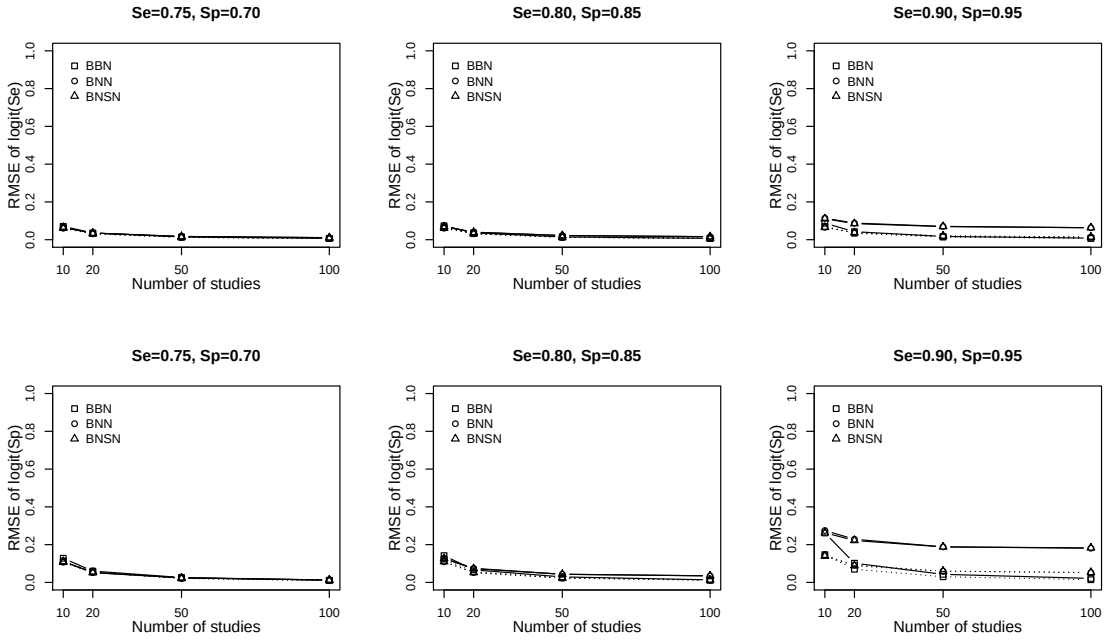

Figure 2.4: RMSE of μ_1 and μ_2 for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,0.35), (α_1,α_2) = (-0.56,-0.37) and BNSN is the true model.

CP than the BNN model, regardless of the simulation scenarios. Similar results had been observed when the true model is the BBN (Figure 2.7).

In terms of bias of the between-study variances σ_1^2 and σ_2^2 (Figure 2.8), all methods underestimated the true parameters in most of the scenarios. This result is expected since the maximum likelihood method was used to obtain parameter estimates. When Se and Sp are small, the proposed BNSN model yielded less biased estimates than the standard models particularly when the number of studies varies from small to medium. For the same scenarios but a large number of studies, the BNSN and BBN

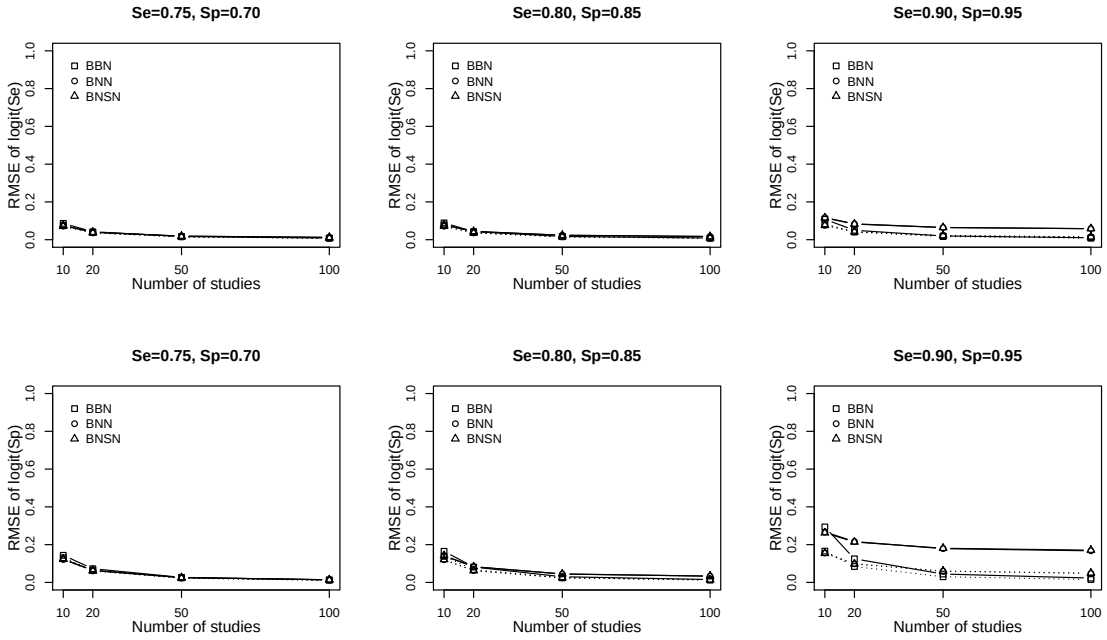

Figure 2.5: RMSE of μ_1 and μ_2 for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,-0.59) and BBN is the true model.

method produced comparable estimates. For large Se and Sp, the BNSN and BBN yielded similar bias when sample size and number of studies is large. Overall, the BNSN and BNN model resulted in the least and most biased estimates, respectively, regardless of the parameters varied in the simulation. When data are generated from the bivariate normal distribution, the BBN model yielded the smallest biased estimates followed by the BNSN model. Particularly, the BNSN and BBN model produced similar bias as the sample size and number of studies increases. Once again, the BNN model yielded the most biased estimates, regardless of the scenarios.

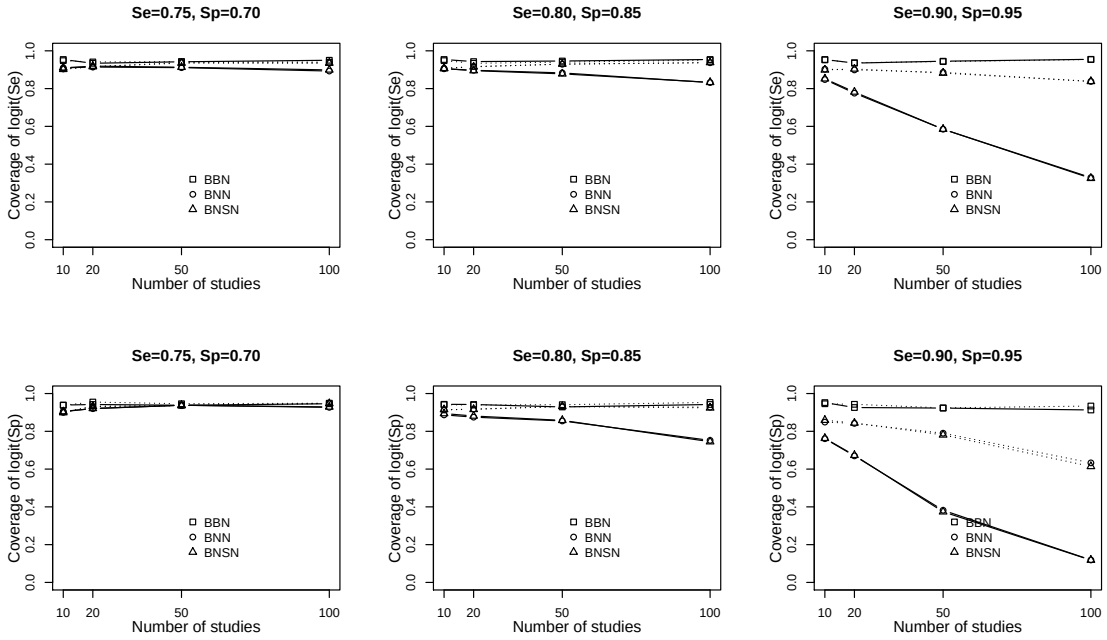

Figure 2.6: The 95% coverage probability of μ_1 and μ_2 for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,-0.59), (α_1,α_2) = (-1.55,-2.06) and BNSN is the true model.

Concerning RMSE of σ_1^2 and σ_2^2 (Figure 2.9), the BBN model consistently produced smaller RMSEs. The BNSN model yielded larger values when the number of studies and the sample size is small. RMSE of the BNSN model decreases as the number of studies and sample size increases. The BNN model resulted in larger RMSE in σ_2^2 when the sample size is small. When data are generated from the bivariate normal distribution, both BBN and BNN models yielded the smallest RMSE. The RMSE of the BNSN model is the largest regardless of the simulation scenarios.

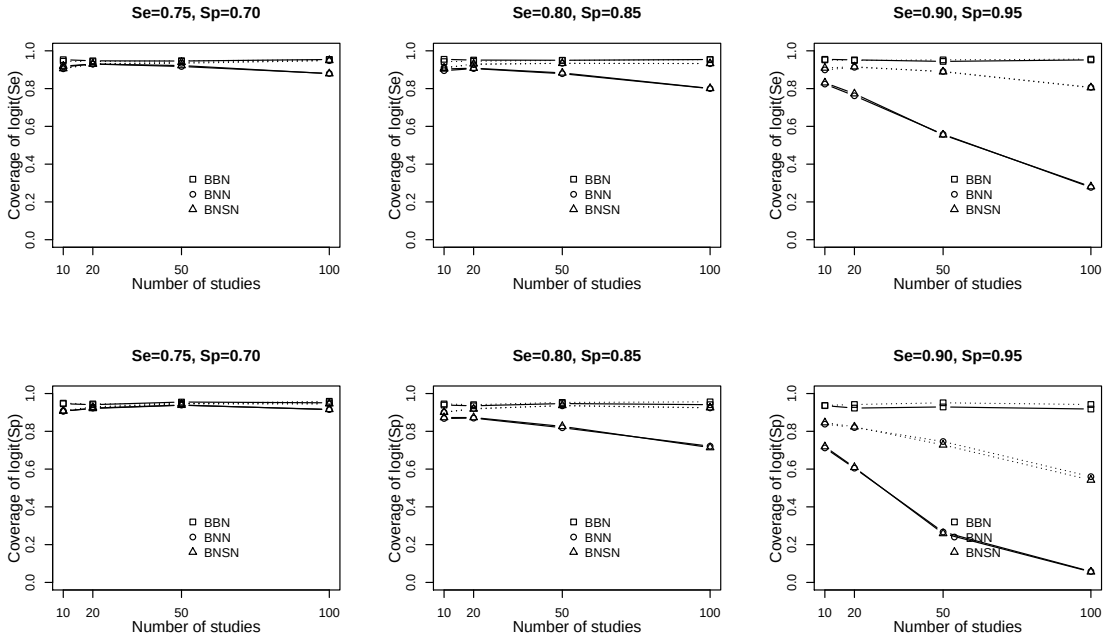

Figure 2.7: The 95% coverage probability of μ_1 and μ_2 for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,-0.59) and BBN is the true model.

Overall, the BBN model produced the smallest RMSE in σ_1^2 and σ_2^2 regardless of the simulation scenarios and assumed true model.

With regards to bias and RMSE (Figure 2.10) of the between-study covariance (σ_{12}), the three methods produced estimates with similar bias in most of the scenarios. The methods underestimated σ_{12} when the skewness parameter ($\boldsymbol{\alpha}$) is large. The RMSE of the three methods agrees as the sample size and number of studies increases. Overall, the BBN and BNN model consistently produced estimates with small bias and RMSE, respectively, regardless of the scenarios considered. Similar

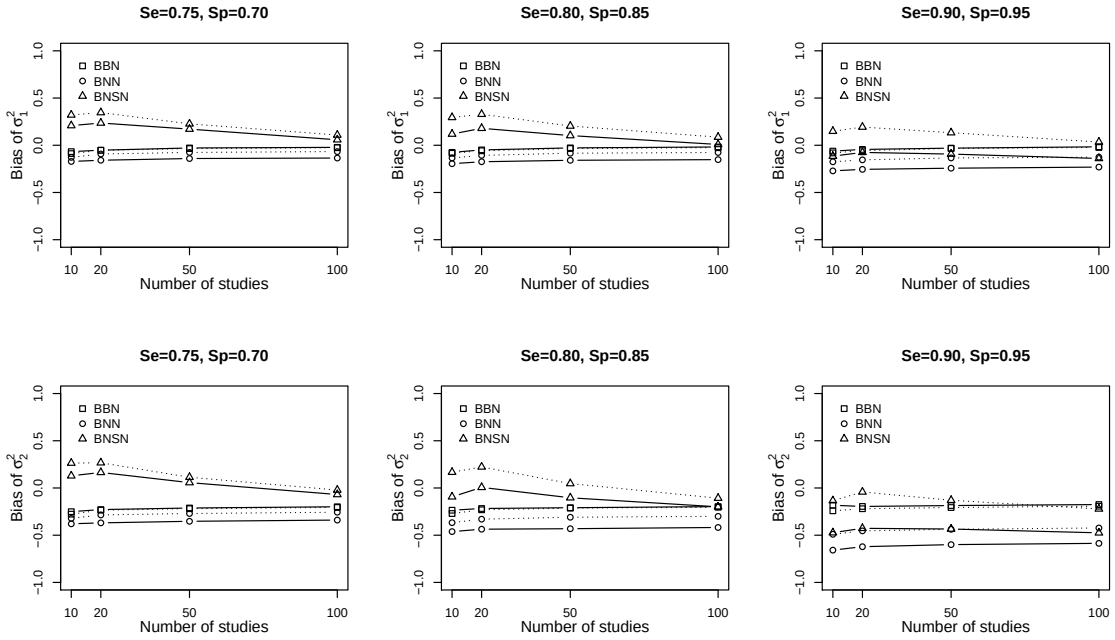

Figure 2.8: Bias of σ_1^2 and σ_2^2 for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,-0.59), (α_1,α_2) = (-1.55,-2.06) and BNSN is the true model.

results were observed when data were generated assuming the BBN model.

2.6 Illustrative example

Results of the three methods, when applied to the dataset introduced in Section 2.3, are presented below. The *EUS* test is one of the four diagnostic tests that Mocellin and Pasquali (2015) evaluated in their review. They fitted the BBN model to the data, and reported the pooled Se (95% CI) and pooled Sp (95% CI) as 83.0 (79.0,

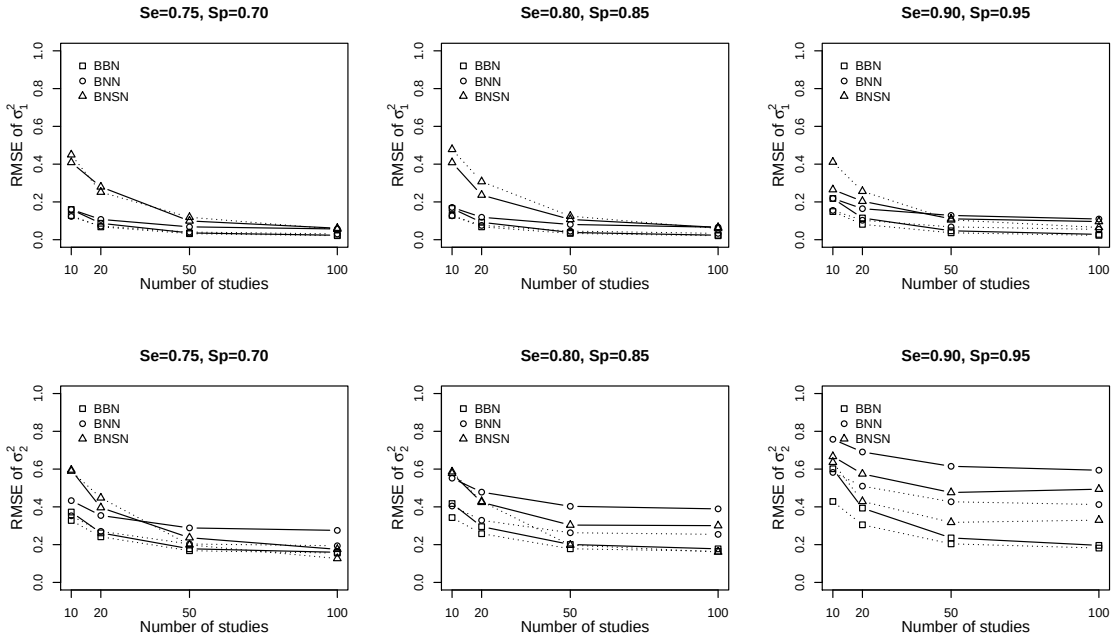

Figure 2.9: RMSE of σ_1^2 and σ_2^2 for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,0.35), (α_1,α_2) = (1.55,2.06) and BNSN is the true model.

87.0) and 67.0 (61.0, 72.0), respectively. According to Table 2.2, the BBN method's estimated pooled Se is the largest. The pooled Se and pooled Sp of the BNN and BNSN method are comparable which mirrors our simulation study. The 95% CI for true average Se and true average Sp of the BBN method is the widest, and that of the BNSN method is the narrowest which is in line with our simulation results. The estimated skewness parameter of the BNSN method is different from zero and significant (p_b-value = 0.012), providing sufficient evidence for the appropriateness of our proposed method over the BNN method for the *EUS* dataset. Moreover, the

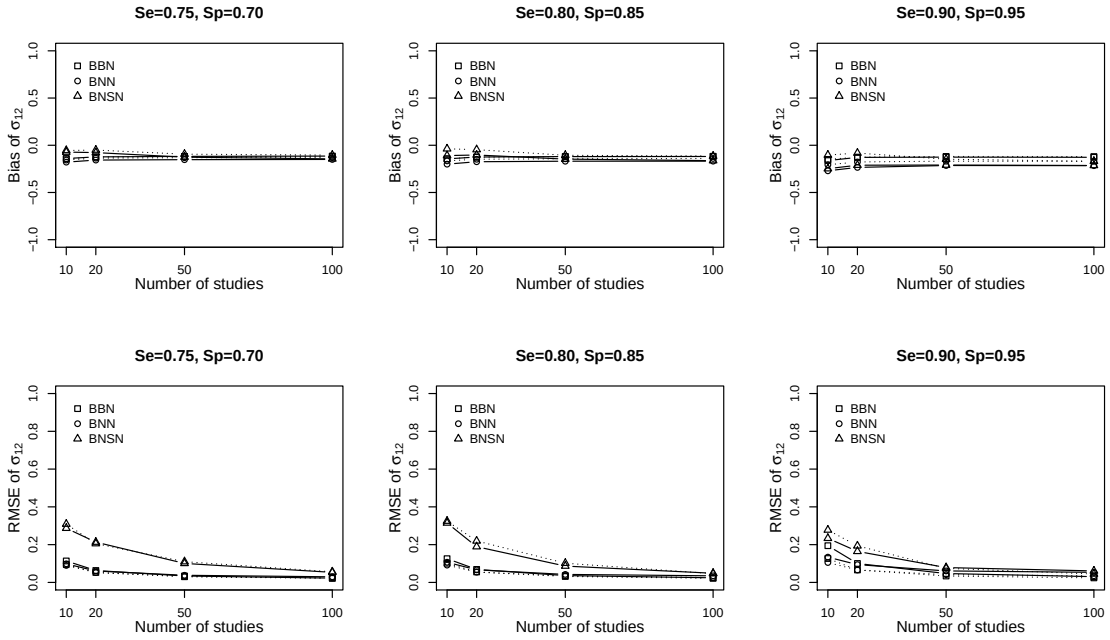

Figure 2.10: Bias (top panel) and RMSE (bottom panel) of σ_{12} for the BBN, BNN and BNSN model for small (solid line) and large (dotted line) sample sizes, different combinations of true (Se,Sp) pair and number of studies when (σ_1^2,σ_2^2,σ_{12}) = (0.72,1.17,0.35), (α_1,α_2) = (-0.56,-0.37) and BNSN is the true model.

AIC of our BNSN method is the smallest of the three methods—further suggesting that our BNSN method is the best fitting model to the *EUS* dataset.

2.7 Discussion

In this paper, we investigated whether a flexible random-effects distribution would be appropriate in meta-analytic models for diagnostic test accuracy studies. The research is motivated by the fact that currently, well-established and widely-used

Table 2.2: Estimates of the pooled Se (95% CI), pooled Sp (95% CI), between-study variances and covariance, skewness parameter and its LRT, and model goodness-of-fit statistic for the *EUS* dataset of Mocellin and Pasquali (2015)

.

Parameters	Model		
	BBN	BNN	BNSN
Se (95% CI)	83.20 (79.03, 86.68)	80.98 (76.90, 84.48)	80.99 (76.97, 84.44)
Sp (95% CI)	66.58 (60.57, 72.08)	66.91 (61.87, 71.58)	66.91 (62.0, 71.48)
σ_1^2	0.48	0.38	0.54
σ_2^2	0.56	0.39	0.95
σ_{12}	-0.22	-0.26	-0.56
$\widehat{\alpha}_1, \widehat{\alpha}_2$	-	-	2.36, -7.77
LRT (p_b-value)	-	-	5.70 (0.012)
AIC	516.430	226.933	225.233

p_b-value stands for a p-value obtained based on 1000 bootstrap samples. See section 2.7 for details of the bootstrap hypothesis test.

methods for meta-analysis of diagnostic test studies do not allow flexible random-effects distribution. Therefore, this paper extended the BNN method of Reitsma et al. (2005), by proposing the bivariate skew-normal as a distribution for the random-effects (true study-specific effect sizes) in a meta-analysis of diagnostic test studies.

The methodology of our proposed model is based on the assumption that the within-study variability can approximately be modelled using the normal distribution (invoking the central limit theorem), but the between-study variability is modelled by the skew-normal distribution. We used the maximum likelihood estimation of the between-study covariance matrix by deriving the marginal distribution of our model (see **Appendix I**), and the pooled Se and pooled Sp are obtained by the usual weighted least-squares approach using the inverse of the sum of the within-and between-study covariance matrices of the proposed model as a weight.

Overall, our proposed BNSN model yielded similar results to the BNN model

with regards to bias, RMSE and CP of the overall Se and Sp. Our result is consistent with those of Lee and Thompson (2008) and Neuhaus et al. (1992) who noted that using flexible random-effects distribution has no significant impact on the fixed-effect estimates. On the other hand, our simulation study revealed that the proposed model performed better than the BNN model in terms of bias and comparable with regards to the RMSE of the between-study (co)variances. This result is also in line with Turner et al. (2001), Pinheiro et al. (2001), and Fernández and Steel (1998) who reported that the variances of the random-effects distribution are affected by misspecification of the random-effects distribution. We want to emphasize more on the implications of the difference in between-study variances among the models. Whereas reporting the overall Se and Sp is one of the primary interest in a meta-analysis of DTA studies for the quantification of diagnostic performance, it has been noted that quantifying the degree of heterogeneity between-studies by estimating the between-study (co)variances, appraising the impact of heterogeneity by calculating statistic such as the I^2 (Jackson et al., 2012; Gasparrini et al., 2012), and constructing a prediction interval for an effect size in a new study (Van Houwelingen et al., 2002; Schwarzer et al., 2015; Higgins et al., 2009), in which the estimated between-study (co)variances play a vital role, is very important to make a complete inference in random-effects meta-analysis (Rücker et al., 2008; Higgins et al., 2009). As a result, we suggest using the proposed flexible random-effects model since it performed favorably in terms of the bias and RMSE of the between-study (co)variances compared to the standard BNN model. Although results are not shown here, we have also observed that the BNSN model yielded a narrower CI width than the BBN and BNN

model regardless of the parameters we varied in the simulation.

In our simulation study, we observed that the Nelder-Mead's (Nelder and Mead, 1965) simplex algorithm degenerates for at most 1 out of 1000 (0.1%) of the data in 17 out of 576 scenarios when BNSN is the true model, and in 2 out of 144 scenarios when BBN is the true model. This degeneracy of the Nelder-Mead algorithm occurs when the points that constitute the initial simplex lie in the same hyperplane, for example, when there are more than two vertices on a line (Lagarias et al., 1998). We excluded those datasets and reported the simulation results using the remaining 99.9% of the data. In general, the three models resulted in better optimality measures as the number of studies and sample size increases. Moreover, the performance of the proposed BNSN model improves as Se and Sp moves away from 100%. Our simulations also revealed that the BBN and BNSN models are robust to model misspecification despite other parameters varied in the simulations and the BNN is the least robust model to model misspecification. However, the performance of the BNN model improves as the elements of the between-study covariance matrix and skewness parameter decrease in magnitude. Applying the methods to a published data revealed results matching with our simulations. The methods, particularly the BNN and BNSN, produced similar point estimates of the pooled Se and Sp. However, the proposed model yielded narrower CIs and demonstrated superior goodness-of-fit. These results resemble those of Lee and Thompson (2008) when they assumed skewed and heavy-tailed distribution for the random-effects in univariate meta-analysis.

The proposed BNSN method requires estimating two more parameters than the BNN and BBN model, which may cause difficulty fitting meta-analyses with a small

number of studies. However, we have shown in Section 2.5 with meta-analyses having as few as ten studies that the proposed method can fit small meta-analyses without a problem. Another potential limitation of the BNSN model is that it requires slightly longer computational time than the BNN and BBN methods. Lastly, estimation of the skewness parameter of the proposed model may not be robust to starting values. Arellano-Valle et al. (2005) and others suggested the use of alternative approaches such as the expectation-maximization (EM) algorithm to estimate parameters of the multivariate skew-normal distribution. Therefore, when testing the significance of the skewness parameter, we implemented the parametric bootstrap to obtain the distribution for the LRT statistic and an appropriate p-value as shown in Section 2.6. To compute the bootstrap p-value, we first fit the BNN and BNSN models to the observed DTA data and obtain the observed LRT statistic, LRT_o. Then, for a fixed bootstrap sample $b = 1, 2, ..., 1000$, we generate new DTA data using the parameter estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ obtained from the BNN model (i.e., under the null hypothesis that there is no skewness) and data characteristics (i.e., n and k) as explained in Subsection 2.5.1. We then fit both the BNN and BNSN models to each bootstrap data and compute the bootstrap LRT statistic, LRT_b, $b = 1, 2, ..., k$. Finally, we calculate the bootstrap p-value using the expression

$$p_b\text{-value} = \frac{1 + \sum_{b=1}^{1000} I(LRT_b \geq LRT_o)}{1 + 1000},$$

where I is the indicator variable. We implemented our proposed method in the R programming language, and the R code used to produce the results presented in this chapter is available upon request from the author.

In conclusion, although the skewness parameter is not of primary interest in meta-analytic models, when the distribution of the data is not symmetric, implementing a robust method like the proposed BNSN model that accommodates asymmetry, would significantly contribute towards valid and reliable inference. It is important to note that our proposed model can be fitted to DTA datasets innocently since the standard BNN model is a special case of the proposed model. For practitioners, we suggest fitting both the standard BNN and our proposed model and choose the better model based on goodness-of-fit (AIC) and hypothesis test for the skewness parameter (LRT). Overall, the findings in this paper support the importance of robust bivariate random-effects model with flexible random-effects distribution for meta-analysis of diagnostic test accuracy studies.

Appendix I: Proof of (2.17)

Result 1: The marginal model of the BNSN model defined in (2.15)–(2.16), is given by

$$\mathbf{Y}_i \sim SN_2(\boldsymbol{\eta}, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i) \tag{2.22}$$

where

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} + \boldsymbol{\Psi}_i;\ \ \boldsymbol{\omega}_i = \frac{\boldsymbol{\Sigma}_i^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}}{\sqrt{1+\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}}};\ \ \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\alpha};\ \text{ and } \boldsymbol{\Gamma}_i = \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Psi}_i^{-1}.$$

Below, we discuss some matrix results which we make use of to prove **Result 1**.

Result 2: For any nonsingular square matrices $\mathbf{A}$ and $\mathbf{B}$ of dimension $n \times n$, and any conformable matrices $\mathbf{U}$ and $\mathbf{V}$, the Sherman-Morrison-Woodbury identity for the inverse of sums of matrices is

$$(\mathbf{A}+\mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1} \tag{2.23}$$

Interested users are referred to Henderson and Searle (1981) for details and derivation of (2.23). The following two results, for which we provide the proofs, are also important to prove **Result 1**.

Result 3: For any nonsingular square matrices $\mathbf{A}$ and $\mathbf{B}$ of dimension $n \times n$, the

following identity holds.

$$(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} = \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} \tag{2.24}$$

Note that the left hand side of (2.24) exists by (2.23) assuming that $\mathbf{U} = \mathbf{V} = \mathbf{I}_n$, since $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ are also nonsingular.

Proof of Result 3: It is clear that for any two invertible square matrices, $\mathbf{A}$ and $\mathbf{B}$, the following holds.

$$\begin{aligned}
\mathbf{B}^{-1}(\mathbf{A} + \mathbf{B}) &= \mathbf{B}^{-1}\mathbf{A} + \mathbf{I}_n \\
\Rightarrow \mathbf{B}^{-1} &= (\mathbf{B}^{-1}\mathbf{A} + \mathbf{I}_n)(\mathbf{A} + \mathbf{B})^{-1}, \text{ since } (\mathbf{A} + \mathbf{B})^{-1} \text{ exists by (2.23)} \\
\Rightarrow \mathbf{I}_n &= (\mathbf{B}^{-1}\mathbf{A} + \mathbf{I}_n)(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} \\
\Rightarrow \mathbf{I}_n &= (\mathbf{B}^{-1} + \mathbf{A}^{-1})\mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} \\
\Rightarrow (\mathbf{B}^{-1} + \mathbf{A}^{-1})^{-1} &= \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}, \text{ as desired.}
\end{aligned}$$

Result 4: For any two nonsingular square matrices $\mathbf{A}$ and $\mathbf{B}$, the following identity is true.

$$|\mathbf{A} + \mathbf{B}||(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}| = |\mathbf{A}||\mathbf{B}| \tag{2.25}$$

Proof of Result 4: To show that (2.25) is true, consider (2.24)

$$
\begin{aligned}
(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1} &= \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}, \text{ then taking determinants of both sides we have}\\
|(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}| &= |\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}|\\
\Rightarrow |(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}| &= |\mathbf{A}||(\mathbf{A}+\mathbf{B})^{-1}||\mathbf{B}|\\
\Rightarrow |(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}| &= |(\mathbf{A}+\mathbf{B})^{-1}||\mathbf{A}||\mathbf{B}|\\
\Rightarrow |\mathbf{A}+\mathbf{B}||(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}| &= |(\mathbf{A}+\mathbf{B})||(\mathbf{A}+\mathbf{B})^{-1}||\mathbf{A}||\mathbf{B}|\\
\Rightarrow |\mathbf{A}+\mathbf{B}||(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}| &= |\mathbf{A}||\mathbf{B}|, \text{ as required.}
\end{aligned}
$$

Now we can provide the proof of our main result below.

Proof of Result 1: The marginal model can be obtained from (2.15)–(2.16) by integrating out random-effects from the joint distribution of $\mathbf{Y}_i$ and $\mathbf{b}_i$ as shown below.

$$
\begin{aligned}
f_{\mathbf{Y}_i}(\mathbf{y}_i) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{Y}_i,\mathbf{b}_i}(\mathbf{y}_i,\mathbf{b}_i)d\mathbf{b}_i, i=1,...,k;\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{Y}_i|\mathbf{b}_i}(\mathbf{y}_i|\mathbf{b}_i)f_{\mathbf{b}_i}(\mathbf{b}_i)d\mathbf{b}_i\\
&= \int_{\mathbb{R}^2} \phi_2(\mathbf{y}_i;\boldsymbol{\eta}+\mathbf{b}_i,\boldsymbol{\Psi}_i)2\phi_2(\mathbf{b}_i;\boldsymbol{\eta},\boldsymbol{\Sigma})\Phi_1(\boldsymbol{\alpha}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{b}_i-\mathbf{0}))d\mathbf{b}_i\\
&= 2\int_{\mathbb{R}^2} \frac{1}{2\pi|\boldsymbol{\Psi}_i|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(\mathbf{y}_i-\boldsymbol{\eta}-\mathbf{b}_i)^T\boldsymbol{\Psi}^{-1}(\mathbf{y}_i-\boldsymbol{\eta}-\mathbf{b}_i)\}\\
&\times \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(\mathbf{b}_i-\mathbf{0})^T\boldsymbol{\Sigma}^{-1}(\mathbf{b}_i-\mathbf{0})\}\Phi_1(\boldsymbol{\alpha}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{b}_i-\mathbf{0}))d\mathbf{b}_i
\end{aligned}
$$

$$f_{\mathbf{Y}_i}(\mathbf{y}_i) = \frac{1}{2\pi^2|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\boldsymbol{\Psi}_i|^{\frac{1}{2}}} \int_{\mathbb{R}^2} \exp\{-\frac{1}{2}(\mathbf{x}_i - \mathbf{b}_i)^T\boldsymbol{\Psi}^{-1}(\mathbf{x}_i - \mathbf{b}_i)\} \exp\{-\frac{1}{2}(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)\}$$
$$\times \Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i, \tag{2.26}$$
$$\text{where } \mathbf{x}_i = \mathbf{y}_i - \boldsymbol{\eta} \text{ and } \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\alpha} \tag{2.27}$$

Now consider the quadratic form

$$\begin{aligned}(\mathbf{x}_i - \mathbf{b}_i)^T\boldsymbol{\Psi}_i^{-1}(\mathbf{x}_i - \mathbf{b}_i) &= (\mathbf{x}_i^T - \mathbf{b}_i^T)\boldsymbol{\Psi}_i^{-1}(\mathbf{x}_i - \mathbf{b}_i) \\ &= (\mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1} - \mathbf{b}_i^T\boldsymbol{\Psi}_i^{-1})(\mathbf{x}_i - \mathbf{b}_i) \\ &= \mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{x}_i - \mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i - \mathbf{b}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{x}_i + \mathbf{b}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i \\ &= \mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{x}_i - 2\mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i + \mathbf{b}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i \end{aligned} \tag{2.28}$$

Then substituting back (2.28) into (2.26), we get

$$\begin{aligned} f_{\mathbf{Y}_i}(\mathbf{y}_i) &= \frac{1}{2\pi^2|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\boldsymbol{\Psi}_i|^{\frac{1}{2}}} \int_{\mathbb{R}^2} \exp\{-\frac{1}{2}(\mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{x}_i - 2\mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i + \mathbf{b}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i)\} \exp\{-\frac{1}{2}\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i\} \\ &\quad \times \Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i \\ &= \frac{\exp\{-\frac{1}{2}(\mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{x}_i)\}}{2\pi^2|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\boldsymbol{\Psi}_i|^{\frac{1}{2}}} \int_{\mathbb{R}^2} \exp\{-\frac{1}{2}(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i + \mathbf{b}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i - 2\mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i)\}\Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i \\ &= F_1(\mathbf{x}_i) \int_{\mathbb{R}^2} \exp\{-\frac{1}{2}\left(\mathbf{b}_i^T(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Psi}_i^{-1})\mathbf{b}_i - 2\mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i\right)\}\Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i \\ &\text{where } F_1(\mathbf{x}_i) = \frac{\exp\{-\frac{1}{2}(\mathbf{x}_i^T\boldsymbol{\Psi}_i^{-1}\mathbf{x}_i)\}}{2\pi^2|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\boldsymbol{\Psi}_i|^{\frac{1}{2}}} \end{aligned} \tag{2.29}$$

$$f_{\mathbf{Y}_i}(\mathbf{y}_i) = F_1(\mathbf{x}_i) \int_{\mathbb{R}^2} \exp\{-\frac{1}{2}\left(\mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - 2\mathbf{x}_i^T \mathbf{\Psi}_i^{-1} \mathbf{b}_i\right)\} \Phi_1(\mathbf{\Lambda}^T \mathbf{b}_i) d\mathbf{b}_i \tag{2.30}$$

where $\mathbf{\Gamma}_i = \mathbf{\Sigma}^{-1} + \mathbf{\Psi}_i^{-1}$ is a 2×2 positive definite matrix since $\mathbf{\Sigma}$ and $\mathbf{\Psi}_i$ are positive definite. Now, consider the exponent in the above integral,

$$\begin{aligned}
\mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - 2\mathbf{x}_i^T \mathbf{\Psi}_i^{-1} \mathbf{b}_i &= \mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - 2\mathbf{x}_i^T \mathbf{\Psi}_i^{-1} \mathbf{I}_2 \mathbf{b}_i \\
&= \mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - 2\mathbf{x}_i^T \mathbf{\Psi}_i^{-1} \mathbf{\Gamma}_i^{-1} \mathbf{\Gamma}_i \mathbf{b}_i, \text{ since } \mathbf{\Gamma}_i^{-1} \text{ exists by (2.23) or by definition.} \\
&= \mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - 2\mathbf{x}_i^T \mathbf{\Psi}_i^{-1} \mathbf{\Gamma}_i^{-1} \mathbf{\Gamma}_i \mathbf{b}_i + (\mathbf{\Gamma}_i^{-1} \mathbf{\Psi}_i^{-1} \mathbf{x}_i)^T \mathbf{\Gamma}_i (\mathbf{\Gamma}_i^{-1} \mathbf{\Psi}_i^{-1} \mathbf{x}_i) \\
&\quad - (\mathbf{\Gamma}_i^{-1} \mathbf{\Psi}_i^{-1} \mathbf{x}_i)^T \mathbf{\Gamma}_i (\mathbf{\Gamma}_i^{-1} \mathbf{\Psi}_i^{-1} \mathbf{x}_i) \\
&= \mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - 2F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i \mathbf{b}_i + F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i) - F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i) \\
&\quad \text{where } F_2^T(\mathbf{x}_i) = \mathbf{\Gamma}_i^{-1} \mathbf{\Psi}_i^{-1} \mathbf{x}_i \text{ is a vector of length two.} \\
&= \{\mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - 2F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i \mathbf{b}_i + F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i)\} - F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i) \\
&= \{\mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - 2F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i \mathbf{b}_i + F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i)\} - F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i) \\
&= \{\mathbf{b}_i^T \mathbf{\Gamma}_i \mathbf{b}_i - F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i \mathbf{b}_i - \mathbf{b}_i^T \mathbf{\Gamma}_i F_2(\mathbf{x}_i) + F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i)\} \\
&\quad - F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i) \\
&= \{(\mathbf{b}_i^T \mathbf{\Gamma}_i - F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i)(\mathbf{b}_i - F_2(\mathbf{x}_i))\} - F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i) \\
&= \{(\mathbf{b}_i - F_2(\mathbf{x}_i))^T \mathbf{\Gamma}_i (\mathbf{b}_i - F_2(\mathbf{x}_i)) - F_2^T(\mathbf{x}_i) \mathbf{\Gamma}_i F_2(\mathbf{x}_i)\}
\end{aligned} \tag{2.31}$$

Then, substituting (2.31) back into (2.30), we get

$$\begin{aligned} f_{\mathbf{Y}_i}(\mathbf{y}_i) &= F_1(\mathbf{x}_i)\int_{\mathbb{R}^2}\exp\{-\frac{1}{2}\left((\mathbf{b}_i - F_2(\mathbf{x}_i))^T\mathbf{\Gamma}_i(\mathbf{b}_i - F_2(\mathbf{x}_i)) - F_2^T(\mathbf{x}_i)\mathbf{\Gamma}_i F_2(\mathbf{x}_i)\right)\} \\ &\times \Phi_1(\mathbf{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i \\ &= F_1(\mathbf{x}_i)\exp\{\frac{1}{2}F_2^T(\mathbf{x}_i)\mathbf{\Gamma}_i F_2(\mathbf{x}_i)\}\int_{\mathbb{R}^2}\exp\{-\frac{1}{2}\left((\mathbf{b}_i - F_2(\mathbf{x}_i))^T\mathbf{\Gamma}_i(\mathbf{b}_i - F_2(\mathbf{x}_i))\right)\} \\ &\times \Phi_1(\mathbf{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i \\ &= F_3(\mathbf{x}_i)\int_{\mathbb{R}^2}\exp\{-\frac{1}{2}\left((\mathbf{b}_i - F_2(\mathbf{x}_i))^T\mathbf{\Gamma}_i(\mathbf{b}_i - F_2(\mathbf{x}_i))\right)\}\Phi_1(\mathbf{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i \end{aligned} \tag{2.32}$$

$$\begin{aligned} \text{where } F_3(\mathbf{x}_i) &= F_1(\mathbf{x}_i)\exp\{\frac{1}{2}F_2^T(\mathbf{x}_i)\mathbf{\Gamma}_i F_2(\mathbf{x}_i)\} \\ &= \frac{\exp\{-\frac{1}{2}(\mathbf{x}_i^T\mathbf{\Psi}_i^{-1}\mathbf{x}_i)\}}{2\pi^2|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{\Psi}_i|^{\frac{1}{2}}}\exp\{\frac{1}{2}\mathbf{x}_i^T\mathbf{\Psi}_i^{-1}\mathbf{\Gamma}_i^{-1}\mathbf{\Gamma}_i\mathbf{\Gamma}_i^{-1}\mathbf{\Psi}_i^{-1}\mathbf{x}_i\} \\ &= \frac{\exp\{-\frac{1}{2}\left(\mathbf{x}_i^T\mathbf{\Psi}_i^{-1}\mathbf{x}_i - \mathbf{x}_i^T\mathbf{\Psi}_i^{-1}\mathbf{\Gamma}_i^{-1}\mathbf{\Psi}_i^{-1}\mathbf{x}_i\right)\}}{2\pi^2|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{\Psi}_i|^{\frac{1}{2}}} \\ &= \frac{\exp\{-\frac{1}{2}\mathbf{x}_i^T\left(\mathbf{\Psi}_i^{-1} - \mathbf{\Psi}_i^{-1}\mathbf{\Gamma}_i^{-1}\mathbf{\Psi}_i^{-1}\right)\mathbf{x}_i\}}{2\pi^2|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{\Psi}_i|^{\frac{1}{2}}} \\ &= \frac{\exp\{-\frac{1}{2}\mathbf{x}_i^T\left(\mathbf{\Psi}_i^{-1} - \mathbf{\Psi}_i^{-1}(\mathbf{\Sigma}^{-1} + \mathbf{\Psi}_i^{-1})^{-1}\mathbf{\Psi}_i^{-1}\right)\mathbf{x}_i\}}{2\pi^2|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{\Psi}_i|^{\frac{1}{2}}}, \\ &= \frac{\exp\{-\frac{1}{2}\mathbf{x}_i^T(\mathbf{\Sigma} + \mathbf{\Psi}_i)^{-1}\mathbf{x}_i\}}{2\pi^2|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{\Psi}_i|^{\frac{1}{2}}}, \text{ by (2.23)} \\ &= \frac{\exp\{-\frac{1}{2}\mathbf{x}_i^T\mathbf{\Sigma}_i^{-1}\mathbf{x}_i\}}{2\pi^2|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{\Psi}_i|^{\frac{1}{2}}} \end{aligned} \tag{2.33}$$

Next, substituting (2.33) into (2.32), we have

$$
\begin{aligned}
f_{\mathbf{Y}_i}(\mathbf{y}_i) &= \frac{\exp\{-\frac{1}{2}\mathbf{x}_i^T\boldsymbol{\Sigma}_i^{-1}\mathbf{x}_i\}}{2\pi^2|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\boldsymbol{\Psi}_i|^{\frac{1}{2}}}\int_{\mathbb{R}^2}\exp\{-\frac{1}{2}\left((\mathbf{b}_i-F_2(\mathbf{x}_i))^T\boldsymbol{\Gamma}_i(\mathbf{b}_i-F_2(\mathbf{x}_i))\right)\}\\
&\quad\times\Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i\\
&= \int_{\mathbb{R}^2}\frac{2\exp\{-\frac{1}{2}\mathbf{x}_i^T\boldsymbol{\Sigma}_i^{-1}\mathbf{x}_i\}}{2\pi 2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\boldsymbol{\Psi}_i|^{\frac{1}{2}}}\exp\{-\frac{1}{2}\left((\mathbf{b}_i-F_2(\mathbf{x}_i))^T\boldsymbol{\Gamma}_i(\mathbf{b}_i-F_2(\mathbf{x}_i))\right)\}\Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i\\
&= \int_{\mathbb{R}^2}2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)\phi_2(\mathbf{b}_i|F_2(\mathbf{x}_i),\boldsymbol{\Gamma}_i^{-1})\Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i \quad \text{as a consequence of (2.25)}\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)\int_{\mathbb{R}^2}\phi_2(\mathbf{b}_i|F_2(\mathbf{x}_i),\boldsymbol{\Gamma}_i^{-1})\Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)d\mathbf{b}_i\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)\mathbb{E}\left[\Phi_1(\boldsymbol{\Lambda}^T\mathbf{b}_i)\right], \text{ where } \mathbf{b}_i\sim N_2(F_2(\mathbf{x}_i),\boldsymbol{\Gamma}_i^{-1})\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)\mathbb{E}\left[P(Z\le\boldsymbol{\Lambda}^T\mathbf{b}_i)\right], \text{ where } Z\sim N(0,1)\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)P(Z\le\boldsymbol{\Lambda}^T\mathbf{b}_i)\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)P(W\le 0), \text{ where } W=Z-\boldsymbol{\Lambda}^T\mathbf{b}_i\sim N(-\boldsymbol{\Lambda}^TF_2(\mathbf{x}_i),1+\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda})\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)P\left(\frac{W-\mathbb{E}(W)}{\sqrt{\text{var}(W)}}\le\frac{0-\mathbb{E}(W)}{\sqrt{\text{var}(W)}}\right)\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)P\left(Z\le\frac{\boldsymbol{\Lambda}^TF_2(\mathbf{x}_i)}{\sqrt{1+\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}}}\right)\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)\Phi_1\left(\frac{\boldsymbol{\Lambda}^TF_2(\mathbf{x}_i)}{\sqrt{1+\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}}}\right)\\
&= 2\phi_2(\mathbf{x}_i;\mathbf{0},\boldsymbol{\Sigma}_i)\Phi_1\left(\frac{\boldsymbol{\Lambda}^T(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Psi}_i^{-1})^{-1}\boldsymbol{\Psi}_i^{-1}\mathbf{x}_i}{\sqrt{1+\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}}}\right)
\end{aligned}
$$

$$
\begin{aligned}
f_{\mathbf{Y}_i}(\mathbf{y}_i) &= 2\phi_2(\mathbf{x}_i; \mathbf{0}, \boldsymbol{\Sigma}_i)\Phi_1\left(\frac{\boldsymbol{\Lambda}^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}+\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}_i^{-1}\mathbf{x}_i}{\sqrt{1+\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}}}\right), \text{ which is a consequence of (2.24)} \\
&= 2\phi_2(\mathbf{x}_i; \mathbf{0}, \boldsymbol{\Sigma}_i)\Phi_1\left(\frac{\boldsymbol{\Lambda}^T\boldsymbol{\Sigma}\boldsymbol{\Sigma}_i^{-1}\mathbf{x}_i}{\sqrt{1+\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}}}\right) \\
&= 2\phi_2(\mathbf{x}_i; \mathbf{0}, \boldsymbol{\Sigma}_i)\Phi_1(\boldsymbol{\omega}_i^T\boldsymbol{\Sigma}_i^{-\frac{1}{2}}\mathbf{x}_i), \text{ where } \boldsymbol{\omega}_i = \frac{\boldsymbol{\Sigma}_i^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}}{\sqrt{1+\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}}}
\end{aligned}
$$

Finally, the distribution of our marginal model becomes

$$
\begin{aligned}
f_{\mathbf{Y}_i}(\mathbf{y}_i; \boldsymbol{\eta}, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i) &= 2\phi_2(\mathbf{y}_i - \boldsymbol{\eta}; \mathbf{0}, \boldsymbol{\Sigma}_i)\Phi_1(\boldsymbol{\omega}_i^T\boldsymbol{\Sigma}_i^{-\frac{1}{2}}(\mathbf{y}_i - \boldsymbol{\eta})), \text{ by (2.27)} \\
&= 2\phi_2(\mathbf{y}_i; \boldsymbol{\eta}, \boldsymbol{\Sigma}_i)\Phi_1(\boldsymbol{\omega}_i^T\boldsymbol{\Sigma}_i^{-\frac{1}{2}}(\mathbf{y}_i - \boldsymbol{\eta})) \\
\Rightarrow \mathbf{Y}_i &\sim SN_2(\boldsymbol{\eta}, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i), \text{ proving } \textbf{Result 1.}
\end{aligned}
$$

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pages 215–222. Springer.

Arellano-Valle, R., Bolfarine, H., and Lachos, V. (2005). Skew-normal linear mixed models. *Journal of Data Science*, 3(4):415–438.

Arnold, B. C., Beaver, R. J., Azzalini, A., Balakrishnan, N., Bhaumik, A., Dey, D., Cuadras, C., and Sarabia, J. M. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, 11(1):7–54.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):579–602.

Azzalini, A. and Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726.

Baker, R. and Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health Care Management Science*, 11(2):121–131.

Baker, R. and Jackson, D. (2016). New models for describing outliers in meta-analysis. *Research Synthesis Methods*, 7(3):314–328.

Beath, K. J. (2014). A finite mixture method for outlier detection and robustness in meta-analysis. *Research Synthesis Methods*, 5(4):285–293.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley and Sons.

Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*, 59(12):1331.

Doebler, P. (2017). *mada: Meta-Analysis of Diagnostic Accuracy*. R package version 0.5.8.

Doebler, P., Holling, H., and Böhning, D. (2012). A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. *Psychological Methods*, 17(3):418.

Fernández, C. and Steel, M. F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2012). Multivariate meta-analysis for non-linear and other multi-parameter associations. *Statistics in Medicine*, 31(29):3821–3839.

Hamza, T. H., van Houwelingen, H. C., and Stijnen, T. (2008). The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of Clinical Epidemiology*, 61(1):41–51.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1):53–60.

Henze, N. and Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10):3595–3617.

Higgins, J. P., Thompson, S. G., and Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1):137–159.

Huang, H. (2018). Metatron: meta-analysis for classification data and correction to imperfect reference. R package version 0.1-1. 2014.

Jackson, D., White, I. R., and Riley, R. D. (2012). Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Statistics in Medicine*, 31(29):3805–3820.

Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2):151–162.

Kovalchik, S. (2013). Tutorial on meta-analysis in R. *R useR*.

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147.

Lee, K. J. and Thompson, S. G. (2008). Flexible parametric models for random-effects distributions. *Statistics in Medicine*, 27(3):418–434.

Mecklin, C. J. and Mundfrom, D. J. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation*, 75(2):93–107.

Mocellin, S. and Pasquali, S. (2015). Diagnostic accuracy of endoscopic ultrasonography (EUS) for the preoperative locoregional staging of primary gastric cancer. *Cochrane Database of Systematic Reviews*, (2).

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Neuhaus, J. M., Hauck, W. W., and Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79(4):755–762.

Paul, M., Riebler, A., Bachmann, L., Rue, H., and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine*, 29(12):1325–1339.

Pinheiro, J. C., Liu, C., and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, 10(2):249–276.

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58(10):982–990.

Rücker, G., Schwarzer, G., Carpenter, J. R., and Schumacher, M. (2008). Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Medical Research Methodology*, 8(1):79.

Schwarzer, G., Carpenter, J. R., and Rücker, G. (2015). *Meta-Analysis with R*, volume 4724. New York: Springer.

Turner, R. M., Omar, R. Z., and Thompson, S. G. (2001). Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine*, 20(3):453–472.

Van Houwelingen, H. C., Arends, L. R., and Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in medicine*, 21(4):589–624.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D., and Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, 7(1):55–79.

Viechtbauer, W. et al. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3):1–48.

Chapter 3

Statistical methods for detecting outlying and influential studies in meta-analysis of diagnostic test accuracy studies

3.1 Abstract

Bivariate random-effects models are currently widely used to synthesize pairs of test sensitivity and specificity across studies. Inferences drawn based on these models may be distorted in the presence of outlying or influential studies. Currently, subjective methods such as inspection of forest plots are used to identify outlying studies in meta-analysis of diagnostic test accuracy studies. We propose objective methods

based on solid statistical reasoning for identifying outlying and/or influential studies. The proposed methods are validated using simulation study and illustrated on two published meta-analysis data. Our methods outperform and avoid the subjectivity of the currently used ad hoc methods. The proposed methods can be used as a sensitivity analysis tool concurrently with the current bivariate random-effects models or as a preliminary analysis tool for robust models that accommodate outlying and/or influential studies in meta-analysis of diagnostic test accuracy studies.

Keywords: Diagnostic test accuracy studies, Influential studies, Meta-analysis, Outlying studies, Sensitivity and Specificity

3.2 Introduction

Meta-analysis of diagnostic test accuracy (DTA) studies is a statistical method that allows synthesis of test characteristics from several independent studies addressing the same research question. Meta-analysis of DTA studies is an active area of research which is attracting new methodological developments which include the hierarchical bivariate random-effects models (Rutter and Gatsonis, 2001; Reitsma et al., 2005; Chu and Cole, 2006), copula-based mixed-effects models (Nikoloulopoulos, 2015; Kuss et al., 2014; Nikoloulopoulos, 2017, 2018), network meta-analysis models (Nyaga et al., 2018a,b), and a composite likelihood method of estimation (Chen et al., 2017). Typically, pairs of test sensitivity (Se) and test specificity (Sp) are combined. Hierarchical bivariate random-effects models (Rutter and Gatsonis,

2001; Reitsma et al., 2005; Chu and Cole, 2006) are widely used to synthesize pairs of Se and Sp. However, inferences and conclusions drawn based on these models may be distorted in the presence of outlying or influential studies. Therefore, it is vital to examine whether or not the studies included in the meta-analysis are outlying and/or influential.

The impact of outlying and/or influential studies has been widely studied in the conventional meta-analysis of treatment or intervention studies (Viechtbauer and Cheung, 2010; Gumedze and Jackson, 2011; Hedges and Olkin, 1985; Baker and Jackson, 2008; Beath, 2014; Baker and Jackson, 2016). Viechtbauer and Cheung (2010) and Gumedze and Jackson (2011) proposed different methods of identifying outlying or influential studies in meta-analysis of intervention studies. Viechtbauer and Cheung proposed methods based on residuals obtained from the univariate random-effects meta-analysis model by extending the method proposed by Hedges and Olkin (1985), which was based on the univariate fixed-effect meta-analysis model. Gumedze and Jackson proposed a 'variance shift outlier model' based on a univariate random-effects meta-analysis model aiming to detect outlying studies. Their methodology makes use of the likelihood ratio test (LRT) as a test statistic and the parametric bootstrap approach to get the sampling distribution of the observed LRT statistic.

Although the study of identifying outlying and influential studies has been given emphasis in the context of meta-analysis of intervention studies, it has not been investigated in the context of meta-analysis of diagnostic test accuracy studies. Currently, ad hoc methods used to detect outlying studies in meta-analysis of DTA studies include: inspection of forest plots (Devillé et al., 2002; Doria et al., 2006;

Petignat et al., 2007; Singal et al., 2009), bivariate box plots (Zhou et al., 2016; Pormohammad et al., 2017) and scatter plots of Se and Sp (de Jesus et al., 2009; Kriston et al., 2008).

However, these subjective methods of identifying outlying studies have limitations. They are based only on raw data and the observed effect sizes (Se and Sp) not on a statistical model which takes into account either the within-study variability or the between-study heterogeneity. Furthermore, the methods do not tell if an outlying study is influential or not.

To tackle the above-mentioned shortcomings of the ad hoc methods of detecting outlying studies, we propose two objective approaches based on solid statistical reasoning for detecting outlying and/or influential studies in meta-analysis of diagnostic test accuracy studies. In our first approach, we propose residual-based method that uses bivariate residuals from a bivariate random-effects model to examine the presence of outlying studies. In the second approach, we develop a bivariate random-effects mean-shift outlier model (BMSOM) and propose the likelihood ratio test (LRT) to test for the presence of outlying studies. The parametric bootstrap is used to obtain the sampling distribution of the LRT statistic and the corresponding rejection region. Moreover, we propose two objective functions or indexes to identify influential studies. Therefore, our proposed methods take into account both the within-and between-study variability in sensitivity and specificity since the methods are based on a bivariate random-effects statistical model.

The chapter is organized as follows. In Section 3.3, we motivate the proposed methods using two published meta-analyses of DTA studies data. In Section 3.4 we

discuss the statistical methodology of the proposed methods and we illustrate the proposed methods using two published meta-analysis of DTA studies in Section 3.5. In Section 3.6, the proposed methods are compared and validated using a simulation study. Finally, we conclude the chapter with a discussion and conclusion in Section 3.7.

3.3 Motivating examples

In this Section, we provide two real meta-analyses data to motivate the proposed methods. The data are chosen to demonstrate how the methods perform when the meta-analysis contains one or more potentially outlying or influential studies as detected by the ad hoc methods and when the number of studies and sample size varies.

3.3.1 US-Children data

The first data is about a ultrasonography test. Doria et al. (2006) studied the effectiveness of an ultrasonography test in diagnosing appendicitis in children and adults. The children version of the ultrasonography data (*US-Children*) consists of 23 studies and an average number of 77 and 254 children with and without appendicitis, respectively. The forest plot of the *US-Children* data is given in Figure 3.1. Examining the forest plot suggests that study 13 could be potentially outlying in sensitivity. Nonetheless, the forest plot does not tell if there are other potential outlying studies or if study 13 is influential. We reanalyze this data using our proposed methods and

present the results in Section 3.5.

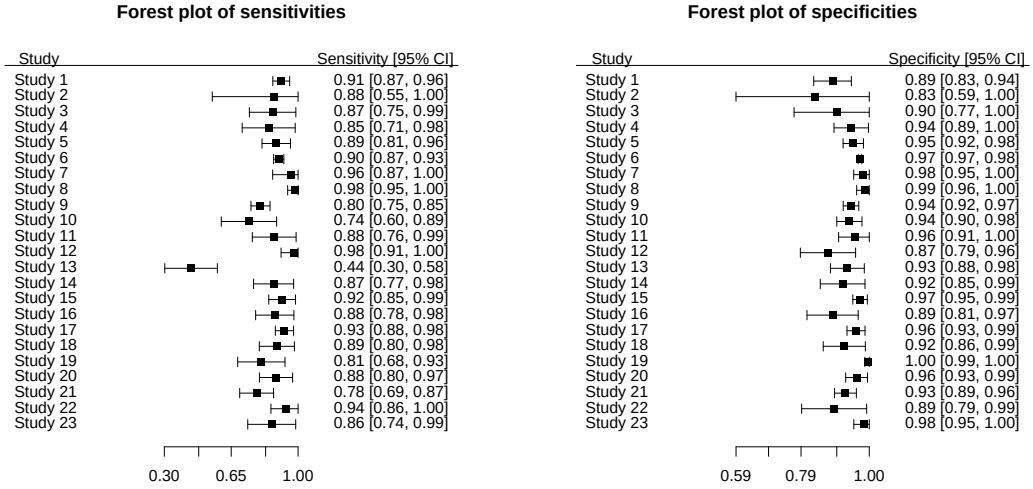

Figure 3.1: Forest plot for sensitivity (left) and specificity (right) of the *US-Children* data.

3.3.2 AUDIT-C data

The accuracy of the alcohol use disorder identification test (AUDIT), which consists of ten questions, in screening unhealthy alcohol consumption in adults is compared to the alcohol use disorder identification test-consumption ($AUDIT\text{-}C$) questionnaire, which has only three questions, by Kriston et al. (2008). The $AUDIT\text{-}C$ data consists of 14 studies and an average of 148 and 1,162 participants with and without alcohol use disorder, respectively. Figure 3.2 shows the forest plot of the $AUDIT\text{-}C$ dataset. It is not obvious which study is outlying, if any, based on the forest plot. At first glance, it might seem study 6 and 8 are potentially outlying in specificity. In Section

3.5, we have present results of our proposed approaches when applied to the *AUDIT-C* data.

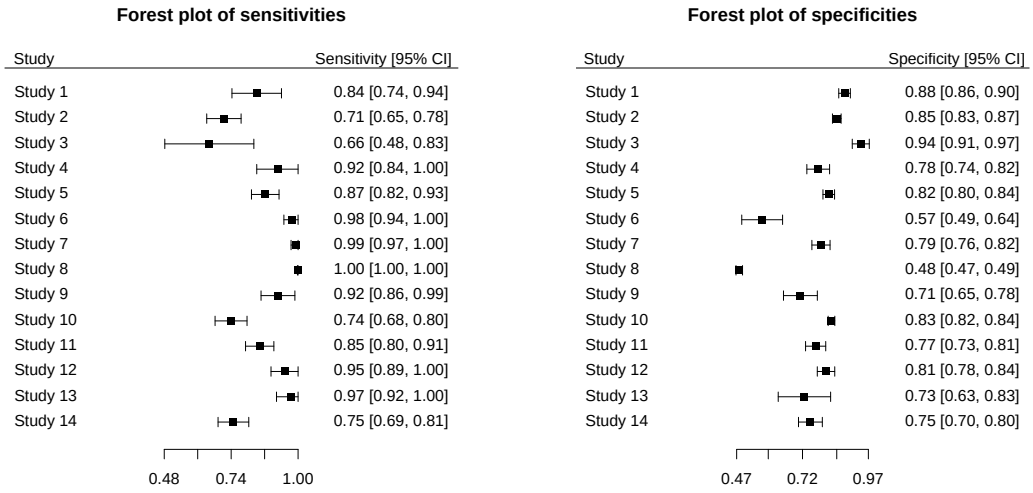

Figure 3.2: Forest plot for sensitivity (left) and specificity (right) of the *AUDIT-C* data.

3.4 Methods

In the following notations, we want to note that the observed study-specific sensitivities (Se_i) and specificities (Sp_i) are defined as functions of a threshold value, c, which separates positive and negative test results in the context of continuous biomarkers (see (Pepe, 2003), §4.2). However, for simplicity of notation, we omit the threshold c when describing the models. On the other hand, it is important to distinguish the notation when one needs to directly model the thresholds as in the hierarchical

summary receiver operating characteristic (HSROC) model of Rutter and Gatsonis (2001) which had been shown by Harbord et al. (2007) to be equivalent to the bivariate random-effects model discussed in this chapter in the absence of covariates.

3.4.1 Detecting outlying studies—residual-based approach

We first discuss in detail the bivariate linear mixed model of Reitsma et al. (2005) because it is the basis of our proposed method under the residual-based approach. The standard bivariate linear mixed model of Reitsma et al., which we call the bivariate normal-normal (BNN) model because it models both the within-and between-study variability using the bivariate normal distribution, is defined as follows. The BNN model assumes that within-study, the observed vector of response $\mathbf{y}_i = (y_{1i}, y_{2i})^T = (\text{logit}(\widehat{Se}_i), \text{logit}(\widehat{Sp}_i))^T$ is modeled according to

$$\mathbf{y}_i|\boldsymbol{\mu}_i \sim N_2(\boldsymbol{\mu}_i, \boldsymbol{\Psi}_i), \;\; i = 1, ..., k, \tag{3.1}$$

where $\boldsymbol{\mu}_i$ denotes the true study-specific logit(Se) and logit(Sp) and

$$\boldsymbol{\Psi}_i = \begin{pmatrix} \psi_{1i}^2 & 0 \\ 0 & \psi_{2i}^2 \end{pmatrix}, i = 1, ..., k,$$

represents the within-study covariance matrix of the observed response vector and it is assumed to be known. The assumed to be known within-study variances of the logit-transformed sensitivities and specificities are estimated from the data using

the first-order delta method. Accordingly, for $i = 1, ..., k$, ψ_{1i}^2= var(logit(Se_i)) = $1/(n_{1i}Se_i(1 - Se_i))$ and ψ_{2i}^2= var(logit(Sp_i)) = $1/(n_{2i}Sp_i(1 - Sp_i))$, where n_{1i} and n_{2i} denote the total number of subjects in the diseased and non-diseased group, respectively.

On the second level, the BNN model assumes heterogeneity among the true study-specific logit-transformed Se and Sp. Thus,

$$\boldsymbol{\mu}_i \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}), i = 1, ..., k, \tag{3.2}$$

where $\boldsymbol{\mu}$ denotes the overall or average true logit(Se) and logit(Sp) and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

is the between-study covariance matrix which denotes the heterogeneity parameter.

By combining (3.1) and (3.2) and assuming that the studies are independent, we arrive at the marginal model,

$$\mathbf{y}_i \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i), \;\; i = 1, ..., k, \tag{3.3}$$

where $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} + \boldsymbol{\Psi}_i$, and the restricted log-likelihood function,

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \text{const.} - \frac{1}{2}\sum_{i=1}^{k} \log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{k}(\mathbf{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) - \frac{1}{2}\log\left|\sum_{i=1}^{k}\boldsymbol{\Sigma}_i^{-1}\right|.$$

Once the estimate of the between-study covariance matrix, $\widehat{\boldsymbol{\Sigma}}$, is obtained using

iterative numerical methods, the parameter of interest, $\boldsymbol{\mu}$, is estimated using weighted average method, which is equivalent to the maximum likelihood estimate:

$$\widehat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{k}(\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i)^{-1}\right)^{-1} \sum_{i=1}^{k}(\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i)^{-1}\mathbf{y}_i.$$

Asymptotically, the estimated average effect size, $\widehat{\boldsymbol{\mu}}$, is approximately distributed as bivariate normal with mean $\boldsymbol{\mu}$ and covariance, VARCOV $= \left(\sum_{i=1}^{k}\widehat{\mathbf{W}}_i\right)^{-1} = \left(\sum_{i=1}^{k}(\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i)^{-1}\right)^{-1}$. Therefore, the approximate 100(1-α)% confidence interval for μ_j can be obtained as $\hat{\mu}_j \mp Z_{\frac{\alpha}{2}}\sqrt{\text{VARCOV}(j,j)}$, where $Z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$th percentile of the standard normal distribution and VARCOV(j,j) is the j-th diagonal element of VARCOV.

We now describe our residual-based method of detecting outlying studies in meta-analysis of diagnostic test accuracy studies. We generalized methods of identifying outliers in regression analysis and univariate meta-analysis developed by Hedges and Olkin (1985) and Viechtbauer and Cheung (2010), respectively. Therefore, our methodology reduces to their univariate approach when $\sigma_{12} = 0$. Furthermore, our approach can also be applied in any multivariate meta-analysis.

The raw residuals for study i can be defined as $\boldsymbol{\epsilon}_i = \mathbf{y}_i - \widehat{\boldsymbol{\mu}}$, where $\mathbf{y}_i$ and $\widehat{\boldsymbol{\mu}}$ are as defined before. Since $\text{var}(\boldsymbol{\epsilon}_i) = \widehat{\mathbf{W}}_i - (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1}$ (see **Appendix II**), the internally studentized residual for study i is

$$\mathbf{r}_i = \text{var}(\boldsymbol{\epsilon}_i)^{-\frac{1}{2}}\boldsymbol{\epsilon}_i = \left(\widehat{\mathbf{W}}_i - (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1}\right)^{-\frac{1}{2}}\boldsymbol{\epsilon}_i.$$

Since an observed effect size that deviates from the true fitted model (3.3) complicates the process of identifying outlying studies using the internally studentized residuals, because it pulls the estimated average effect size ($\widehat{\boldsymbol{\mu}}$) towards itself and yields a smaller $\boldsymbol{\epsilon}_i$, the studentized deleted residuals also known as the externally studentized residuals are recommended for identifying outlying studies (Viechtbauer and Cheung, 2010; Hedges and Olkin, 1985). The studentized deleted residuals which is computed by leaving out one study at a time is defined as

$$\mathbf{t}_i = \left(\text{var}(\boldsymbol{\epsilon}_{i(-i)})\right)^{-\frac{1}{2}} \boldsymbol{\epsilon}_{i(-i)}, \;\; i = 1, ..., k, \tag{3.4}$$

where $\boldsymbol{\epsilon}_{i(-i)} = \mathbf{y}_i - \widehat{\boldsymbol{\mu}}_{i(-i)}$ is the deleted residual for study i, $\widehat{\boldsymbol{\mu}}_{i(-i)}$ is the expected average logit(Se) and logit(Sp) for study i when the i-th study is excluded from the meta-analysis.

Since $\mathbf{y}_i$ and $\widehat{\boldsymbol{\mu}}_{i(-i)}$ are uncorrelated, the variance of the deleted residuals is (see **Appendix II**),

$$\text{var}(\boldsymbol{\epsilon}_{i(-i)}) = \widehat{\mathbf{W}}^{-1}_{i(-i)} + \left(\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)}\right)^{-1},$$

where $\widehat{\mathbf{W}}_{i(-i)} = \left(\widehat{\boldsymbol{\Sigma}}_{(-i)} + \boldsymbol{\Psi}_i\right)^{-1}$, $i = 1, ..., k$, is the weight for study i when the i-th study is excluded from the meta-analysis.

It can be shown that (see **Appendix II**), if the assumed model (3.3) is true, the studentized deleted residuals, $\mathbf{t}_i$, follow the standard bivariate normal distribution.

That is,

$$\mathbf{t}_i \sim N_2(\mathbf{0}, \mathbf{I}_2), \text{ equivalently, } \mathrm{t}_{ij} \sim N(0,1), \ j = 1, 2.$$

Because studies that do not fit the assumed model (3.3) tend to yield effect sizes ($\mathbf{y}_i$) which deviate more from $\widehat{\boldsymbol{\mu}}_{i(-i)}$ than would be expected due to the three sources of variability: $\widehat{\boldsymbol{\Sigma}}_{(-i)}$, $\boldsymbol{\Psi}_i$, and $\mathrm{var}(\widehat{\boldsymbol{\mu}}_{(-i)}) = \left(\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)}\right)^{-1}$, the corresponding studentized deleted residuals will become large. Since the studentized deleted residuals are independent and marginally distributed as the standard normal when the assumed model (3.3) is true, the check for outliers can proceed similar to Viechtbauer and Cheung (2010) and Hedges and Olkin (1985). That is, studies with studentized deleted residuals larger than 1.96 in absolute value can be considered as potentially outlying. However, as a solution to the multiple testing issue, we employ the parametric bootstrap to obtain the sampling distribution of the standardized deleted residuals and the corresponding rejection regions, following Gumedze et al. (2010) and Gumedze and Jackson (2011) (See the conclusion of Subsection 3.4.2 for details on the parametric bootstrap approach). Accordingly, up to m studies can be labeled as potentially outlying if the observed standardized deleted residuals of the m studies are larger than the 95-th percentile of the first m ordered bootstrap standardized deleted residuals.

3.4.2 Detecting outlying studies—likelihood ratio test approach

Under this approach, we propose a random-effects bivariate mean shift outlier model (BMSOM) with the aim of identifying outlying studies in meta-analysis of diagnostic test accuracy studies.

Similar to the BNN model, the BMSOM describes the within-study variation in Se and Sp as:

$$\mathbf{y}_i|\boldsymbol{\mu}_i \sim N_2(\boldsymbol{\mu}_i, \boldsymbol{\Psi}_i), \;\; i = 1, ..., k, \tag{3.5}$$

where $\mathbf{y}_i$, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Psi}_i$ are as defined in (3.1).

However, to describe the between-study variability in the study-specific true Se and Sp, the BMSOM assumes a shifted or inflated mean vector for the j-th study, $j = 1, ..., k$. Mathematically, the BMSOM describes the between-study variation as

$$\boldsymbol{\mu}_i \sim N_2(\boldsymbol{\mu} + \boldsymbol{\eta}_j, \boldsymbol{\Sigma}), i = j = 1, ..., k, \tag{3.6}$$

$$\text{where } \boldsymbol{\eta}_j = \begin{cases} (\eta_1, \eta_2)^T, & \text{if } i = j \\ \mathbf{0}, & \text{if } i \neq j, \end{cases}$$

which, upon combining (3.5) and (3.6), leads to the BMSOM marginal model,

$$\mathbf{y}_i \sim N_2(\boldsymbol{\mu} + \boldsymbol{\eta}_j, \boldsymbol{\Sigma}_i), \;\; i = j = 1, ..., k,$$

where $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\eta}_j$ are as defined earlier.

Therefore, by setting $\boldsymbol{\eta}_j = \mathbf{0}$, for $i \neq j$, the log-likelihood of the BMSOM is given by

$$l(\boldsymbol{\mu}, \boldsymbol{\eta}_j, \boldsymbol{\Sigma}) = \text{const.} - \frac{1}{2}\sum_{i=1}^{k} \log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{k}(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\eta}_j)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\eta}_j)$$

where const. is a constant that does not depend on model parameters. As in the BNN model, the parameters of the BMSOM model are estimated iteratively using numerical methods.

To detect outlying studies, we first fit the BMSOM to the meta-analysis data. Then, if the j-th study provides a large $\widehat{\boldsymbol{\eta}}_j$, the magnitude of $\widehat{\boldsymbol{\eta}}_j$ may reflect that the j-th study could potentially be outlying.

However, the shifted mean, $\widehat{\boldsymbol{\eta}}_j$, could be large due to chance alone. Therefore, we can formally test for the significance of the shifted mean vector using the likelihood ratio test (LRT) approach. The LRT statistic for the hypothesis we are interested in,

$$\text{H}_0: \;\; \boldsymbol{\eta}_j = \mathbf{0} \;\text{ vs }\; \text{H}_1: \;\; \boldsymbol{\eta}_j \neq \mathbf{0}, \;\text{ for study } j,$$

is given by

$$\text{LRT}_j = -2\left(l(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) - l(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\eta}}_j, \hat{\boldsymbol{\Sigma}})\right),$$

where $l(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ is the maximized log-likelihood of the BNN model, which is true under H_0, and $l(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\eta}}_j, \hat{\boldsymbol{\Sigma}})$ is the maximized log-likelihood of the BMSOM, which is true under H_1.

Under the null hypothesis, the asymptotic distribution of LRT_j is a chi-square distribution with two degrees of freedom, i.e. $\text{LRT}_j \sim \chi^2(2)$. However, following Gumedze et al. (2010) and Gumedze and Jackson (2011), we propose the parametric bootstrap to get the sampling distribution of LRT_j under H_0 and the corresponding critical value (the 95-th percentile of the first m studies) as a solution to the problem of multiple testing.

The following steps have been followed to carry out the parametric bootstrap approach.

1. Fit the null BNN model to the observed DTA data and obtain $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$.

2. Generate new DTA data using the $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ obtained from step 1 following the procedure described by Negeri et al. (2018). That is, generate k logit-transformed Se and Sp pairs from the bivariate normal distribution using $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$; randomly simulate the study-specific sample sizes from a Poisson distribution, and, the study-specific number of positive and negative test results from the Binomial distribution.

3. Fit both the BNN model and BMSOM to the newly generated DTA data from

Step 2, compute LRT_j, $j = 1, ..., k.$, and sort them in decreasing order.

4. For a fixed bootstrap size B, repeat Steps 2 and 3 to generate the empirical distribution of the bootstrap LRT test statistic, LRT_j^b, $b = 1, ..., B$, $j = 1, ..., k$.

5. Finally, compute the 95-th percentile of the first m ordered bootstrap likelihood ratio test statistics, LRT_j^b, which can be used as critical values to conclude up to m studies as outlying (Gumedze and Jackson, 2011; Gumedze et al., 2010).

Therefore, if the observed likelihood ratio test statistics of m studies are bigger than the 95-th percentile of the first m ordered bootstrap likelihood ratio test statistics, the m studies can be categorized as potentially outlying.

3.4.3 Detecting influential studies

All outlying studies are not necessarily influential. A study can be considered influential if removing that study from the meta-analysis significantly changes results of the fitted model. We propose two indexes which assist in identifying whether a study is influential or not.

The impact of the i-th study can be examined by the amount of relative change in the variance-covariance matrix of the average effect size when removing the i-th study from the meta-analysis compared to when all studies are included (Viechtbauer and Cheung, 2010; Belsley et al., 1980). To examine this effect, we propose the VARCOVRATIO index as

$$\begin{aligned}\text{VARCOVRATIO}_i &= \frac{\det(\text{VARCOV}_{(-i)})}{\det(\text{VARCOV})} \\ &= \frac{\det\left((\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1}\right)}{\det\left((\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i})^{-1}\right)}.\end{aligned}$$

$\text{VARCOVRATIO}_i \in (0, \infty)$, and it gets closer to 0 if study i is potentially influential (i.e. excluding study i increases precision), whereas, it tends to ∞ if the ith study is not influential (i.e., removing study i decreases precision, which is an undesirable property in meta-analysis). And an VARCOVRATIO_i equal to or closer to one indicates that removing study i has no effect on the precision by which the overall effect size is estimated, thus, is not influential.

VARCOVRATIO_i measures the relative change in the variance-covariance matrix of the overall effect size when study i is excluded from the meta-analysis compared to when all studies are included. Therefore, a VARCOVRATIO_i below one indicates that exclusion of study i decreases the overall variance-covariance matrix, or equivalently, removing study i from the meta-analysis yields more precise overall effect size estimates. Similarly, a less than one VARCOVRATIO_i value indicates that including study i in the meta-analysis reduces precision. Therefore, a study with VARCOVRATIO_i value of less than one can be regarded as potentially influential.

The influence of the i-th study can also be assessed by the amount of relative change in the heterogeneity parameter when study i is removed compared to when all studies are included in the meta-analysis. The second index, SIGMARATIO_i determines the relative change in the amount of heterogeneity when study i is removed

from the meta-analysis relative to when all studies are included. SIGMARATIO is defined as

$$\text{SIGMARATIO}_i = \frac{\det(\widehat{\boldsymbol{\Sigma}}_{(-i)})}{\det(\widehat{\boldsymbol{\Sigma}})}.$$

$\text{SIGMARATIO}_i \in (0, \infty)$, and it goes to 0 if the ith study is potentially influential (i.e. excluding study i decreases the amount of heterogeneity), whereas, it raises to ∞ if the study i is not influential (i.e., removing study i increases heterogeneity between-studies, which is an unwanted characteristic in meta-analysis). And an SIGMARATIO_i equal to or closer to one signifies that removing the ith study has no effect on the amount of heterogeneity, therefore, is not influential.

A $\text{SIGMARATIO}_i < 1$ indicates that study i could potentially be influential because removing study i from the meta-analysis decreases the amount of heterogeneity among the studies, or equivalently, including study i increases the amount of heterogeneity. We propose a parametric bootstrap to obtain the empirical distribution of the two proposed indexes and the corresponding critical values (the 95-th percentile of the first m order statistics) as discussed in the previous subsection.

3.5 Illustrative examples

In this section, we illustrate the proposed methods of detecting outlying and influential studies using the two datasets introduced in Section 3.3. The two data are

selected to demonstrate our methods because the data are different in terms of number of studies, sample size, number of outlying and influential studies, and the pooled sensitivity and specificity.

3.5.1 US-Children data

When the BNN model is fitted to the *US-Children* data, the estimated pooled sensitivity and specificity were 86.4% and 94.3%, respectively. And the proportion of variability attributed to the between-study heterogeneity parameter as quantified by I^2 statistic (Jackson et al., 2012; Gasparrini et al., 2012) was 73.2% (p-value < 0.001). Figure 3.3 shows the observed standardized deleted residuals and the corresponding 95-th percentile of the first three order statistics for the *US-Children* data. As seen in the forest plot in Figure 3.1, study 13 is detected as outlying in Se. Additionally, study 1, 6, 8, 12 and 19 are also identified as potentially outlying according to the residual-based approach.

When the likelihood ratio test-based approach is applied to the *US-Children* data (Figure 3.4), similar to the result based on forest plots, only study 13 is detected as potentially outlying.

Figure 3.5 displays methods used to detecting influential studies. The VARCOVRATIO$_i$ suggests that out of the six potentially outlying studies detected by the residual-based approach, five of them: study 1, 6, 8, 12 and 13 are potentially influential. Similarly, the SIGMARATIO$_i$ index identifies study 1, 6, 12 and 13 as influential studies.

We now study the impact of these five influential studies on the pooled Se, pooled Sp and the heterogeneity parameter. We present the impact as a percentage change

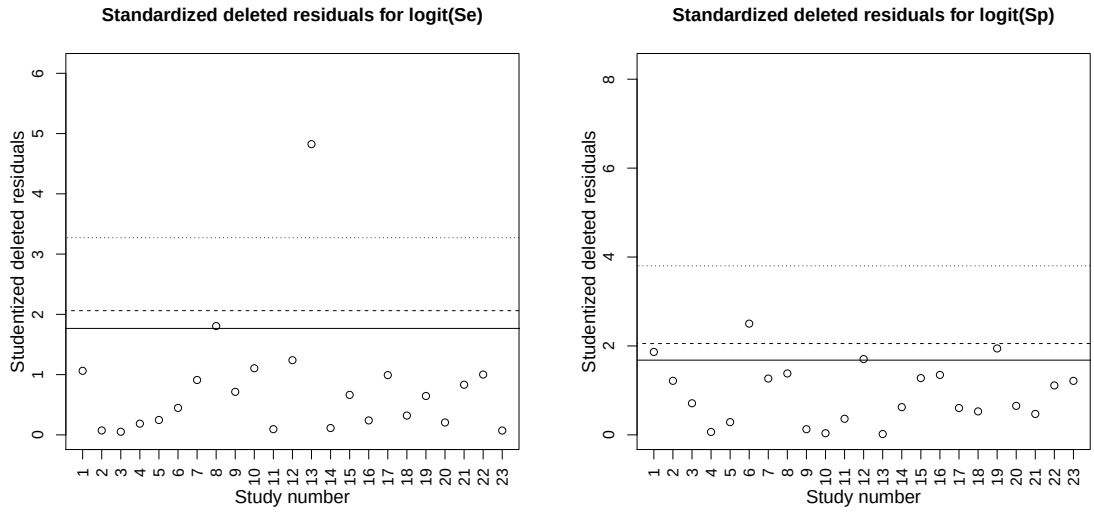

Figure 3.3: Standardized deleted residuals for logit-transformed Se (left) and Sp (right) of the *US-Children* data. Dotted, dashed and solid lines represent the 95-th percentile of the first-, second- and third-order statistics of the standardized deleted residuals, respectively.

and as an (absolute change) as deemed necessary. Removing study 13 increases the pooled Se by 1.04% (0.9), and removing studies 1, 6, 8 and 12 has a combining effect of reducing the pooled Se by 1.85% (-1.6). This results reveal that including study 13, which has very small observed Se, pulls the overall Se towards itself that it yields an underestimated average Se. On the other hand, including the other four influential studies, 1, 6, 8 and 12, which have large study-specific Se values, produces an overestimated pooled Se by pulling it towards themselves. Moreover, removing the influential studies reduces the I^2 statistic by about 19%.

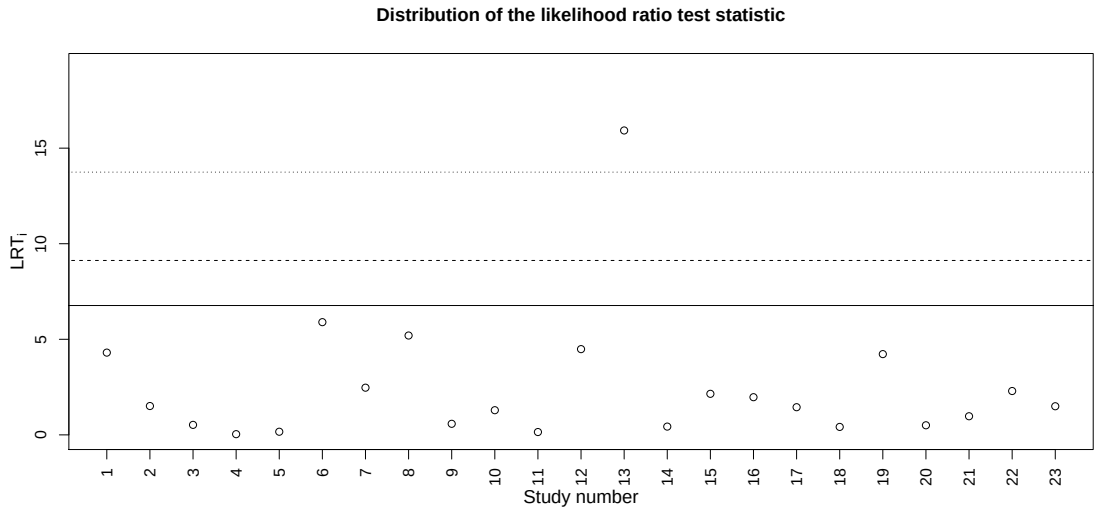

Figure 3.4: Likelihood ratio test statistics and its 95-th percentile for the *US-Children* data. Dotted, dashed and solid lines represent the 95-th percentile of the first-, second- and third-order statistics of the likelihood ratio test statistic, respectively.

3.5.2 AUDIT-C data

For the *AUDIT-C* data, the estimated pooled Se and Sp are 89.9% and 78.0%, respectively; and the I^2 statistic for heterogeneity is 98.8% (p-value < 0.001). According to the residual-based approach (Figure 3.6), study 7, 8 and 14 are identified as potentially outlying in Se and study 3, 8, and 14 are detected as outlying studies in Sp.

Figure 3.7 shows results of the likelihood ratio test-based approach for the *AUDIT-C* dataset. According to the LRT-based approach, study 3 and 8 are identified as potentially outlying studies.

Furthermore, the three outlying studies, study 3, 8 and 14, are also detected

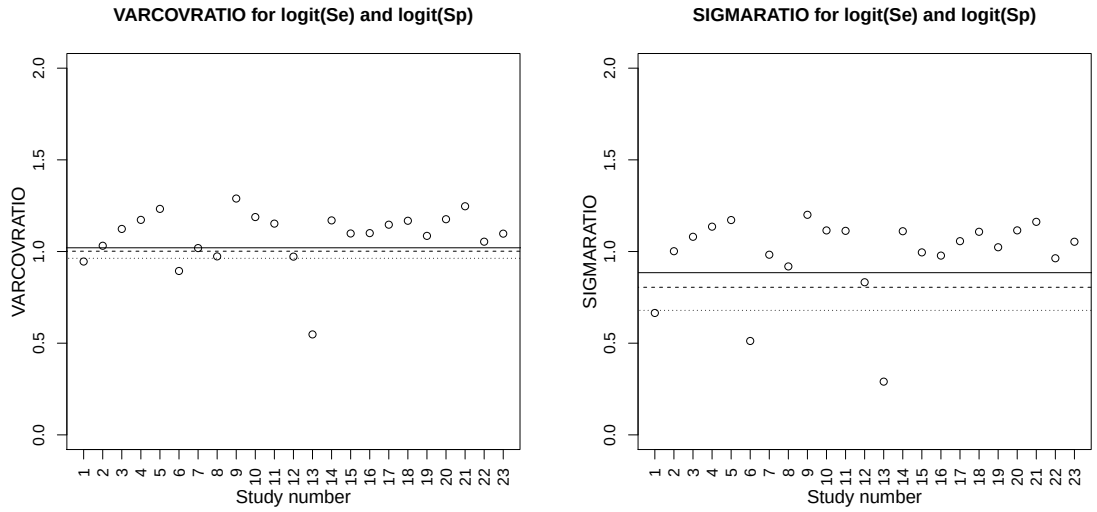

Figure 3.5: VARCOVRATIO$_i$ (left) and SIGMARATIO$_i$ (right) indexes for the *US-Children* data. Dotted, dashed and solid lines represent the 95-th percentile of the first-, second- and third-order statistics of the indexes, respectively.

as potentially influential by both the VARCORATIO$_i$ and SIGMARATIO$_i$ indexes (Figure 3.8).

Again the impact of removing the outlying or influential studies is presented as a percentage change (absolute change). Study 3 influences the overall Sp by pulling towards itself and overestimating it. When the study is excluded from the meta-analysis, it reduces the overall Sp by 2.05% (-1.6). Removing study 8 results in a reduction of the overall Se by 3.45% (-3.1), an increase in average Sp by 2.31% (1.8), and reduction in the I^2 statistic by 9.82%. Study 14 had the least influence as its exclusion from the meta-analysis increases the pooled Se and Sp by 0.89% (0.8) and 0.26% (0.2), respectively. Excluding study 8 and 14, the two studies identified as

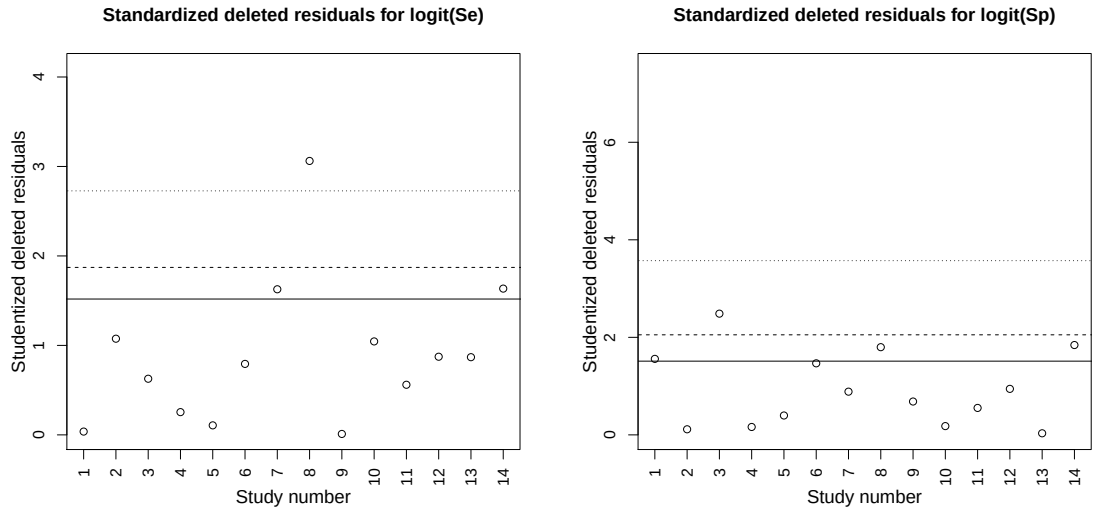

Figure 3.6: Standardized deleted residuals for logit-transformed Se (left) and Sp (right) of the *AUDIT-C* data. Dotted, dashed and solid lines represent the 95-th percentile of the first-, second- and third-order statistics of the standardized deleted residuals, respectively.

outlying in Se, decreases the pooled Se by 2.6% (-2.3). And removing the two studies detected as outlying in Sp, study 3 and 14, decreases the pooled Sp by 1.9% (-1.5).

3.6 Simulation study

3.6.1 Simulation design

We have also performed a small-scale simulation study to compare the performance of the two proposed methods of identifying outlying and influential studies, by mimicking the two datasets discussed in Section 3.3.

We simulated outlying studies (studies with large or small mean vector than

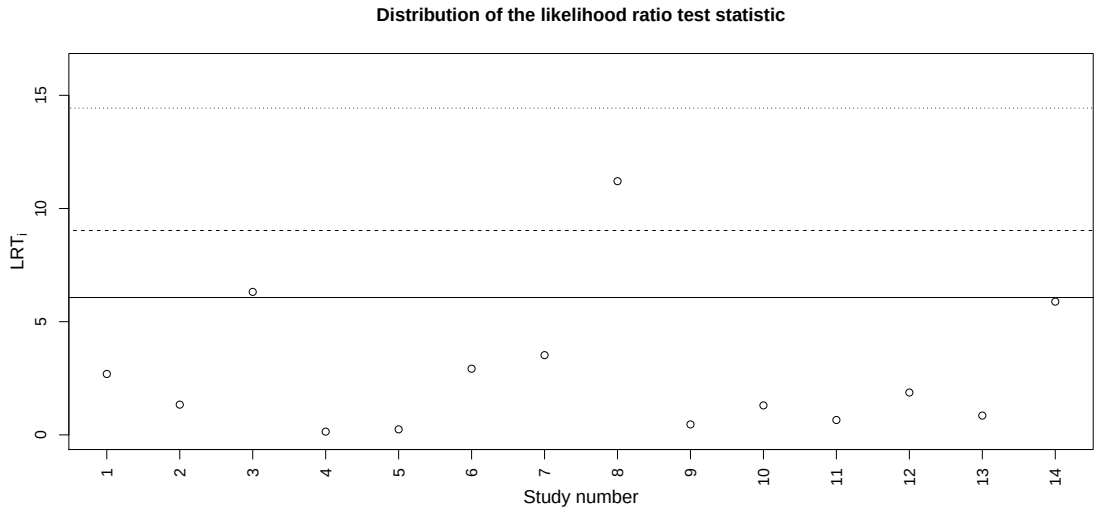

Figure 3.7: Likelihood ratio test statistics and its 95-th percentile for the *AUDIT-C* data. Dotted, dashed and solid lines represent the 95-th percentile of the first-, second- and third-order statistics of the likelihood ratio test statistic, respectively.

the rest of the studies) and influential studies (studies with an inflated between-study covariance matrix), and designed our simulation according to the following four settings where data are generated as described in Negeri et al. (2018) using either

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 0.5 & -0.2 \\ -0.2 & 0.6 \end{pmatrix} \text{ or } \boldsymbol{\Sigma}_2 = \begin{pmatrix} 4.5 & -0.2 \\ -0.2 & 6.0 \end{pmatrix}$$

as the between-study covariance parameter.

1. Setting 1: Randomly generate meta-analysis of diagnostic test accuracy studies with the parameters $(Se, Sp) = (0.6, 0.7)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$, $n_1 = 100$, $n_2 = 200$ and

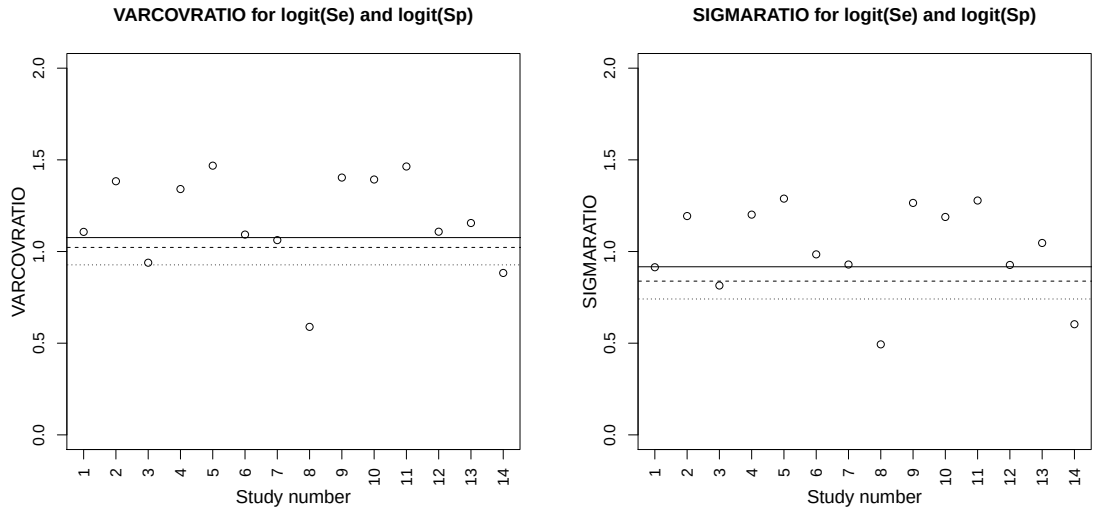

Figure 3.8: VARCOVRATIO$_i$ (left) and SIGMARATIO$_i$ (right) indexes for the *AUDIT-C* data. Dotted, dashed and solid lines represent the 95-th percentile of the first-, second- and third-order statistics of the indexes, respectively.

$k = 14$. Next, randomly simulate three outlying and influential studies with the parameters $(Se, Sp) = (0.95, 0.95)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$, $n_1 = 100$, $n_2 = 200$ and $k = 14$, and, substitute them for the first three studies generated above. This setting represents when outlying studies are defined as studies with large mean vector when $k = 14$.

2. Setting 2: Randomly simulate meta-analysis of diagnostic test accuracy studies with the parameters $(Se, Sp) = (0.95, 0.95)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$, $n_1 = 100$, $n_2 = 200$ and $k = 14$. Then, randomly generate three outlying and influential studies with the parameters $(Se, Sp) = (0.6, 0.7)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$, $n_1 = 100$, $n_2 = 200$ and $k = 14$, and, replace them with the first three studies generated above. This setting

indicates when outlying studies are defined as studies with small mean vector when $k = 14$.

3. Setting 3: Randomly draw meta-analysis of diagnostic test accuracy studies with the parameters $(Se, Sp) = (0.6, 0.7)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$, $n_1 = 100$, $n_2 = 200$ and $k = 23$. Next, randomly simulate three outlying and influential studies with the parameters $(Se, Sp) = (0.95, 0.95)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$, $n_1 = 100$, $n_2 = 200$ and $k = 23$, and, substitute them for the first three studies generated above. This setting represents when outlying studies are defined as studies with large mean vector when $k = 23$.

4. Setting 4: Randomly generate meta-analysis of diagnostic test accuracy studies with the parameters $(Se, Sp) = (0.95, 0.95)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$, $n_1 = 100$, $n_2 = 200$ and $k = 23$. Then, randomly generate three outlying and influential studies with the parameters $(Se, Sp) = (0.6, 0.7)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$, $n_1 = 100$, $n_2 = 200$ and $k = 23$, and, replace them with the first three studies generated above. This setting indicates when outlying studies are defined as studies with small mean vector when $k = 23$.

3.6.2 Simulation results

We replicated the simulation 100 times since the methods are computationally intensive and compared the performance of the residual-based and LRT-based approaches under the four simulation settings. Performance of the methods for identifying influential studies have also been assessed.

Table 3.1: Proportion of times the residual-based and likelihood ratio test-based approaches identified the three outlying studies correctly as outlying for the four simulation scenarios.

		Methods of identifying outlying studies	
Scenario	No. of outlying studies	Residual-based	LRT-based
1	0	0.01	0.17
	1	0.36	0.57
	2	0.57	0.26
	3	0.06	0.00
2	0	0.80	0.09
	1	0.20	0.48
	2	0.00	0.42
	3	0.00	0.01
3	0	0.00	0.02
	1	0.25	0.45
	2	0.57	0.45
	3	0.18	0.08
4	0	0.89	0.03
	1	0.11	0.35
	2	0.00	0.54
	3	0.00	0.08

Table 3.1 displays the frequency and proportion of outlying studies out of the three outlying studies systematically introduced in the simulation as identified by the residual-based and LRT-based approaches. For Scenarios 1 and 3, the residual-based approach correctly detected at least one of the three outlying studies 99% and 100% of the time, respectively; for the same scenario, the LRT-based approach correctly predicted at least one of the three outlying studies 83% and 98% times. In both scenarios, the residual-based approach correctly identified two of the outlying studies 57% of the time, whereas, the LRT-based approach correctly detected two of the three outlying studies 26% and 45% times, respectively. All three outlying

studies are truly identified as outlying 6% of the time in Scenario 1 and 18% of the time in Scenario 3 by the residual-based approach. The LRT-based approach correctly detected all three outlying studies 0% and 8% of the time in Scenarios 1 and 3, respectively. This result indicates that the performance of both approaches of identifying outlying studies improve as the number of studies increases and when outlying studies are defined as studies with large mean vector from the remaining studies.

Table 3.2: Proportion of times the VARCOVRATIO and SIGMARATIO indexes identified the three influential studies correctly as influential for the four simulation scenarios.

		Methods of identifying influential studies	
Scenario	No. of influential studies	VARCOVRATIO	SIGMARATIO
1	0	0.00	0.00
	1	0.05	0.07
	2	0.51	0.54
	3	0.44	0.39
2	0	0.00	0.00
	1	0.17	0.18
	2	0.59	0.59
	3	0.24	0.23
3	0	0.00	0.00
	1	0.04	0.07
	2	0.41	0.43
	3	0.55	0.50
4	0	0.00	0.00
	1	0.16	0.16
	2	0.56	0.53
	3	0.28	0.31

For scenarios 2 and 4, the residual-based method correctly detected one of the

three outlying studies 20% and 11% of the time, respectively. The LRT-based approach truly identified one of the three outlying studies 48% and 35% times. For the same scenario, at least two out of the three outlying studies are correctly identified as outlying 0% and 0%, and, 43% and 62% of the time by the residual-based and LRT-based approaches, respectively. And three of the outlying studies were wrongly classified as non-outlying 80% and 89% of the time, respectively, by the residual-based approach; while the proportion of misclassification was only 9% and 3%, respectively, for the LRT-based approach.

These results reveal that the residual-based approach performs poorly when outlying studies are defined as studies with smaller mean vector than the remaining studies in the meta-analysis. When outlying studies are defined like that, this happens because the deleted raw residual ($\boldsymbol{\epsilon}_{i(-i)}$) becomes a small negative number resulting in a smaller negative deleted standardized residuals (3.4) which will be much smaller than the -1.96 critical value. Furthermore, performance of the LRT-based approach gets better as the number of studies increases while that of the residual-based approach worsen.

Results of the methods for identifying influential studies are given in Table 3.2. Neither the SIGMARATIO nor VARCOVRATIO index misclassified the three influential studies as non-influential. The VARCOVRATIO and SIGMARATIO index correctly identified at least two of the three influential studies as influential 95% and 93%, 83% and 82%, 96% and 93%, and, 84% and 84% times in scenario 1, 2, 3 and 4, respectively. This result tells that the two indexes are equally effective and robust to the simulation scenarios.

3.7 Discussion

The aim of this chapter is to propose methods for identifying outlying and influential studies which are based on strong statistical foundation unlike the ad hoc approaches currently being used in meta-analysis of diagnostic test accuracy studies literature. Contrary to the subjective approaches, the proposed methods consider both the within- and between-study variation in Se and Sp when deciding whether a study is outlying or influential. Moreover, since the proposed methods are based on statistical reasoning, they bypass the subjectivity of the ad hoc approaches.

The proposed methods can be generalized to multivariate meta-analysis and the residual-based approach of detecting outlying studies will reduce to the univariate method of Viechtbauer and Cheung (2010) when the covariance matrix is set to zero. In implementing our methods, we proposed the parametric bootstrap approach to deal with the issue of multiple testing following Gumedze et al. (2010) and Gumedze and Jackson (2011).

Two published meta-analysis data and a simulation study are used to demonstrate and validate the proposed methods. Accordingly, the residual-based and LRT-based approaches of detecting outlying studies yielded similar results, however, as demonstrated in our simulations, the residual-based approach is more effective at detecting outlying studies that have large mean vector and the LRT-based approach is preferable at detecting outlying studies with small mean vector. SIGMARATIO and VARCOVRATIO indexes produced comparable results when detecting influential studies.

Although we created awareness on the prevalence of outlying and influential studies in meta-analysis of DTA studies and proposed sound statistical methods to identify them, we do not, however, promote the removal of outlying or influential studies from the meta-analysis since we might lose some useful information. But we are noting that when outlying or influential studies are present, inferences and conclusions drawn from the meta-analysis may mislead clinicians or policymakers who rely on evidence-based medical or policy decision making. Motivated by the results of this chapter, in chapter 4, we propose a bivariate random-effects meta-analysis model that robustly synthesizes diagnostic test accuracy data without removing the outlying or influential studies.

In conclusion, we recommend that the residual-based approach should be used to screen for outlying studies when studies with large Se and Sp are suspected as potentially outlying, and, the likelihood ratio test-based approach is ideal for detecting studies with small Se and Sp as potential outliers. Either of the proposed methods of identifying influential studies can be used effectively. Our methods can be used as a sensitivity analysis tool concurrently with the current bivariate random-effects model or as a preliminary analysis tool for robust models that accommodate outlying and/or influential studies. Although the proposed methods are solely based on the usual binary (no disease, disease) diagnostic test assumption, a future work would aim to study the performance of the methods in the context of the three-way (e.g., no disease, mild disease, advanced disease) diagnostic design as in, for example, Inácio de Carvalho et al. (2018). More future work might investigate ways to quantify uncertainty by assigning the probability of actually being outlying to each study

instead of dichotomizing the state of the studies. One of the potential approaches this could be achieved is to use the Bayesian approach of modeling as in Zhang et al. (2015) or to fit a mixture model from a Frequentist approach as in, for example, Beath (2014).

Appendix II

Since $\mathbf{y}_i$ is involved in the calculation of $\widehat{\boldsymbol{\mu}}$, the variance of the i-th raw residual, $\boldsymbol{\epsilon}_i$, is given by

$$
\begin{aligned}
\text{var}(\boldsymbol{\epsilon}_i) &= \text{var}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}) = \text{var}(\mathbf{y}_i) + \text{var}(\widehat{\boldsymbol{\mu}}) - 2\text{cov}(\mathbf{y}_i, \widehat{\boldsymbol{\mu}}) \\
&= \widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i + (\sum_{i=1}^{k}(\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i)^{-1})^{-1} - 2\text{cov}(\mathbf{y}_i, (\sum_{i=1}^{k}(\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i)^{-1})^{-1} \sum_{i=1}^{k}(\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i)^{-1}\mathbf{y}_i) \\
&= \widehat{\boldsymbol{\Sigma}}_i + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} - 2\text{cov}(\mathbf{y}_i, (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} \sum_{i=1}^{k}\widehat{\mathbf{W}}_i\mathbf{y}_i) \\
&= \widehat{\boldsymbol{\Sigma}}_i + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} - 2(\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1}\text{cov}(\mathbf{y}_i, \sum_{i=1}^{k}\widehat{\mathbf{W}}_i\mathbf{y}_i)(\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-T} \\
&= \widehat{\boldsymbol{\Sigma}}_i + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} - 2(\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} \sum_{i=1}^{k}\widehat{\mathbf{W}}_i\text{cov}(\mathbf{y}_i, \mathbf{y}_i)\widehat{\mathbf{W}}_i^{T}(\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-T} \\
&= \widehat{\boldsymbol{\Sigma}}_i + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} - 2(\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} \sum_{i=1}^{k}\widehat{\mathbf{W}}_i\widehat{\mathbf{W}}_i^{-1}\widehat{\mathbf{W}}_i^{T}(\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-T} \\
&= \widehat{\mathbf{W}}_i^{-1} + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} - 2(\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1} \\
&= \widehat{\mathbf{W}}_i^{-1} - (\sum_{i=1}^{k}\widehat{\mathbf{W}}_i)^{-1}, \text{ as required.}
\end{aligned}
$$

Now because $\mathbf{y}_i$ is not involved in the computation of $\widehat{\boldsymbol{\mu}}_{i(-i)}$ and thus $\mathbf{y}_i$ and $\widehat{\boldsymbol{\mu}}_{i(-i)}$ are uncorrelated, variance of the deleted residuals is given by

$$\begin{aligned}
\text{var}(\boldsymbol{\epsilon}_{i(-i)}) &= \text{var}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_{i(-i)}) = \text{var}(\mathbf{y}_i) + \text{var}(\widehat{\boldsymbol{\mu}}_{i(-i)}) \\
&= \text{var}(\mathbf{y}_i) + \text{var}((\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1} \sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)} \mathbf{y}_{i(-i)}) \\
&= \widehat{\boldsymbol{\Sigma}}_{i(-i)} + (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1} \sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)} \text{var}(\mathbf{y}_{i(-i)}) \widehat{\mathbf{W}}_{i(-i)} (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1} \\
&= \widehat{\boldsymbol{\Sigma}}_{i(-i)} + (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1} \sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)} \widehat{\mathbf{W}}_{i(-i)}^{-1} \widehat{\mathbf{W}}_{i(-i)} (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1} \\
&= \widehat{\boldsymbol{\Sigma}}_{i(-i)} + (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1} \\
&= \widehat{\mathbf{W}}_{i(-i)}^{-1} + (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1},
\end{aligned}$$

where $\widehat{\mathbf{W}}_{i(-i)} = (\widehat{\boldsymbol{\Sigma}}_{i(-i)} + \boldsymbol{\Psi}_i)^{-1}$, $i = 1, ..., k$.

To show that the studentized deleted residuals are distributed as the standard bivariate normal distribution, note that

$$\mathbf{t}_i = \left(\text{var}(\boldsymbol{\epsilon}_{i(-i)})\right)^{-\frac{1}{2}} \boldsymbol{\epsilon}_{i(-i)} \sim N_2[\mathbb{E}\{\left(\text{var}(\boldsymbol{\epsilon}_{i(-i)})\right)^{-\frac{1}{2}} \boldsymbol{\epsilon}_{i(-i)}\}, \text{var}\{\left(\text{var}(\boldsymbol{\epsilon}_{i(-i)})\right)^{-\frac{1}{2}} \boldsymbol{\epsilon}_{i(-i)}\}]$$

since $\boldsymbol{\epsilon}_{i(-i)}$'s are a linear combination of a bivariate normally distributed random vector, $\mathbf{Y}_i$. Now we show that

$$\begin{aligned}
\mathbb{E}(\mathbf{t}_i) &= \mathbb{E}\left[\left(\text{var}(\boldsymbol{\epsilon}_{i(-i)})\right)^{-\frac{1}{2}} \boldsymbol{\epsilon}_{i(-i)}\right] \\
&= \mathbb{E}\left[\left(\widehat{\mathbf{W}}^{-1}_{i(-i)} + (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1}\right)^{-\frac{1}{2}} \boldsymbol{\epsilon}_{i(-i)}\right] \\
&= \left(\widehat{\mathbf{W}}^{-1}_{i(-i)} + (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1}\right)^{-\frac{1}{2}} \mathbb{E}(\boldsymbol{\epsilon}_{i(-i)}) \\
&= \left(\widehat{\mathbf{W}}^{-1}_{i(-i)} + (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1}\right)^{-\frac{1}{2}} \mathbb{E}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_{i(-i)}) \\
&= \left(\widehat{\mathbf{W}}^{-1}_{i(-i)} + (\sum_{i=1}^{k} \widehat{\mathbf{W}}_{i(-i)})^{-1}\right)^{-\frac{1}{2}} \mathbf{0} \\
&= \mathbf{0},
\end{aligned}$$

and

$$
\begin{aligned}
\text{var}(\mathbf{t}_i) &= \text{var}\left[\left(\text{var}(\boldsymbol{\epsilon}_{i(-i)})\right)^{-\frac{1}{2}}\boldsymbol{\epsilon}_{i(-i)}\right] \\
&= \text{var}\left[\left(\widehat{\mathbf{W}}^{-1}_{i(-i)} + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_{i(-i)})^{-1}\right)^{-\frac{1}{2}}\boldsymbol{\epsilon}_{i(-i)}\right] \\
&= \left(\widehat{\mathbf{W}}^{-1}_{i(-i)} + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_{i(-i)})^{-1}\right)^{-\frac{1}{2}}\text{var}(\boldsymbol{\epsilon}_{i(-i)})\left((\sum_{i=1}^{k}\widehat{\mathbf{W}}_{i(-i)})^{-1} + \widehat{\mathbf{W}}^{-1}_{i(-i)}\right)^{-\frac{1}{2}} \\
&= \left(\widehat{\mathbf{W}}^{-1}_{i(-i)} + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_{i(-i)})^{-1}\right)^{-\frac{1}{2}}\left(\widehat{\mathbf{W}}^{-1}_{i(-i)} + (\sum_{i=1}^{k}\widehat{\mathbf{W}}_{i(-i)})^{-1}\right) \\
&\times \left((\sum_{i=1}^{k}\widehat{\mathbf{W}}_{i(-i)})^{-1} + \widehat{\mathbf{W}}^{-1}_{i(-i)}\right)^{-\frac{1}{2}} \\
&= \mathbf{I}_2.
\end{aligned}
$$

Therefore, $\mathbf{t}_i \sim N_2(\mathbf{0}, \mathbf{I}_2)$, or equivalently, $\text{t}_{ij} \sim N(0,1)$, $i = 1, ..., k$, $j = 1, 2$, as desired.

Bibliography

Baker, R. and Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health Care Management Science*, 11(2):121–131.

Baker, R. and Jackson, D. (2016). New models for describing outliers in meta-analysis. *Research Synthesis Methods*, 7(3):314–328.

Beath, K. J. (2014). A finite mixture method for outlier detection and robustness in meta-analysis. *Research Synthesis Methods*, 5(4):285–293.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics.* New York: Wiley.

Chen, Y., Liu, Y., Ning, J., Nie, L., Zhu, H., and Chu, H. (2017). A composite likelihood method for bivariate meta-analysis in diagnostic systematic reviews. *Statistical Methods in Medical Research*, 26(2):914–930.

Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*, 59(12):1331.

de Jesus, J. O., Parker, L., Frangos, A. J., and Nazarian, L. N. (2009). Accuracy of MRI, MR Arthrography, and Ultrasound in the Diagnosis of Rotator Cuff Tears: A Meta-Analysis. *American Journal of Roentgenology*, 192:1701–1707.

Devillé, W. L., Buntinx, F., Bouter, L. M., Montori, V. M., De Vet, H. C., Van der

Windt, D. A., and Bezemer, P. D. (2002). Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Medical Research Methodology*, 2(1):9.

Doria, A. S., Moineddin, R., Kellenberger, C. J., and et al. (2006). US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis. *Radiology*, 241(1):83–94.

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2012). Multivariate meta-analysis for non-linear and other multi-parameter associations. *Statistics in Medicine*, 31(29):3821–3839.

Gumedze, F. N. and Jackson, D. (2011). A random effects variance shift model for detecting and accommodating outliers in meta-analysis. *BMC Medical Research Methodology*, 11(1):19.

Gumedze, F. N., Welham, S. J., Gogel, B. J., and Thompson, R. (2010). A variance shift model for detection of outliers in the linear mixed model. *Computational Statistics and Data Analysis*, 54:2128–2144.

Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P., and Sterne, J. A. (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, 8(2):239–51.

Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Aanalysis*. New York: Academic Press.

Inácio de Carvalho, V., de Carvalho, M., and Branscum, A. J. (2018). Bayesian Bootstrap Inference for the receiver operating characteristic surface. *Stat*, 7:e211.

Jackson, D., White, I. R., and Riley, R. D. (2012). Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Statistics in Medicine*, 31(29):3805–3820.

Kriston, L., HÖlzel, L., Weiser, A. K., Berner, M. M., and Härter, M. (2008). Meta-analysis: Are 3 Questions Enough to Detect Unhealthy Alcohol Use? *Annals of Internal Medicine*, 149:879–888.

Kuss, O., Hoyer, A., and Solms, A. (2014). Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. *Statistics in Medicine*, 33(1):17–30.

Negeri, Z. F., Shaikh, M., and Beyene, J. (2018). Bivariate Random-Effects Models for Meta-analysis of Diagnostic Test Accuracy Studies Using Arcsine-Based Transformations. *Biometrical Journal*, 60(4):827–844.

Nikoloulopoulos, A. K. (2015). A mixed effect model for bivariate meta-analysis of diagnostic test accuracy studies using a copula representation of the random effects distribution. *Statistics in Medicine*, 34(29):3842–3865.

Nikoloulopoulos, A. K. (2017). A vine copula mixed effect model for trivariate meta-analysis of diagnostic test accuracy studies accounting for disease prevalence. *Statistical Methods in Medical Research*, 26(5):2270–2286.

Nikoloulopoulos, A. K. (2018). Hybrid copula mixed models for combining case-control and cohort studies in meta-analysis of diagnostic tests. *Statistical Methods in Medical Research*, 27(8):2540–2553.

Nyaga, V. N., Aerts, M., and Arbyn, M. (2018a). ANOVA model for network meta-analysis of diagnostic test accuracy data. *Statistical Methods in Medical Research*, 27(6):1766–1784.

Nyaga, V. N., Arbyn, M., and Aerts, M. (2018b). Beta-binomial analysis of variance model for network meta-analysis of diagnostic test accuracy data. *Statistical Methods in Medical Research*, 27(8):2554–2566.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction.* New York: Oxford University Press.

Petignat, P., Faltin, D. L., and Bruchim, I. (2007). Are self-collected samples comparable to physician-collected cervical specimens for human papillomavirus DNA testing? A systematic review and metaanalysis. *Gynecologic Oncology*, 105:530–535.

Pormohammad, A., Riahi, S. M., Nasiri, M. J., and et al. (2017). Diagnostic test accuracy of adenosine deaminase for tuberculous meningitis: A systematic review and meta-analysis. *Journal of Infection*, 74:545–554.

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58(10):982–990.

Rutter, C. M. and Gatsonis, C. A. (2001). A hierarchical regression approach

to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, 20(19):2865–2884.

Singal, A., Volk, M. L., Waljee, and et al. (2009). Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Alimentary Pharmacology & Therapeutics*, 30:37–47.

Viechtbauer, W. and Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2):112–125.

Zhang, J., Fu, H., and Carlin, B. P. (2015). Detecting outlying trials in network meta-analysis. *Statistics in Medicine*, 34(19):2695–2707.

Zhou, Y., Wang, C., Zhu, H., Lin, Y., Pan, B., Zhang, X., and et al. (2016). Diagnostic Accuracy of PIK3CA Mutation Detection by Circulating Free DNA in Breast Cancer: A Meta-Analysis of Diagnostic Test Accuracy. *PLoS ONE*, 11(6).

Chapter 4

Robust bivariate random-effects model for accommodating outlying and influential studies in meta-analysis of diagnostic test accuracy studies

4.1 Abstract

Due to the inevitable correlation between test sensitivity (Se) and test specificity (Sp), hierarchical or bivariate random-effects models are widely used to perform a

meta-analysis of diagnostic test accuracy studies. Conventionally, these models assume that the random-effects follow the bivariate normal distribution. However, the inference made using the well-established bivariate random-effects models, when outlying and influential studies are present, may lead to misleading conclusions, since outlying or influential studies can influence highly parameter estimates due to their disproportional weight. Therefore, we develop a new robust bivariate random-effects model which accommodates outlying and influential observations and gives robust statistical inference by down-weighting the effect of outlying and influential studies. The marginal model and the Monte Carlo expectation-maximization (MCEM) algorithm for our proposed model are derived. A simulation study is carried out to validate the proposed method and compare it against the standard methods. Regardless of the parameters varied in our simulations, the proposed model produces robust point estimates of Se and Sp compared to the standard models. Moreover, our proposed model results in more precise estimates as it yielded the narrowest confidence intervals. The proposed model also generates similar point and interval estimates of Se and Sp as the standard models when there are no outlying and influential studies. Two published meta-analyses are used to illustrate the methods.

Keywords: Influential studies, Meta-analysis, Outlying studies, Robust random-effects model, Sensitivity, Specificity

4.2 Introduction

A meta-analysis of diagnostic test accuracy (DTA) allows for the synthesis of diagnostic test characteristics, such as test Se and Sp, across several independent studies. Due to the inevitable correlation between Se and Sp, hierarchical random-effects model (Rutter and Gatsonis, 2001) or bivariate random-effects models (Reitsma et al., 2005; Chu and Cole, 2006) are widely used to perform meta-analysis of DTA studies. Conventionally, these models assume that the random-effects, which describe the variation among the true study-specific sensitivities and specificities, follow the bivariate normal distribution. However, in the presence of outlying and/or influential studies, this assumption may not adequately describe the variation in true Se and Sp across studies. Moreover, due to their weights, outlying and/or influential studies have the potential to greatly influence the estimated overall Se and Sp. Therefore, simply fitting standard bivariate random-effects models when outlying and/or influential studies are present may lead to misleading inferences (Negeri and Beyene, 2019).

In the conventional meta-analysis of intervention studies, a substantial number of research articles (Baker and Jackson, 2008; Lee and Thompson, 2008; Viechtbauer and Cheung, 2010; Gumedze and Jackson, 2011; Beath, 2014; Baker and Jackson, 2016) which advocate for the use of robust statistical methods in the presence of outlying or influential studies have been published. Methods of identifying outlying and/or influential studies had been proposed independently by Viechtbauer and Cheung (2010) and Gumedze and Jackson (2011). Lee and Thompson (2008) proposed several heavy-tailed distributions including the t, skew-normal and skew-t for

the random-effects in the presence of outlying studies. On the other hand, both long- and short-tailed distributions like the t and the Beta, respectively, have been proposed to accommodate outlying studies by Baker and Jackson (2008). A model that uses a mixture of normal distributions for the random-effects, identifies and down-weights the effect of outlying studies, was developed by Beath (2014). As an improvement on their previous work, Baker and Jackson (2016) proposed new and mathematically tractable skewed marginal distributions which are also simpler to fit than the models proposed by Baker and Jackson (2008). In their recent review article, Jackson and White (2018) recommended the use of robust meta-analytic models that relax the assumption of normally distributed random-effects.

Despite the attention it has gained in the conventional meta-analysis of interventions, tackling the effect of outlying and/or influential studies has not been studied in the context of meta-analysis of DTA studies. To the best of our knowledge, the proposal of advanced statistical methods for detecting the presence of outlying or influential studies in a meta-analysis of DTA studies was only recently studied by Negeri and Beyene (2019) (Chapter 3 in this thesis). That work demonstrated that the well-established bivariate random-effects models can be significantly affected by outlying or influential studies and may lead to misleading inferences. Although having a sound statistical method for identifying outlying and/or influential studies is a promising and necessary start, there is no clear methodology that guides how to deal with outlying and influential studies in a meta-analysis of DTA studies. There has been a general understanding among researchers that outlying or influential observations should not be discarded from the analyses due to their nature. However, a more

robust statistical method that can accommodate those outlying and/or influential studies by minimizing the effect of such studies is desired.

Since the current hierarchical and bivariate random-effects models for meta-analysis of DTA studies do not account for outlying and/or influential studies and may lead to unreliable conclusions as pointed out in Chapter 3, we aim to fill this gap by developing a new alternative bivariate random-effects model which can accommodate outlying and/or influential studies by down-weighting their effect and gives a robust statistical inference. More specifically, we propose the bivariate Laplace distribution (Kotz et al., 2001) for the random-effects to model the variation in Se and Sp across studies with the aim of accounting for potentially outlying and/or influential studies. We chose the bivariate Laplace distribution over the frequently used distributions like the t for two reasons. First, the Laplace distribution has not been implemented in the conventional meta-analysis of treatments or in a meta-analysis of DTA studies, although there have been other proposed robust models that accommodate outlying studies in the former literature as discussed above. Second, since the bivariate Laplace distribution has exactly the same number of parameters as the bivariate normal, it leads to a meaningful model comparison using information criteria approaches. By assuming the bivariate normal distribution to model the within-study variation in Se and Sp, we derive our marginal model to be the bivariate normal-Laplace (NL) distribution. That led us to implement the Monte Carlo expectation-maximization (MCEM) algorithm to obtain the maximum likelihood estimates of our model parameters. The standard errors of the parameter estimators and information criteria for model selection are obtained at the last iteration of the

MCEM algorithm once the algorithm has converged.

The rest of this chapter is organized as follows. Section 4.3 motivates the proposed model using two published datasets. Section 4.4 gives a comprehensive description of the standard as well as the proposed methods. Section 4.5 presents a simulation study results. In Section 4.6, we illustrate the methods using two real DTA meta-analyses, and, we close the chapter with a discussion and conclusion in Section 4.7.

4.3 Motivating examples

This section discusses two published meta-analyses of DTA studies to motivate the proposed model. These data are selected because they contain one or more outlying and/or influential studies according to, among others, the methods proposed in Chapter 3. The performance of the different methods when fitted to these datasets will be discussed in Section 4.6.

4.3.1 Naked eye examination (NEE) data

The first example, which is based on the NEE test for the diagnosis of primary melanoma compared to the dermoscopy test, is obtained from Vestergaard et al. (2008). The forest plot for the NEE data is given in Figure 4.1. Vestergaard et al. (2008) identified and discussed the impact of two outlying studies on the diagnostic odds ratio of dermoscopy relative to the NEE test. Both Studies 2 and 6 were detected as outlying and influential in specificity, and Study 3 is identified to be outlying and influential in sensitivity.

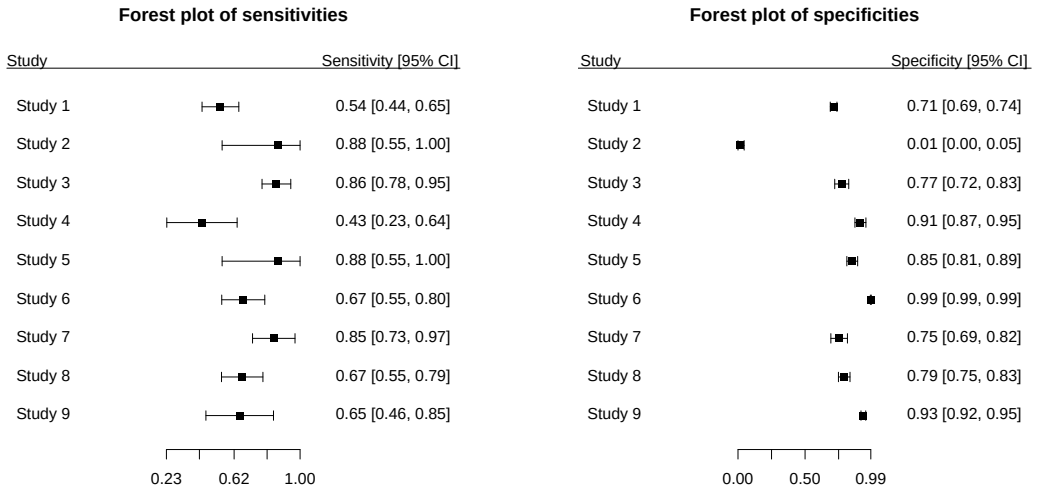

Figure 4.1: Forest plot for sensitivity (left) and specificity (right) of the *NEE* data.

4.3.2 Mini-mental state examination (MMSE) data

The second meta-analysis is from the review of Arevalo-Rodriguez et al. (2015) about the MMSE test for detecting Alzheimer's disease and dementia in people with mild cognitive impairment. Figure 4.2 displays the forest plot of the MMSE data. Studies 1 and 2 are detected as outlying and influential in specificity.

4.4 Methods

In this section, we present three bivariate random-effects models used to synthesize a pair of Se and Sp obtained from a set of k independent DTA studies. Before discussing

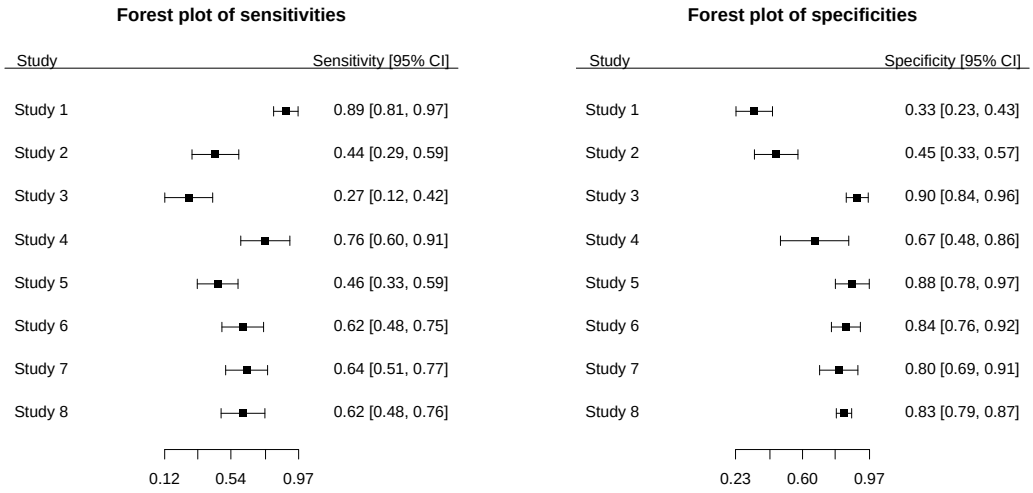

Figure 4.2: Forest plot for sensitivity (left) and specificity (right) of the *MMSE* data.

the models, we define the following notations. Let $\mathbf{y}_i = (y_{1i}, y_{2i})^T = \{\text{logit}(\widehat{Se}_i), \text{logit}(\widehat{Sp}_i)\}^T$ be a vector of the observed logit sensitivities and logit specificities for each study, i=1,...,k. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the location and scale parameters of a bivariate distribution, respectively.

4.4.1 The bivariate normal-normal (BNN) model

The standard bivariate random-effects model of Reitsma et al. (2005) assumes that the within-study variation in sensitivity and specificity is modelled according to

$$\mathbf{Y}_i|\mathbf{b}_i \sim N_2(\boldsymbol{\mu} + \mathbf{b}_i, \boldsymbol{\Psi}_i), i = 1, ..., k \tag{4.1}$$

where $\boldsymbol{\mu}$ is the fixed-effect and represents the pooled or overall logit-transformed Se and Sp, $\mathbf{b}_i$ denotes the study-specific true underlying effect (random-effects), and

$$\boldsymbol{\Psi}_i = \begin{pmatrix} \psi_{1i}^2 & 0 \\ 0 & \psi_{2i}^2 \end{pmatrix}, i = 1, ..., k$$

represents the assumed to be known within-study covariance matrix.

On the second level, the BNN model assumes that the true effect, $\mathbf{b}_i$, varies across studies, and models this variability as

$$\mathbf{b}_i \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}), i = 1, ..., k. \tag{4.2}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

denotes the between-study covariance matrix.

As in any hierarchical model, maximum likelihood estimation of parameters of the BNN model is also based on the marginal model. The marginal model obtained from (4.1) and (4.2) by integrating out the random-effects is

$$\mathbf{Y}_i \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i), i = 1, ..., k; \qquad \text{where } \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} + \boldsymbol{\Psi}_i. \tag{4.3}$$

The likelihood and log-likelihood function of the BNN model which follows from

(4.3) are given as

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{y}) = \prod_{i=1}^{k} \frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu})^T {\boldsymbol{\Sigma}_i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\}$$

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{y}) = -k\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{k}(\mathbf{y}_i - \boldsymbol{\mu})^T {\boldsymbol{\Sigma}_i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}), \quad (4.4)$$

respectively.

From (4.4) it follows that the maximum likelihood estimates (MLE) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ do not have closed-form solutions and should be obtained iteratively using numerical methods. R packages such as the `mada` (Doebler, 2017) and the `mvmeta` (Gasparrini et al., 2012), which employ both the MLE and restricted MLE methods, can be used to fit the BNN model.

Once the parameters are estimated, inference regarding the parameter of interest, $\boldsymbol{\mu}$, is made by first deriving its variance (or standard error). From (4.4), the solution to the score function with respect to $\boldsymbol{\mu}$ is

$$\boldsymbol{\mu} = \left(\sum_{i=1}^{k} \mathbf{W}_i\right)^{-1} \sum_{i=1}^{k} \mathbf{W}_i \mathbf{Y}_i \quad (4.5)$$

where $\mathbf{W}_i = (\boldsymbol{\Sigma} + \boldsymbol{\Psi}_i)^{-1} = \boldsymbol{\Sigma}_i^{-1}$

Therefore, the estimated variance-covariance matrix of the pooled logit(Se) and

logit(Sp) estimator, $\boldsymbol{\mu}$, can easily be obtained from (4.5) as

$$\widehat{\text{var}(\boldsymbol{\mu})} = \left(\sum_{i=1}^{k} \widehat{\mathbf{W}}_i\right)^{-1} = \left(\sum_{i=1}^{k} \widehat{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1} \tag{4.6}$$

where $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i$, and $\widehat{\boldsymbol{\Sigma}}$ is the maximum likelihood estimate of $\boldsymbol{\Sigma}$ obtained by maximizing (4.4). Finally, the desired $(1-\alpha)100\%$ Wald-type confidence interval for the unknown pooled logit(Se), μ_1, and pooled logit(Sp), μ_2, can be obtained using (4.5) and (4.6).

4.4.2 The bivariate binomial-normal (BBN) model

The bivariate generalized random-effects model of Chu and Cole (Chu and Cole, 2006; Chu et al., 2010), is defined by assuming the Binomial distribution for modelling the within-study variability and the bivariate normal distribution for modelling the between-study variability in sensitivity and specificity. That is, the BBN assumes

$$\begin{aligned} TP_i|b_{1i} &\sim \text{Binomial}(n_{1i}, Se_i); y_{1i} = \mu_1 + b_{1i}; \\ TN_i|b_{2i} &\sim \text{Binomial}(n_{2i}, Sp_i); y_{2i} = \mu_2 + b_{2i}; \\ \mathbf{b}_i &\sim N_2(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned}$$

where TP_i and TN_i are the study-specific number of true positives and true negatives, n_{1i} and n_{2i} are the study-specific total number of diseased and non-diseased subjects, respectively; $y_{1i}, y_{2i}, b_{1i}, b_{2i}, \mu_1, \mu_2$ are such that $\mathbf{y}_i = (y_{1i}, y_{2i})^T, \mathbf{b}_i = (b_{1i}, b_{2i})^T$, and $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$.

The marginal likelihood function of the BBN model has no closed-form expression since the integral in (4.7) does not have a closed-form solution. However, the R package `Metatron` (Huang, 2018) has recently implemented the BBN model by using the adaptive Gaussian quadrature algorithm (Chu and Cole, 2006; Paul et al., 2010) to approximate the likelihood numerically.

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{y}) = \int_{\mathbb{R}^2} \prod_{i=1}^{k} f_{\mathbf{y}_i|\mathbf{b}_i}(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\mu}) f_{\mathbf{b}_i}(\mathbf{b}_i|\boldsymbol{\Sigma}) d\mathbf{b}_i \tag{4.7}$$

4.4.3 The proposed model

This section is devoted to our proposed bivariate random-effects model. Before that, we review the symmetric multivariate Laplace (or double-exponential) distribution.

4.4.3.1 Review of the symmetric multivariate Laplace distribution

Kotz et al. (2001) defined a p-variate random vector $\mathbf{X}$ to have a symmetric multivariate Laplace distribution with a location parameter $\mathbf{0}$ and a $p \times p$ positive-definite dispersion matrix $\boldsymbol{\Sigma}$, say $\mathbf{X} \sim L_p(\mathbf{0}, \boldsymbol{\Sigma})$ if it has a characteristic function

$$\varphi_X(\mathbf{t}) = \frac{1}{1 + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}},$$

and density function

$$f(\mathbf{x}|\mathbf{0}, \boldsymbol{\Sigma}) = \frac{2}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left(\frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}{2}\right)^{\nu/2} K_\nu\left(\sqrt{2\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}\right)$$

where $\nu = (2-p)/2$ and $K_\nu(.)$ is the modified Bessel function of the third kind (see Kotz et al. (2001) for details).

As discussed in Eltoft et al. (2006) and Wang et al. (2008), the multivariate Laplace distribution can be approximated by

$$f(\mathbf{x}|\mathbf{0},\mathbf{\Sigma}) \sim \frac{\exp\left(-\sqrt{2\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x}}\right)}{2^{\frac{2p-1}{4}}\pi^{\frac{p-1}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}\left(\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x}\right)^{\frac{1}{4}}}, \tag{4.8}$$

upon using the asymptotic formula for the Bessel function:

$$K_\nu(y) \sim \sqrt{\frac{\pi}{2y}}\exp(-y), \text{ as } y \longrightarrow \infty.$$

The mean and covariance matrix of $\mathbf{X}$ found from its characteristics function is given, respectively, to be

$$\mathbb{E}(\mathbf{X}) = \mathbf{0} \text{ and } \text{var}(\mathbf{X}) = \mathbf{\Sigma}.$$

4.4.3.2 The bivariate normal-Laplace (BNL) model

Our proposed BNL model generalizes the BNN model by proposing a heavy-tailed distribution instead of the commonly used normal distribution for the random-effects. More specifically, we model the within-study variability by the bivariate normal distribution and the between-study variability by the symmetric bivariate Laplace distribution. Assuming that $\mathbf{b}_i$ and $\mathbf{e}_i$ are independent for each study, $i = 1, ..., k$,

the BNL model can be defined as

$$\begin{aligned}
&\mathbf{Y}_i = \boldsymbol{\mu} + \mathbf{b}_i + \mathbf{e}_i, \;\; i = 1, ..., k, \\
&\mathbf{b}_i \sim L_2(\mathbf{0}, \boldsymbol{\Sigma}), \;\; i = 1, ..., k, \text{is the random-effects}, \\
&\mathbf{e}_i \sim N_2(\mathbf{0}, \boldsymbol{\Psi}_i), \;\; i = 1, ..., k, \text{is the random-error term},
\end{aligned} \tag{4.9}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}_i$ are as defined earlier.

Deriving the marginal distribution of the response variable in (4.9) would let one to estimate parameters of the BNL model by the maximum likelihood method. To do so, we define a new random vector $\boldsymbol{\eta}_i = \mathbf{b}_i + \mathbf{e}_i$ and rewrite (4.9) as

$$\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\eta}_i.$$

Then using the assumption that $\mathbf{b}_i$ and $\mathbf{e}_i$ are independent for each study and by employing properties of characteristic functions, we obtain the characteristic function of $\boldsymbol{\eta}_i$ to be

$$\begin{aligned}
\varphi_{\eta_i}(\mathbf{t}) &= \varphi_{\mathbf{b}_i + \mathbf{e}_i}(\mathbf{t}); \\
&= \varphi_{\mathbf{b}_i}(\mathbf{t}) \varphi_{\mathbf{e}_i}(\mathbf{t}); \\
&= \left(\frac{1}{1 + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \right) \exp(-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t}); \\
&= \frac{\exp(-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})}{1 + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}.
\end{aligned} \tag{4.10}$$

Using the uniqueness of characteristic functions, (4.10) is recognized as the characteristic function of the symmetric bivariate Normal-Laplace distribution, $NL_2(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}_i)$. That is, $\boldsymbol{\eta}_i \sim NL_2(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}_i)$. Therefore, using *Property 1* of Jose and Thomas (2014), $\mathbf{Y}_i \sim NL_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}_i)$ with characteristic function

$$\varphi_{\mathbf{y}_i}(\mathbf{t}) = \frac{\exp(i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})}{1 + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \tag{4.11}$$

As shown in **Appendix III.A**, the mean vector and covariance matrix of the marginal model of our BNL random-effects model can easily be obtained from (4.11) to be

$$\mathbb{E}(\mathbf{Y}_i) = \boldsymbol{\mu}, \text{ and } \text{var}(\mathbf{Y}_i) = \boldsymbol{\Sigma} + \boldsymbol{\Psi}_i. \tag{4.12}$$

We note that our marginal model has no explicit density function since there is no closed-form density function for the multivariate NL distribution in general. Therefore, we cannot directly maximize the likelihood function to get estimates of our model parameters. That leads us to the next section—methods of estimating the parameters of our BNL model.

4.4.3.3 Maximum likelihood estimation of the parameters of the BNL model

Our BNL model can also be defined in a hierarchical way as

$$\begin{aligned}\mathbf{Y}_i|\mathbf{b}_i &\sim N_2(\boldsymbol{\mu}+\mathbf{b}_i, \boldsymbol{\Psi}_i), \;\; i=1,...,k;\\ \mathbf{b}_i &\sim L_2(\mathbf{0}, \boldsymbol{\Sigma}), \;\; i=1,...,k.\end{aligned}$$

leading to the following likelihood function for the marginal model

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{y}_i) = \prod_{i=1}^{k} \int_{\mathbb{R}^2} f_{\mathbf{y}_i|\mathbf{b}_i}(\mathbf{y}_i|\mathbf{b}_i; \boldsymbol{\mu}, \boldsymbol{\Psi}_i) f_{\mathbf{b}_i}(\mathbf{b}_i; \boldsymbol{\Sigma}) d\mathbf{b}_i. \tag{4.13}$$

Due to the above-mentioned reasons, the likelihood in (4.13) does not have a closed-form expression. Therefore, by treating the random-effects ($\mathbf{b}_i$) as missing data, we propose the Monte Carlo (MC) version of the expectation-maximization (EM) algorithm known as the MCEM algorithm, to estimate the five parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_{12}, \sigma_2^2)$ of our BNL model.

4.4.3.4 The MCEM algorithm

Suppose that $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_k)^T$ and $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_k)^T$ denote the observed data and latent or unobserved data, respectively, and let $\mathbf{y}_c = (\mathbf{y}, \mathbf{b})$ represent the complete data. Let $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the set of parameters of the BNL model. The

complete-data likelihood and log-likelihood function are respectively given as

$$L_c(\boldsymbol{\theta}; \mathbf{y_c}) = \prod_{i=1}^{k} f_{\mathbf{y}|\mathbf{b}}(\mathbf{y}_i|\mathbf{b}_i; \boldsymbol{\mu}, \boldsymbol{\Psi}_i) f_{\mathbf{b}}(\mathbf{b}_i; \boldsymbol{\Sigma}), \text{ and}$$
$$l_c(\boldsymbol{\theta}; \mathbf{y_c}) = \sum_{i=1}^{k} \left[\log\{f_{\mathbf{y}|\mathbf{b}}(\mathbf{y}_i|\mathbf{b}_i; \boldsymbol{\mu}, \boldsymbol{\Psi}_i)\} + \log\{f_{\mathbf{b}}(\mathbf{b}_i; \boldsymbol{\Sigma})\}\right].$$

The EM algorithm of Dempster et al. (1977) is an iterative process that involves computing the Q-function — the expected value of the complete-data log-likelihood given the observed data and the current estimate of $\boldsymbol{\theta}$ at the E-step; and iteratively maximizing the resultant Q-function until convergence at the M-step of the algorithm.

Thus, the expected complete-data log-likelihood function given the observed data and the current estimate of $\boldsymbol{\theta}$, denoted by $Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}})$ is given by

$$\begin{aligned}
Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) &= \mathbb{E}\{l_c(\boldsymbol{\theta}; \mathbf{y_c})|\mathbf{y}, \widehat{\boldsymbol{\theta}}\} \\
&= \mathbb{E}\left[\sum_{i=1}^{k} \log\{f_{\mathbf{y}|\mathbf{b}}(\mathbf{y}_i|\mathbf{b}_i; \boldsymbol{\mu}, \boldsymbol{\Psi}_i)\}|\mathbf{y}, \widehat{\boldsymbol{\theta}}\right] + \mathbb{E}\left[\sum_{i=1}^{k} \log\{f_{\mathbf{b}}(\mathbf{b}_i; \boldsymbol{\Sigma})\}|\mathbf{y}, \widehat{\boldsymbol{\theta}}\right] \\
&= \sum_{i=1}^{k} \mathbb{E}\{\log(f_{\mathbf{y}|\mathbf{b}}(\mathbf{y}_i|\mathbf{b}_i; \boldsymbol{\mu}, \boldsymbol{\Psi}_i))|\mathbf{y}, \widehat{\boldsymbol{\theta}}\} + \sum_{i=1}^{k} \mathbb{E}\{\log(f_{\mathbf{b}}(\mathbf{b}_i; \boldsymbol{\Sigma}))|\mathbf{y}, \widehat{\boldsymbol{\theta}}\} \\
&= \sum_{i=1}^{k} \int_{\mathbb{R}^2} \log\{f_{\mathbf{y}|\mathbf{b}}(\mathbf{y}_i|\mathbf{b}_i; \boldsymbol{\mu}, \boldsymbol{\Psi}_i)\} f_{\mathbf{b}|\mathbf{y}}(\mathbf{b}_i|\mathbf{y}_i; \widehat{\boldsymbol{\theta}}) d\mathbf{b}_i \\
&+ \sum_{i=1}^{k} \int_{\mathbb{R}^2} \log\{f_{\mathbf{b}}(\mathbf{b}_i; \boldsymbol{\Sigma})\} f_{\mathbf{b}|\mathbf{y}}(\mathbf{b}_i|\mathbf{y}_i; \widehat{\boldsymbol{\theta}}) d\mathbf{b}_i \qquad (4.14)
\end{aligned}$$

However, as shown in (4.14), we cannot directly compute the Q–function since the expectation (or integral) requires the knowledge of the marginal likelihood function of

the response variable $\mathbf{y}$—which does not exist in closed-form. Therefore, we propose the following MCEM algorithm where we approximate the E-step of the deterministic EM algorithm using MC methods.

Specifically, for a given Monte Carlo sample size R, we generate samples $\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, ..., \mathbf{b}_{i,R}$, $i = 1, .., k$, using Metropolis-Hastings algorithm from the target density $f_{\mathbf{b}|\mathbf{y}}(\mathbf{b}_i|\mathbf{y}_i; \widehat{\boldsymbol{\theta}})$ by specifying some candidate density say $f(\mathbf{b}_i; .)$. The acceptance probability of the Metropolis-Hastings algorithm has a nice form when the marginal density of the random-effects $f_{\mathbf{b}_i}(\mathbf{b}_i; \widehat{\boldsymbol{\Sigma}})$ is used as a candidate density as shown in **Appendix III.B**.

Once the MC samples $\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, ..., \mathbf{b}_{i,R}, i = 1, .., k$ are obtained, the Q–function in the E-step of our MCEM algorithm can be approximated by

$$Q_R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = \frac{1}{R}\sum_{i=1}^{k}\sum_{r=1}^{R}\log\{f_{\mathbf{y}|\mathbf{b}}(\mathbf{y}_i|\mathbf{b}_{i,r}; \boldsymbol{\mu}, \boldsymbol{\Psi}_i)\} + \frac{1}{R}\sum_{i=1}^{k}\sum_{r=1}^{R}\log\{f_{\mathbf{b}}(\mathbf{b}_{i,r}; \boldsymbol{\Sigma})\}. \quad (4.15)$$

By the law of large numbers, the estimator in (4.15) converges to the theoretical expectation given in (4.14).

In the maximization step of the algorithm, we maximize $Q_R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Maximizing $Q_R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\mu}$ yields a closed form solution

$$\widehat{\boldsymbol{\mu}}^{(j)} = \left(\sum_{i=1}^{k}\boldsymbol{\Psi}_i^{-1}\right)^{-1}\frac{1}{R}\sum_{i=1}^{k}\sum_{r=1}^{R}\boldsymbol{\Psi}_i^{-1}(\mathbf{y}_i - \mathbf{b}_{i,r}^{(j)}) \quad (4.16)$$

at the j-th iteration of the algorithm. Maximizing $Q_R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\Sigma}$ requires application of numerical methods. In what follows, we use the *optim* function

in the R-programming language to obtain the MLE, $\widehat{\boldsymbol{\Sigma}}^{(j)}$, of $\boldsymbol{\Sigma}$ at the j-th iteration. Point estimates of the five parameters of the BNN model, obtained after fitting the BNN model to DTA data by excluding the outlying and influential studies, is used to initiate the MCEM algorithm. If the between-study covariance matrix from the BNN model is not positive-definite, we use the 2×2 identity matrix as a starting value for $\boldsymbol{\Sigma}$ instead.

4.4.3.5 Monte Carlo sample size and convergence criteria

When implementing the MCEM algorithm, the MC sample size, R, has been given special consideration in the literature. Tanner (1993) recommends to start the MCEM algorithm with a small R, say $R = 10$, when the approximations $\widehat{\boldsymbol{\theta}}^{(j)}$ are not close to the MLE estimate $\widehat{\boldsymbol{\theta}}$ and to increase R when the iterates become closer to the MLE. Different researchers used different approaches when increasing R from iteration to iteration. McCulloch (1997) employed a predetermined value of R for different categories of iterations. He suggested to use $R = 50, 200$ and 500 for iterations 1-19, 20-39 and 40-50, respectively, and to terminate the algorithm after the 50-th iteration. Booth and Hobert (1999), Levine and Casella (2001) and Levine and Fan (2004) all developed an automated approach that would increase the MC sample size only if the previous estimate $\widehat{\boldsymbol{\theta}}^{(j-1)}$ lies in an approximate confidence interval about the current estimate $\widehat{\boldsymbol{\theta}}^{(j)}$.

Though increasing the MC sample size, R, at each iteration of the MCEM algorithm is needed as pointed out by Neath et al. (2013), it would be practical to come up with an inexpensive approach, both in terms of programming and computational

time. Robert and Casella (2010), Section 5.4.4, demonstrates the MCEM algorithm by increasing the current MC sample size linearly by doubling the value of R in the previous iteration (i.e. $R^{(j)} = 2 \times R^{(j-1)}$). Although this approach is straightforward to apply and easy to use in practice, the MC sample size will grow exponentially and leads to wastage of the MC sample sizes in the early stages of the MCEM algorithm where the current iterates $\widehat{\boldsymbol{\theta}}^{(j)}$ are far from the true MLE value $\widehat{\boldsymbol{\theta}}$ (Levine and Casella, 2001).

In this chapter, we suggest increasing the MC sample size at every iteration of the algorithm in a linear way, say $R^{(j)} = R^{(j-1)} + \lfloor R^{(j-1)}/c \rfloor$, where $\lfloor . \rfloor$ means the integer part and c is a positive constant. This approach has been previously used by Booth and Hobert (1999) and Levine and Casella (2001) using $c = 3, 4, 5$ but only when the approximate confidence interval of the current estimate $\widehat{\boldsymbol{\theta}}^{(j)}$ includes the previous estimate value $\widehat{\boldsymbol{\theta}}^{(j-1)}$. We have successfully implemented our approach using a starting MC sample size of $R = 10$ and $c = 5$.

Another point that needs to be discussed is when to stop the algorithm and declare the current parameter estimates as the MLEs. Although several methods have been suggested in the literature, the criteria based on relative error suggested by Booth and Hobert (1999) and Neath et al. (2013) is appealing. For some user-specified positive constants δ and ϵ, the ordinary EM algorithm will report convergence at the j-th iteration if the following relative error criterion is satisfied;

$$\max_{1 \leq i \leq 5} \left(\frac{\left| \boldsymbol{\theta}_i^{(j)} - \boldsymbol{\theta}_i^{(j-1)} \right|}{\left| \boldsymbol{\theta}_i^{(j)} \right| + \delta} \right) < \epsilon. \tag{4.17}$$

For the MCEM algorithm, due to the introduction of MC error in the E-step of the algorithm, both Booth and Hobert (1999) and Neath et al. (2013) recommend to declare convergence only if criteria (4.17) is met for three successive iterations. In this paper, we used criteria (4.17) with $\delta = 0.001$ and $\epsilon = 0.005$ following Booth and Hobert (1999); and we would declare convergence if criteria (4.17) is met for three consecutive iterations, covariance matrix of the estimates (see section 4.4.3.6) is positive-definite and maximum number of iterations is not reached. **Appendix III.C** summarizes the MCEM algorithm used in this article.

4.4.3.6 Estimation of standard errors and model selection criteria

Since the EM algorithm does not provide standard errors as a byproduct, several researchers have developed alternative methods that approximate standard errors of parameters estimated by the EM algorithm. Among these methods, the Oakes' approach (Oakes, 1999) is relatively simple to apply. It is well understood that, asymptotically, the inverse of the observed Fisher's information matrix, denoted by $I(\boldsymbol{\theta})$, approximates the variance of the maximum likelihood estimator. That is

$$\text{var}(\widehat{\boldsymbol{\theta}}) = I(\widehat{\boldsymbol{\theta}})^{-1} = \left[-\frac{\partial^2 \log\{L(\boldsymbol{\theta}|\mathbf{y})\}}{\partial \boldsymbol{\theta}^2} \right]^{-1} \mid \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}.$$

Oakes (1999) provided an approximation to the matrix of second-order derivatives by approximating the Q-function using the following MC sum

$$\frac{\partial^2 \log\{L(\boldsymbol{\theta}|\mathbf{y})\}}{\partial \boldsymbol{\theta}^2} = \frac{1}{R}\sum_{r=1}^{R} \frac{\partial^2 \log\{L(\boldsymbol{\theta}'|\mathbf{y};\mathbf{b}_r)\}}{\partial \boldsymbol{\theta}^2} + \frac{1}{R}\sum_{r=1}^{R}\left[\frac{\partial \log\{L(\boldsymbol{\theta}'|\mathbf{y};\mathbf{b}_r)\}}{\partial \boldsymbol{\theta}} - \frac{1}{R}\sum_{r=1}^{R}\frac{\partial \log\{L(\boldsymbol{\theta}'|\mathbf{y};\mathbf{b}_r)\}}{\partial \boldsymbol{\theta}}\right]^2, \tag{4.18}$$

using the random-effects $\mathbf{b}_{i,1}$, $\mathbf{b}_{i,2}$,..., $\mathbf{b}_{i,R}, i = 1, ..., k$, which have been already simulated at the last iteration of the MCEM algorithm (Robert and Casella, 2004). In **Appendix III.D**, we provide the first- and second-order partial derivatives of the complete-data log-likelihood required to compute (4.18).

We may compare goodness-of-fit of the proposed model against the well-established ones via AIC and BIC by approximating the marginal log-likelihood $L(\widehat{\boldsymbol{\theta}}; \mathbf{y})$ through Monte Carlo. That is, since

$$\begin{aligned} \log L(\widehat{\boldsymbol{\theta}}; \mathbf{y}) = \log \prod_{i=1}^{k} f(\mathbf{y}_i; \widehat{\boldsymbol{\theta}}) &= \log \prod_{i=1}^{k} \int_{\mathbb{R}^2} f(\mathbf{y}_i|\mathbf{b}_i; \widehat{\boldsymbol{\theta}}) f(\mathbf{b}_i; \widehat{\boldsymbol{\theta}}) d\mathbf{b}_i \\ &= \log \prod_{i=1}^{k} \mathbb{E}\{f(\mathbf{y}_i|\mathbf{b}_i; \widehat{\boldsymbol{\theta}})\}, \end{aligned}$$

at the last iteration of the MCEM algorithm, following Chen et al. (2002) we can generate large MC samples $\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, ..., \mathbf{b}_{i,R}, i = 1, ..., k$ from $f(\mathbf{b}_i; \widehat{\boldsymbol{\theta}})$ and approximate

the marginal likelihood by the MC average

$$\log L(\widehat{\boldsymbol{\theta}}; \mathbf{y}) = \sum_{i=1}^{k} \log \left\{ \frac{1}{R} \sum_{r=1}^{R} f(\mathbf{y}_i | \mathbf{b}_{i,r}; \widehat{\boldsymbol{\theta}}) \right\}.$$

Using the value of R and the MC samples $\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, ..., \mathbf{b}_{i,R}, i = 1, ..., k$ obtained at the last iteration of the MCEM algorithm, we compute the AIC and BIC as

$$\text{AIC} = -2\log L(\widehat{\boldsymbol{\theta}}; \mathbf{y}) - 2m,$$
$$\text{BIC} = -2\log L(\widehat{\boldsymbol{\theta}}; \mathbf{y}) - \log(2k)m,$$

where m, k and $\log L(\widehat{\boldsymbol{\theta}}; \mathbf{y})$ are the number of free parameters to be estimated, number of studies in the meta-analysis, and the maximized log-likelihood value, respectively. The smaller the AIC or BIC, the better the model.

4.5 Simulation study

4.5.1 Simulation design

A simulation study is performed to demonstrate the performance of the methods discussed in Section 4.4. A total of 56 settings has been considered by varying the random-effects distribution: bivariate normal or bivariate Laplace, model parameters: $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and data characteristics: n and k. To inform our simulation, we reanalyzed 165 empirical meta-analyses obtained from the Cochrane database for diagnostic test reviews from 2011–2015. Accordingly, the following set of parameters

are considered in our simulation study:

1. The overall Se and Sp for simulating DTA data with small mean vectors: (Se=0.5, Sp=0.6).

2. The overall Se and Sp for generating DTA data with large mean vectors: (Se=0.90, Sp=0.95).

3. The between-study variances and covariance parameter for simulating DTA data with no outlying studies: (σ_1^2=0.5, σ_2^2=0.6, σ_{12}=-0.2).

4. The between-study variances and covariance parameter for generating DTA data with outlying studies: (σ_1^2=(0.2,1.5), σ_2^2=(0.25,1.6), σ_{12}=-0.2).

5. The number of subjects in the diseased (n_1) and non-diseased (n_2) category: (n_1=100, n_2=200).

6. The number of studies in the meta-analysis: (k=10, 20, 30, and 40).

Seven different scenarios have been considered to systematically introduce outlying and influential studies by varying $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Our definition of 'small' and 'large' between-study variances follows those of Hamza et al. (2008) and Negeri et al. (2018). We define outlying or influential studies as those studies in the meta-analysis which have large Se or Sp, and smaller or larger between-study variances than the rest of the studies. In the first and second scenario, studies that are outlying and influential both in Se and Sp with large ($\sigma_1^2 = 1.5$, $\sigma_2^2 = 1.6$) and small ($\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.25$) between-study variances, respectively, are considered. In the third and fourth scenario, we introduced studies which are outlying and influential only in Se with large

and small between-study variances, respectively. In the fifth and sixth scenario, studies that are outlying and influential only in Sp with large and small between-study variances, respectively, are considered. Finally, studies that are outlying and influential neither in Se nor in Sp are generated. These seven scenarios are considered to illustrate how robust the proposed method is regardless of the presence or absence of outlying studies. We introduced 2, 3, 4, and 5 outlying and influential studies into the meta-analyses with 10, 20, 30, and 40 number of studies, respectively. These seven scenarios for incorporating outlying studies, along with the four k values, and the two distributional assumptions for the random-effects constitute the 56 scenarios we considered in our simulation study.

4.5.2 Simulation results

When outlying and influential studies are introduced both in Se and Sp, the between-studies variances are large, and random-effects are generated from the bivariate normal distribution (Figure 4.3), the proposed BNL model produced the most robust estimates (i.e., least affected by the inflated mean vectors of the outlying studies) whereas the standard models yielded an inflated overall sensitivity and specificity estimates. For these scenarios, the absolute (relative) change in sensitivity between the proposed model and the BNN model ranges between 7% (13%) and 10% (21%), and the absolute (relative) difference in sensitivity between the proposed model and the BBN varies from 0% (0%) to 8% (16%). The absolute (relative) difference in specificity varies from 3% (4%) to 10% (16%), between the proposed BNL model and BNN, and from 1% (2%) to 5% (8%) between the BNL model and BBN model.

For the same scenario but when the random-effects are generated from the bivariate Laplace distribution, the absolute (relative) change in Se between the proposed BNL and BNN model, and between the BNL and BBN model varies from 6% (12%) to 19% (38%), and from 2% (3%) to 12% (26%), respectively. On the other hand, the absolute (relative) change in Sp between the BNL and BNN model, and between the BNL and BBN model, respectively, ranges between 4% (5%) and 10% (16%), and between 1% (1%) and 6% (9%). Similar results have been observed when the small between-study variances have been employed.

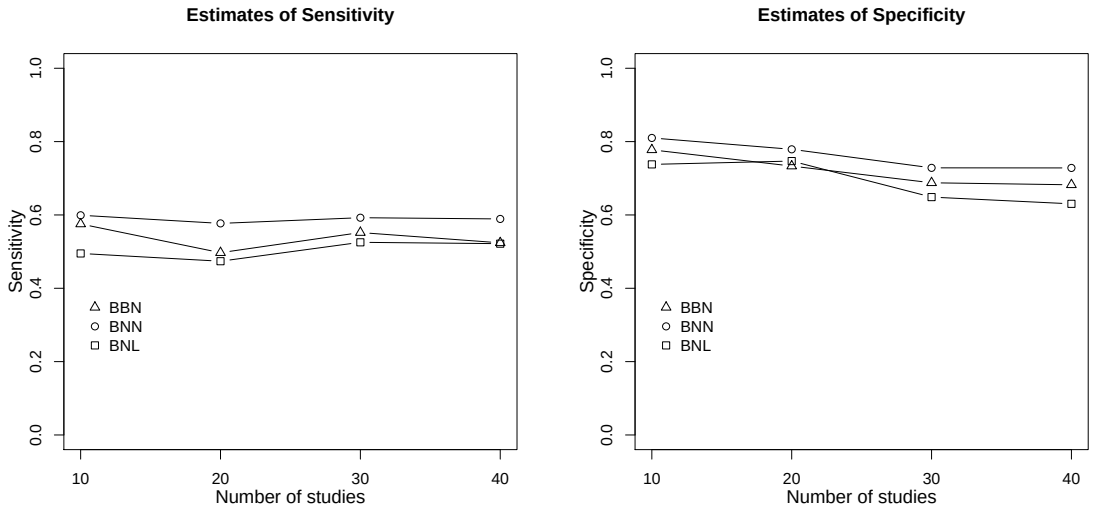

Figure 4.3: Point estimates of sensitivity (left) and specificity (right) of the three methods when outlying and influential studies are introduced both in sensitivity and specificity for different number of studies in the meta-analysis.

Figure 4.4 presents the results when outlying and influential studies are introduced only in Se, the between-study variances are large, and random-effects are generated from the bivariate normal distribution. The proposed BNL model yielded robust estimates of sensitivity and specificity compared to the standard models' estimates. The absolute (relative) change in sensitivity of the BNL model and the BNN model varies from 1% (1%) to 14% (23%), and that of the BNL model and the BBN model ranges between 1% (2%) and 6% (12%). When the bivariate Laplace distribution is used to generate the random-effects, the absolute (relative) change in Se between the BNL and BNN model, and the BNL and BBN model, respectively, ranges from 1% (1%) to 8% (16%), and from 1% (2%) to 6% (13%). Moreover, the robustness of our proposed BNL model is justified as it yielded similar estimates of specificity to the standard models. We observed similar results when the small between-study variances are used.

The results when outlying and influential studies are incorporated only in Sp, the between-study variances are large, and random-effects are simulated from the bivariate normal distribution are displayed in Figure 4.5. The absolute (relative) change in specificity of the BNL model and the BNN model varies from 4% (6%) to 6% (10%), and that of between the BNL model and the BBN model ranges between 1% (1%) and 3% (4%). On the other hand, the absolute (relative) change in Sp between the BNL and BNN model, and the BNL and BBN model was between 2% (2%) and 9% (15%), and between 0% (0%) and 10% (13%), respectively, when the random-effects are simulated from the bivariate Laplace distribution. Once again our proposed

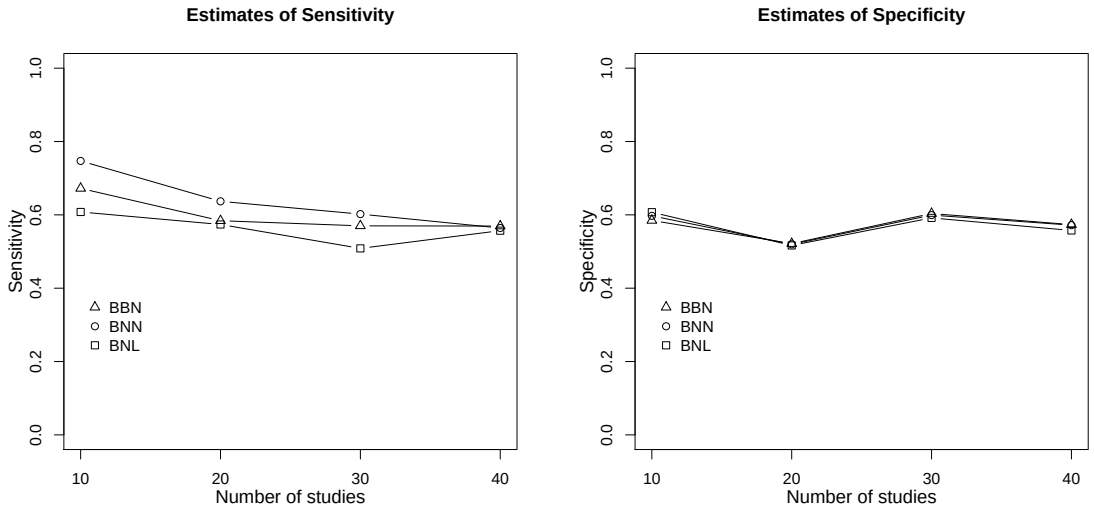

Figure 4.4: Point estimates of sensitivity (left) and specificity (right) of the three methods when outlying and influential studies are introduced only in sensitivity for different number of studies in the meta-analysis.

model produced similar estimates in sensitivity to the standard models suggesting that the BNL model is robust to outlying and influential studies. Analogous results had been noticed when the small between-study variances are assumed.

Finally, Figure 4.6 presents the results of the three models when there are outlying and influential studies neither in Se nor in Sp, and the random-effects are generated from the bivariate normal distribution. Figure 4.6 depicts that the three models yielded similar estimates of sensitivity and specificity when there are no outlying and influential studies. This result further illustrates that our proposed BNL model is robust to outlying and influential studies. Similar results have been observed when

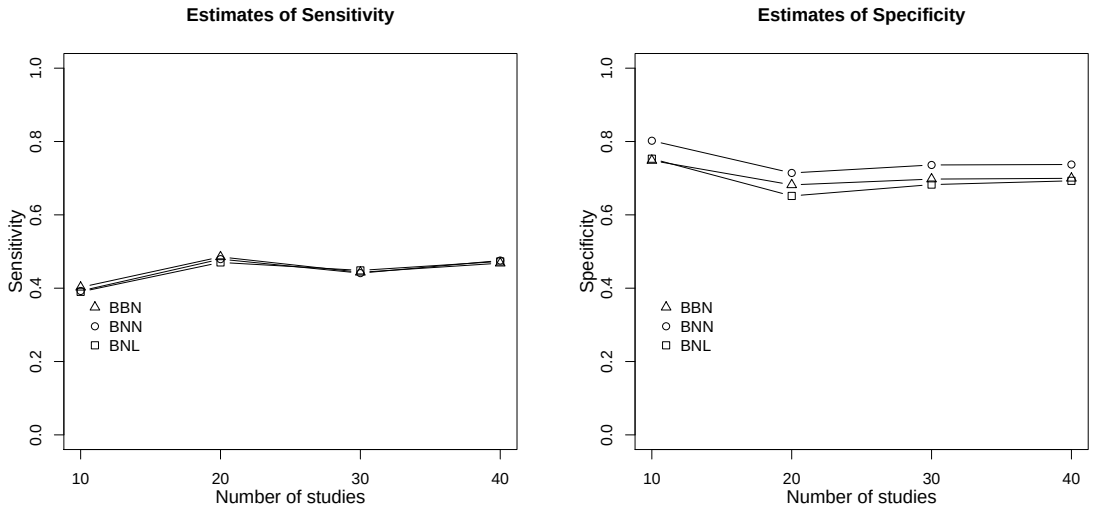

Figure 4.5: Point estimates of sensitivity (left) and specificity (right) of the three methods when outlying and influential studies are introduced only in specificity for different number of studies in the meta-analysis.

the random-effects are generated from the bivariate Laplace distribution.

4.6 Illustrative examples

This section will discuss the results of the standard and proposed methods when applied to the two datasets introduced in section 4.3.

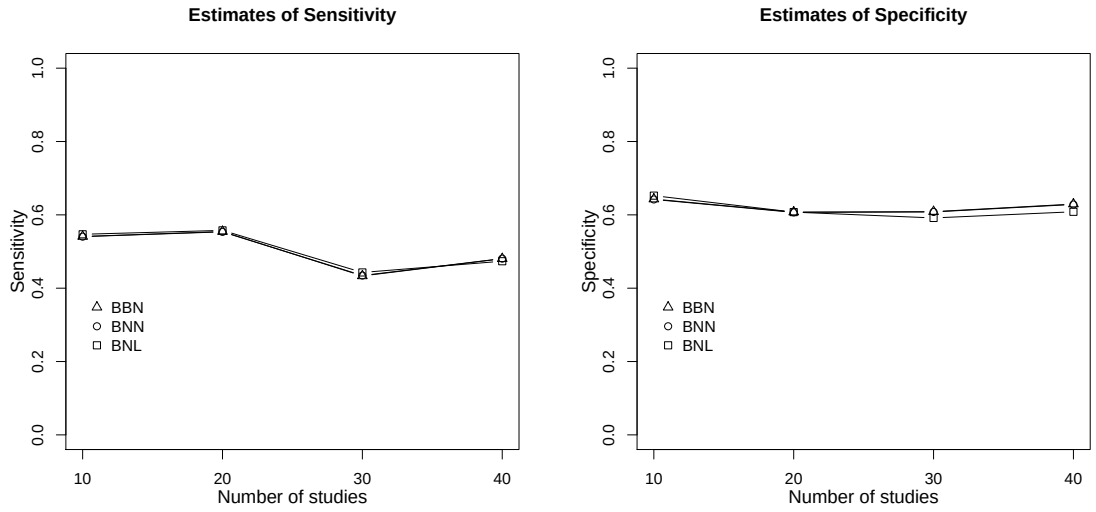

Figure 4.6: Point estimates of sensitivity (left) and specificity (right) of the three methods when there are outlying and influential studies neither in sensitivity nor in specificity for different number of studies in the meta-analysis.

4.6.1 Naked eye examination (NEE) data

Table 4.1 displays results of the three methods when fitted to the *NEE* dataset. The results indicate that the presence of three outlying and influential studies made a significant difference in the point and interval estimates of the pooled sensitivity and specificity among the models. A particularly significant difference is observed between the results of the standard BBN model and the proposed BNL model. Study 3, which was detected as influential in sensitivity, caused the BBN model to yield an overall Se which is larger than that of the proposed model by 4.4% (6.4%) in absolute (relative) value. On the other hand, the two influential studies in specificity resulted in a 5.6% and 6.6% drop, respectively, in the absolute and relative Sp of the

BBN model compared to the BNL model. The impact of these influential studies in absolute (relative) value corresponded to a 1.6% (2.3%) increase in pooled Se and a 2.2% (2.6%) decrease in pooled Sp of the BNN model compared to the BNL model. Moreover, the proposed model produced precise estimates as it yielded narrower confidence intervals. Finally, both the AIC and BIC selected the proposed model as the best model to explain the *NEE* dataset.

Table 4.1: Sensitivity (95% CI), specificity (95% CI) in percent and model comparison statistics for the *NEE* data.

Model	Se (95% CI)	Sp (95% CI)	AIC	BIC
BBN	73.2 (55.6, 85.6)	79.1 (39.4, 95.7)	158.2	162.6
BNN	70.4 (58.9, 79.7)	82.5 (60.4, 93.6)	68.6	73.1
BNL	68.8 (57.5, 78.3)	84.7 (80.6, 88.0)	68.2	72.6

4.6.2 Mini-mental state examination (MMSE) data

We present the meta-analysis results of the methods when applied to the *MMSE* data in Table 4.2. As expected, the three models yielded similar point estimates in the overall sensitivity. However, the two influential studies in specificity caused both the BNN and BBN models to result in a respective 6.0% (7.4%) and 5.8% (7.3%) reduction in absolute (relative) overall Sp compared to the proposed BNL model. Additionally, our proposed model yielded precise estimates as observed from the CIs, and also chosen by the AIC and BIC to be the best model to describe the *MMSE* data.

Table 4.2: Sensitivity (95% CI), specificity (95% CI) in percent and model comparison statistics for the *MMSE* data.

Model	Se (95% CI)	Sp (95% CI)	AIC	BIC
BBN	59.5 (46.9, 68.9)	74.3 (53.1, 88.1)	132.3	136.2
BNN	60.0 (46.7, 78.9)	74.2 (58.9, 85.2)	48.8	52.7
BNL	58.3 (57.5, 78.3)	80.1 (69.5, 87.7)	47.5	51.4

4.7 Discussion

This chapter addresses the need for a robust statistical method for meta-analysis of diagnostic test accuracy data when there are outlying and/or influential studies. With the aim of making a robust inference, a new bivariate random-effects meta-analysis model that automatically downweights the impact of outlying and influential studies on the point and interval estimates of the overall sensitivity and specificity is proposed and compared against the well-established meta-analysis methods for diagnostic test accuracy studies.

The proposed method is motivated by using two real-life meta-analyses and had also been demonstrated and validated using simulated data. The simulation study was designed to demonstrate the performance of our proposed BNL model when outlying and influential studies are present only in Se, only in Sp, both in Se and Sp, and neither in Se nor in Sp. Under each of the four settings, regardless of the other parameters varied, our proposed model produced robust point estimates in Se and Sp compared to the standard models. Of the standard models, the BBN model

yielded relatively robust estimates compared to the BNN model.

In terms of confidence intervals for sensitivity and specificity, our proposed BNL model produced precise confidence intervals relative to the standard models regardless of the parameters varied in our simulation including the distributional assumption for the random-effects. Comparing the standard models, the BBN model yielded narrower confidence intervals than the BNN model. Results of the two published meta-analyses were also in line with our simulation study as the proposed BNL model produced robust and precise point and interval estimates of the overall sensitivity and specificity. Moreover, the proposed model was shown to have demonstrated a better fit according to the information criteria (AIC and BIC).

Although we have chosen to use parameter estimates of the BNN model as starting values to start the MCEM algorithm of our proposed BNL model, we observed that the algorithm converged to similar values when employing different starting values, showing that the proposed method is robust to the choice of starting values. Additionally, the proposed method converged for all the scenarios considered in our simulation study and real-world data examples. Obviously, as for every EM algorithm, fitting our proposed BNL model takes longer time than the standard methods. However, due to its demonstrated superior performance to the standard models and the need for making an accurate inference to inform evidence-based decision-making, it is worth waiting the extra time to use the robust method we proposed here. In fact, the computational time can significantly be reduced by using parallel computing as we did for our study in this paper. We implemented our methodology in the open-source statistical software R, and the R code is available upon request from the

author.

In summary, we propose a flexible bivariate random-effects model for meta-analysis of diagnostic test accuracy studies with the aim of accommodating outlying and influential studies, as a direct response to the study of Negeri and Beyene (2019) which called for such methods to be developed. Our methodology is based on the assumption that the within-and between-study variation in sensitivity and specificity can approximately be described by the bivariate normal and Laplace distribution, respectively. That leads to a new bivariate random-effects model which has the same number of parameters as the standard models but which possesses flexibility (as it avoids removing potential outlying and influential studies) and robustness to the distributional assumption for the random-effects and presence of outlying and influential studies. Due to its flexibility and robustness properties, we believe that our proposed method is a good addition to the meta-analytic literature and an alternative to the standard models for practitioners and researchers in general.

Appendix III.A: Moments of the marginal model of the BNL RE model

The mean vector and covariance matrix of the marginal model of the BNL model is

$$\mathbb{E}(\mathbf{Y}_i) = \boldsymbol{\mu} \text{ and } \text{var}(\mathbf{Y}_i) = \boldsymbol{\Sigma} + \boldsymbol{\Psi}_i.$$

Proof: The moments of $\mathbf{Y}_i$ can be obtained from its characteristic function given in (4.11) using the relation

$$\mathbb{E}(\mathbf{Y}_j^n) = i^{-n}\varphi_{\mathbf{Y}_j}^{(n)}(\mathbf{0}), \;\; j = 1, ..., k, \;\; n = 1, 2, ...,$$

where $i = \sqrt{-1}$ is the complex number and $\varphi_{\mathbf{Y}_j}^{(n)}(.) = \frac{\partial^n \varphi_{\mathbf{Y}_j}(.)}{\partial \mathbf{t}^n}$.

Thus,

$$\begin{aligned}
\mathbb{E}(\mathbf{Y}_j) &= i^{-1}\varphi_{\mathbf{Y}_i}^{(1)}(\mathbf{0}) \\
&= \left\{ \frac{i^{-1}(i\boldsymbol{\mu} - \boldsymbol{\Psi}_i\mathbf{t})\exp(i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})(1 + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}) - \boldsymbol{\Sigma}\mathbf{t}\exp(i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})}{(1 + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})^2} \right\} | \mathbf{t} = \mathbf{0} \\
&= \boldsymbol{\mu}.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
\mathbb{E}(\mathbf{Y}_i\mathbf{Y}_i') &= i^{-2}\varphi^{(2)}_{\mathbf{Y}_i}(\mathbf{0}) \\
&= \left\{\frac{\partial i^{-1}\varphi^{(1)}_{\mathbf{Y}_i}(\mathbf{t})}{\partial \mathbf{t}}\right\}|\mathbf{t}=\mathbf{0} \\
&= i^{-2}\left[\frac{\{-\boldsymbol{\Psi}_i\exp(i\mathbf{t}'\boldsymbol{\mu}-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})+(i\boldsymbol{\mu}-\boldsymbol{\Psi}_i\mathbf{t})^2\exp(i\mathbf{t}'\boldsymbol{\mu}-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})\}(1+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})}{(1+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})^2}\right]|\mathbf{t}=\mathbf{0} \\
&\quad - i^{-2}\left\{\frac{(i\boldsymbol{\mu}-\boldsymbol{\Psi}_i\mathbf{t})\boldsymbol{\Sigma}\mathbf{t}\exp(i\mathbf{t}'\boldsymbol{\mu}-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})}{(1+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})^2}\right\}|\mathbf{t}=\mathbf{0} \\
&\quad -i^{-2}\left[\frac{\{\boldsymbol{\Sigma}\exp(i\mathbf{t}'\boldsymbol{\mu}-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})+\boldsymbol{\Sigma}\mathbf{t}(i\mathbf{t}'\boldsymbol{\mu}-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})\exp(i\mathbf{t}'\boldsymbol{\mu}-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})\}(1+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})^2}{(1+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})^4}\right]|\mathbf{t}=\mathbf{0} \\
&\quad - i^{-2}\left\{\frac{2\boldsymbol{\Sigma}\mathbf{t}\exp(i\mathbf{t}'\boldsymbol{\mu}-\frac{1}{2}\mathbf{t}'\boldsymbol{\Psi}_i\mathbf{t})(1+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})\boldsymbol{\Sigma}\mathbf{t}}{(1+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})^4}\right\}|\mathbf{t}=\mathbf{0} \\
&= \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}'+\boldsymbol{\Psi}_i.
\end{aligned}
$$

Hence, the variance-covariance matrix of the response variable, $\mathbf{Y}_i$ is

$$
\begin{aligned}
\text{var}(\mathbf{Y}_i) &= \mathbb{E}(\mathbf{Y}_i\mathbf{Y}_i') - \mathbb{E}(\mathbf{Y}_i)\mathbb{E}(\mathbf{Y}_i)' \\
&= \boldsymbol{\Sigma}+\boldsymbol{\Psi}_i.
\end{aligned}
$$

Appendix III.B: The Metropolis-Hastings (MH) algorithm to generate the random-effects

In the following, we explain how to generate the unobserved random-effects, $\mathbf{b}_{i,r}, i=1,...,k; r=1,...,R$, using the MH algorithm. Note that our target density in the E-step of the MCEM algorithm is the conditional distribution $f_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{b}_i|\mathbf{y}_i,\boldsymbol{\theta})$. As discussed in Section (4.4.3.4) of this sequel, to generate random samples from

that conditional distribution, we propose the marginal density of the random-effects, $f_{\mathbf{b}_i}(\mathbf{b}_i;\mathbf{0},\boldsymbol{\Sigma})$ as a candidate density since the support of $f_{\mathbf{b}_i}(\mathbf{b}_i;\mathbf{0},\boldsymbol{\Sigma})$ includes that of $f_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{b}_i|\mathbf{y}_i,\boldsymbol{\Sigma})$. Thus, the MH algorithm to generate random samples $\mathbf{b}_{i,r} \sim f_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{b}_i|\mathbf{y}_i;\boldsymbol{\theta})$ can proceed as follows:

1. Fix the MC sample size R and initialize the chain with some starting value $\mathbf{b}_{1,1}^{(0)} \sim f_{\mathbf{b}_i}(\mathbf{b}_i;\mathbf{0},\boldsymbol{\Sigma})$. For $j = 1, ...,$ generate $\mathbf{b}_{i,r}^{(j)}$ as follows:

2. Generate the candidate random vectors $\mathbf{b}_{i,r}^{(j)} \sim f_{\mathbf{b}_i}(\mathbf{b}_i;\mathbf{0},\boldsymbol{\Sigma})$ and set $\mathbf{V}_{i,r}^{(j)} = \mathbf{b}_{i,r}^{(j)}$.

3. Compute the acceptance probability

$$\begin{aligned}
\rho_{i,r}^{(j)} &= \min\left\{\frac{f_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{V}_{i,r}^{(j)}|\mathbf{y}_i;\boldsymbol{\theta})}{f_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{b}_{i,r}^{(j-1)}|\mathbf{y}_i;\boldsymbol{\theta})} \times \frac{f_{\mathbf{b}_i}(\mathbf{b}_{i,r}^{(j-1)};\boldsymbol{\Sigma})}{f_{\mathbf{b}_i}(\mathbf{V}_{i,r}^{(j)};\boldsymbol{\Sigma})}, 1\right\} \\
&= \min\left\{\frac{f_{\mathbf{y}_i|\mathbf{b}_i}(\mathbf{y}_i|\mathbf{V}_{i,r}^{(j)};\boldsymbol{\mu},\boldsymbol{\Psi}_i) f_{\mathbf{b}_i}(\mathbf{V}_{i,r}^{(j)};\boldsymbol{\Sigma})}{f_{\mathbf{y}_i|\mathbf{b}_i}(\mathbf{y}_i|\mathbf{b}_{i,r}^{(j-1)};\boldsymbol{\mu},\boldsymbol{\Psi}_i) f_{\mathbf{b}_i}(\mathbf{b}_{i,r}^{(j-1)};\boldsymbol{\Sigma})} \times \frac{f_{\mathbf{b}_i}(\mathbf{b}_{i,r}^{(j-1)};\boldsymbol{\Sigma})}{f_{\mathbf{b}_i}(\mathbf{V}_{i,r}^{(j)};\boldsymbol{\Sigma})}, 1\right\} \\
&= \min\left\{\frac{f_{\mathbf{y}_i|\mathbf{b}_i}(\mathbf{y}_i|\mathbf{V}_{i,r}^{(j)};\boldsymbol{\mu},\boldsymbol{\Psi}_i)}{f_{\mathbf{y}_i|\mathbf{b}_i}(\mathbf{y}_i|\mathbf{b}_{i,r}^{(j-1)};\boldsymbol{\mu},\boldsymbol{\Psi}_i)}, 1\right\}
\end{aligned}$$

4. Generate uniform random vectors $U_{i,r}^{(j)} \sim u(0,1)$ and set

$$\mathbf{b}_{i,r}^{(j)} = \begin{cases} \mathbf{V}_{i,r}^{(j)}, & \text{if } U_{i,r}^{(j)} \leq \rho_{i,r}^{(j)} \\ \mathbf{b}_{i,r}^{(j-1)}, & \text{if } U_{(i,r)}^{(j)} > \rho_{i,r}^{(j)} \end{cases}$$

Appendix III.C: The MCEM algorithm employed in this article

We summarize our MCEM algorithm as follows.

1. Initialize R, $\boldsymbol{\theta}^{(0)}= (\boldsymbol{\mu}^{(0)}, \boldsymbol{\Sigma}^{(0)})^T$. For $j = 1, ...$

2. Generate $\mathbf{b}_{i,1}^{(j)}, \mathbf{b}_{i,2}^{(j)}, ..., \mathbf{b}_{i,R}^{(j)} \sim f_{\mathbf{b}_i|\mathbf{y}_i}(\mathbf{b}_i|\mathbf{y}_i; \boldsymbol{\theta}^{(0)})$ using the MH algorithm described in **Appendix III.B**.

3. **E-Step:** Estimate $Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(j)})$ by

$$\begin{aligned}
\widehat{Q_R}(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(j)}) &= \frac{1}{R}\sum_{r=1}^{R} \log\{f_{\mathbf{y}|\mathbf{b}}(\mathbf{y}|\mathbf{b}_r^{(j)}; \widehat{\boldsymbol{\theta}}^{(j-1)})\} + \frac{1}{R}\sum_{r=1}^{R} \log\{f_{\mathbf{b}}(\mathbf{b}_r^{(j)}; \widehat{\boldsymbol{\theta}}^{(j-1)})\} \\
&= \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{k} -\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Psi}_i|) \\
&- \frac{1}{2}\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{k} \left(\mathbf{y}_i - \mathbf{b}_r^{(j)} - \boldsymbol{\mu}^{(j-1)}\right)^T \boldsymbol{\Psi}_i^{-1} \left(\mathbf{y}_i - \mathbf{b}_r^{(j)} - \boldsymbol{\mu}^{(j-1)}\right) \\
&+ \frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{k} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}^{(j-1)}|) - \frac{1}{4}\log\left(\mathbf{y}_i^T \boldsymbol{\Sigma}^{-1(j-1)} \mathbf{y}_i\right) \\
&- \sqrt{2}\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{k} \left(\mathbf{y}_i^T \boldsymbol{\Sigma}^{-1(j-1)} \mathbf{y}_i\right)^{\frac{1}{2}}
\end{aligned}$$

4. **M-Step:** Maximize $\widehat{Q_R}(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(j)})$ to obtain $\widehat{\boldsymbol{\mu}}^{(j)}$ and $\widehat{\boldsymbol{\Sigma}}^{(j)}$.

   - Maximizing $\widehat{Q_R}(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(j)})$ with respect to $\boldsymbol{\mu}$ yields

$$\widehat{\boldsymbol{\mu}}^{(j)} = \frac{1}{R}\left(\sum_{i=1}^{k} \boldsymbol{\Psi}_i^{-1}\right)^{-1} \sum_{r=1}^{R}\sum_{i=1}^{k} \boldsymbol{\Psi}_i^{-1} \left(\mathbf{y}_i - \mathbf{b}_{i,r}^{(j)}\right)$$

   - Numerical maximization of $\widehat{Q_R}(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(j)})$ using the *optim()* function in R yields $\widehat{\boldsymbol{\Sigma}}^{(j)}$.

5. i) Set $j = j + 1$

   ii) Set $R = R + \lfloor R/5 \rfloor$

6. Repeat steps two through five until convergence is achieved.

Appendix III.D: The first- and second-order partial derivatives

The first-and second-order partial derivatives of the complete-data log-likelihood, which we use them to construct the approximate Fisher information matrix, is derived below.

Note that the complete-data log-likelihood is given by

$$\begin{aligned} l(\boldsymbol{\theta}|\mathbf{y}_i, \mathbf{b}_i) &= \log L(\boldsymbol{\theta}|\mathbf{y}_c) \\ &= \sum_{i=1}^{k} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Psi}_i|) - \frac{1}{2}\left(\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{b}_i\right)^T \boldsymbol{\Psi}_i^{-1}\left(\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{b}_i\right) \\ &+ \sum_{i=1}^{k} -\frac{1}{2}\log(2^{\frac{3}{2}}\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{4}\log\left(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i\right) - \sqrt{2}\left(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i\right)^{\frac{1}{2}}. \end{aligned}$$

The first-order partial derivatives are

$$\frac{\partial \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_1} = \sum_i^k \frac{y_{1i}}{\psi_{1i}} - \frac{\mu_1}{\psi_{1i}} - \frac{b_{1i}}{\psi_{1i}}.$$

$$\frac{\partial \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_2} = \sum_i^k \frac{y_{2i}}{\psi_{2i}} - \frac{\mu_2}{\psi_{2i}} - \frac{b_{2i}}{\psi_{2i}}.$$

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \sigma_{11}} &= -\frac{1}{2}\sum_{i=1}^k \frac{\sigma_{22}}{|\boldsymbol{\Sigma}|} - \frac{1}{4}\sum_{i=1}^k \frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(2\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \sigma_{12}^2 b_{2i}^2 - \sigma_{22}^2 b_{1i}^2)}{|\boldsymbol{\Sigma}|^2} \\ &\quad - \sqrt{2}\sum_{i=1}^k \frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{12}\sigma_{22}b_{1i}b_{2i} - \frac{1}{2}\sigma_{12}^2 b_{2i}^2 - \frac{1}{2}\sigma_{22}^2 b_{1i}^2)}{|\boldsymbol{\Sigma}|^2}.\end{aligned}$$

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \sigma_{22}} &= -\frac{1}{2}\sum_{i=1}^k \frac{\sigma_{11}}{|\boldsymbol{\Sigma}|} - \frac{1}{4}\sum_{i=1}^k \frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(2\sigma_{12}\sigma_{11}b_{1i}b_{2i} - \sigma_{12}^2 b_{1i}^2 - \sigma_{11}^2 b_{2i}^2)}{|\boldsymbol{\Sigma}|^2} \\ &\quad - \sqrt{2}\sum_{i=1}^k \frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{12}\sigma_{11}b_{1i}b_{2i} - \frac{1}{2}\sigma_{12}^2 b_{1i}^2 - \frac{1}{2}\sigma_{11}^2 b_{2i}^2)}{|\boldsymbol{\Sigma}|^2}.\end{aligned}$$

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \sigma_{12}} &= \sum_{i=1}^k \frac{\sigma_{12}}{|\boldsymbol{\Sigma}|} - \frac{1}{2}\sum_{i=1}^k \frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{22}\sigma_{12}b_{1i}^2 + \sigma_{11}\sigma_{12}b_{2i}^2 - \sigma_{11}\sigma_{22}b_{1i}b_{2i} - \sigma_{12}^2 b_{1i}b_{2i})}{|\boldsymbol{\Sigma}|^2} \\ &\quad - \sqrt{2}\sum_{i=1}^k \frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{22}\sigma_{12}b_{1i}^2 + \sigma_{11}\sigma_{12}b_{2i}^2 - \sigma_{11}\sigma_{22}b_{1i}b_{2i} - \sigma_{12}^2 b_{1i}b_{2i})}{|\boldsymbol{\Sigma}|^2}.\end{aligned}$$

The second-order partial derivatives are

$$\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_1^2} = \sum_{i=1}^{k} -\frac{1}{\psi_{1i}}.$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_2^2} = \sum_{i=1}^{k} -\frac{1}{\psi_{2i}}.$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_1 \partial \mu_2} = \frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_1 \partial \sigma_{11}} = \frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_1 \partial \sigma_{22}} = \frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_1 \partial \sigma_{12}} = 0.$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_2 \partial \mu_1} = \frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_2 \partial \sigma_{11}} = \frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_2 \partial \sigma_{22}} = \frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \mu_2 \partial \sigma_{12}} = 0.$$

$$\begin{aligned}\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \sigma_{11}^2} &= \frac{1}{2}\sum_{i=1}^{k} \frac{\sigma_{22}^2}{|\boldsymbol{\Sigma}|^2} + \frac{1}{4}\sum_{i=1}^{k} \frac{(\mathbf{b}_i^T \boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-2}(2\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \sigma_{12}^2 b_{2i}^2 - \sigma_{22}^2 b_{1i}^2)^2}{|\boldsymbol{\Sigma}|^4} \\ &+ \frac{1}{4}\sum_{i=1}^{k} \frac{(\mathbf{b}_i^T \boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{22}^2\sigma_{12}b_{1i}b_{2i} - \frac{1}{2}\sigma_{22}\sigma_{12}^2 b_{2i}^2 - \frac{1}{2}\sigma_{22}^3 b_{1i}^2)}{|\boldsymbol{\Sigma}|^3} \\ &+ \frac{1}{\sqrt{2}}\sum_{i=1}^{k} \frac{(\mathbf{b}_i^T \boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{3}{2}}(2\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \sigma_{12}^2 b_{2i}^2 - \sigma_{22}^2 b_{1i}^2)^2}{|\boldsymbol{\Sigma}|^4} \\ &+ \sqrt{2}\sum_{i=1}^{k} \frac{(\mathbf{b}_i^T \boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(2\sigma_{22}^2\sigma_{12}b_{1i}b_{2i} - \sigma_{22}\sigma_{12}^2 b_{2i}^2 - \sigma_{22}^3 b_{2i}^2)}{|\boldsymbol{\Sigma}|^3}.\end{aligned}$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \sigma_{22}^2} = \frac{1}{2}\sum_{i=1}^{k}\frac{\sigma_{11}^2}{|\boldsymbol{\Sigma}|^2} + \frac{1}{4}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-2}(2\sigma_{12}\sigma_{11}b_{1i}b_{2i} - \sigma_{12}^2 b_{1i}^2 - \sigma_{11}^2 b_{2i})^2}{|\boldsymbol{\Sigma}|^4}$$
$$+ \frac{1}{4}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{11}^2\sigma_{12}b_{1i}b_{2i} - \frac{1}{2}\sigma_{11}\sigma_{12}^2 b_{1i}^2 - \frac{1}{2}\sigma_{11}^3 b_{2i}^2)}{|\boldsymbol{\Sigma}|^3}$$
$$+ \frac{1}{\sqrt{2}}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{3}{2}}(\sigma_{12}\sigma_{11}b_{1i}b_{2i} - \frac{1}{2}\sigma_{12}^2 b_{1i}^2 - \frac{1}{2}\sigma_{11}^2 b_{2i}^2)^2}{|\boldsymbol{\Sigma}|^4}$$
$$+ \sqrt{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(2\sigma_{11}^2\sigma_{12}b_{1i}b_{2i} - \sigma_{11}\sigma_{12}^2 b_{2i}^2 - \sigma_{11}^3 b_{2i}^2)}{|\boldsymbol{\Sigma}|^3}.$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \sigma_{12}^2} = \sum_{i=1}^{k}\frac{\sigma_{11}\sigma_{22}+\sigma_{12}^2}{|\boldsymbol{\Sigma}|^2} + \frac{1}{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-2}(\sigma_{22}\sigma_{12}b_{1i}^2 + \sigma_{11}\sigma_{12}b_{2i}^2 - \sigma_{11}\sigma_{22}b_{1i}b_{2i} - \sigma_{12}^2 b_{1i}b_{2i})^2}{|\boldsymbol{\Sigma}|^4}$$
$$- \frac{1}{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{22}b_{1i}^2 + \sigma_{11}b_{2i}^2 - 2\sigma_{12}b_{1i}b_{2i})}{|\boldsymbol{\Sigma}|^2}$$
$$- 2\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{22}\sigma_{12}^2 b_{1i}^2 + \sigma_{11}\sigma_{12}^2 b_{2i}^2 - \sigma_{11}\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \sigma_{12}^2 b_{1i}b_{2i})}{|\boldsymbol{\Sigma}|^3}$$
$$+ \frac{1}{\sqrt{2}}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{3}{2}}(\sigma_{22}\sigma_{12}b_{1i}^2 + \sigma_{11}\sigma_{22}b_{1i}b_{2i} - \sigma_{11}\sigma_{22}b_{1i}b_{2i} - \sigma_{12}^2 b_{1i}b_{2i})^2}{|\boldsymbol{\Sigma}|^4}$$
$$- \sqrt{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{22}b_{1i}^2 + \sigma_{11}^2 b_{2i}^2 - 2\sigma_{12}b_{1i}b_{2i})}{|\boldsymbol{\Sigma}|^2}$$
$$- (\sqrt{2})^5\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{22}\sigma_{12}^2 b_{1i}^2 + \sigma_{11}\sigma_{12}^2 b_{2i}^2 - \sigma_{11}\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \sigma_{12}^3 b_{1i}b_{2i})}{|\boldsymbol{\Sigma}|^3}.$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial\sigma_{11}\partial\sigma_{22}} = -\frac{1}{2}\sum_{i=1}^{k}\frac{|\boldsymbol{\Sigma}| - \sigma_{22}\sigma_{11}}{|\boldsymbol{\Sigma}|^2} - \frac{1}{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{12}b_{1i}b_{2i} - \sigma_{22}b_{1i}^2)}{|\boldsymbol{\Sigma}|^2}$$
$$+ \sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{11}\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \frac{1}{2}\sigma_{11}\sigma_{12}^2b_{2i}^2 - \frac{1}{2}\sigma_{11}\sigma_{22}^2b_{1i}^2)}{|\boldsymbol{\Sigma}|^3}$$
$$+\frac{1}{4}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-2}(2\sigma_{12}\sigma_{11}b_{1i}b_{2i} - \sigma_{12}^2b_{1i}^2 - \sigma_{11}^2b_{2i}^2)(2\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \sigma_{12}^2b_{2i}^2 - \sigma_{22}^2b_{1i}^2)}{|\boldsymbol{\Sigma}|^4}$$
$$+\sqrt{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{3}{2}}(\sigma_{12}\sigma_{11}b_{1i}b_{2i} - \frac{1}{2}\sigma_{12}^2b_{1i}^2 - \frac{1}{2}\sigma_{11}^2b_{2i}^2)(\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \frac{1}{2}\sigma_{12}^2b_{2i}^2 - \frac{1}{2}\sigma_{22}^2b_{1i}^2)}{|\boldsymbol{\Sigma}|^4}$$
$$- \sqrt{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{12}b_{1i}b_{2i} - \sigma_{22}b_{1i}^2)}{|\boldsymbol{\Sigma}|^2}$$
$$+ \sqrt{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(2\sigma_{11}\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \sigma_{11}\sigma_{12}^2b_{2i}^2 - \sigma_{11}\sigma_{22}^2b_{1i}^2)}{|\boldsymbol{\Sigma}|^3}.$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial\sigma_{11}\partial\sigma_{12}} = -\sum_{i=1}^{k}\frac{\sigma_{22}\sigma_{12}}{|\boldsymbol{\Sigma}|^2} - \frac{1}{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{22}b_{1i}b_{2i} - \sigma_{12}b_{2i}^2)}{|\boldsymbol{\Sigma}|^2}$$
$$- 2\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{12}^2\sigma_{22}b_{1i}b_{2i} - \frac{1}{2}\sigma_{12}^3b_{2i}^2 - \frac{1}{2}\sigma_{12}\sigma_{22}^2b_{1i}^2)}{|\boldsymbol{\Sigma}|^3}$$
$$+ \frac{1}{4|\boldsymbol{\Sigma}|^4}\sum_{i=1}^{k}(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-2}(2\sigma_{22}\sigma_{12}b_{1i}^2 + 2\sigma_{11}\sigma_{12}b_{2i}^2 - 2\sigma_{11}\sigma_{22}b_{1i}b_{2i} - 2\sigma_{12}^2b_{1i}b_{2i})$$
$$\times (2\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \sigma_{12}^2b_{2i}^2 - \sigma_{22}^2b_{1i}^2)$$
$$+ \frac{\sqrt{2}}{|\boldsymbol{\Sigma}|^4}\sum_{i=1}^{k}(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{3}{2}}(\sigma_{22}\sigma_{12}b_{1i}^2 + \sigma_{11}\sigma_{12}b_{2i}^2 - \sigma_{11}\sigma_{22}b_{1i}b_{2i} - \sigma_{12}^2b_{1i}b_{2i})$$
$$\times (\sigma_{22}\sigma_{12}b_{1i}b_{2i} - \frac{1}{2}\sigma_{12}^2b_{2i}^2 - \frac{1}{2}\sigma_{22}^2b_{1i}^2)$$
$$- \sqrt{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{22}b_{1i}b_{2i} - \sigma_{12}b_{2i}^2)}{|\boldsymbol{\Sigma}|^2}$$
$$- (\sqrt{2})^5\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{22}\sigma_{12}^2b_{1i}b_{2i} - \frac{1}{2}\sigma_{12}^3b_{2i}^2 - \frac{1}{2}\sigma_{22}^2\sigma_{12}b_{1i}^2)}{|\boldsymbol{\Sigma}|^3}.$$

$$
\begin{aligned}
\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial\sigma_{22}\partial\sigma_{12}} &= -\sum_{i=1}^{k}\frac{\sigma_{12}\sigma_{11}}{|\boldsymbol{\Sigma}|^2} - \frac{1}{2}\sum_{i=1}^{2}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{11}b_{1i}b_{2i}-\sigma_{12}b_{1i}^2)}{|\boldsymbol{\Sigma}|^2} \\
&\quad - 2\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-1}(\sigma_{12}^2\sigma_{11}b_{1i}b_{2i}-\frac{1}{2}\sigma_{12}^3b_{1i}^2-\frac{1}{2}\sigma_{12}\sigma_{11}^2b_{2i}^2)}{|\boldsymbol{\Sigma}|^3} \\
&\quad + \frac{1}{4|\boldsymbol{\Sigma}|^4}\sum_{i=1}^{k}(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-2}(2\sigma_{22}\sigma_{12}b_{1i}^2+2\sigma_{11}\sigma_{12}b_{2i}^2-2\sigma_{11}\sigma_{22}b_{1i}b_{2i}-2\sigma_{12}^2b_{1i}b_{2i}) \\
&\quad \times(2\sigma_{12}\sigma_{11}b_{1i}b_{2i}-\sigma_{12}^2b_{1i}^2-\sigma_{11}^2b_{2i}^2) \\
&\quad + \frac{\sqrt{2}}{|\boldsymbol{\Sigma}|^4}\sum_{i=1}^{k}(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{3}{2}}(\sigma_{22}\sigma_{12}b_{1i}^2+\sigma_{11}\sigma_{12}b_{2i}^2-\sigma_{11}\sigma_{22}b_{1i}b_{2i}-\sigma_{12}^2b_{1i}b_{2i}) \\
&\quad \times(\sigma_{12}\sigma_{11}b_{1i}b_{2i}-\frac{1}{2}\sigma_{12}^2b_{1i}^2-\frac{1}{2}\sigma_{11}^2b_{2i}^2) \\
&\quad - \sqrt{2}\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{11}b_{1i}b_{2i}-\sigma_{12}b_{1i}^2)}{|\boldsymbol{\Sigma}|^2} \\
&\quad - (\sqrt{2})^5\sum_{i=1}^{k}\frac{(\mathbf{b}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{b}_i)^{-\frac{1}{2}}(\sigma_{12}^2\sigma_{11}b_{1i}b_{2i}-\frac{1}{2}\sigma_{12}^3b_{1i}^2-\frac{1}{2}\sigma_{12}\sigma_{11}^2b_{2i}^2)}{|\boldsymbol{\Sigma}|^3}.
\end{aligned}
$$

Bibliography

Arevalo-Rodriguez, I., Smailagic, N., i Figuls, M. R., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O. L., Cosp, X. B., and Cullum, S. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, (3).

Baker, R. and Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health Care Management Science*, 11(2):121–131.

Baker, R. and Jackson, D. (2016). New models for describing outliers in meta-analysis. *Research Synthesis Methods*, 7(3):314–328.

Beath, K. J. (2014). A finite mixture method for outlier detection and robustness in meta-analysis. *Research Synthesis Methods*, 5(4):285–293.

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):265–285.

Chen, J., Zhang, D., and Davidian, M. (2002). A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics*, 3(3):347–360.

Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*, 59(12):1331.

Chu, H., Guo, H., and Zhou, Y. (2010). Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Medical Decision Making*, 30(4):499–508.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.

Doebler, P. (2017). *mada: Meta-Analysis of Diagnostic Accuracy*. R package version 0.5.8.

Eltoft, T., Kim, T., and Lee, T.-W. (2006). On the multivariate Laplace distribution. *IEEE Signal Processing Letters*, 13(5):300–303.

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2012). Multivariate meta-analysis for non-linear and other multi-parameter associations. *Statistics in Medicine*, 31(29):3821–3839.

Gumedze, F. N. and Jackson, D. (2011). A random effects variance shift model for detecting and accommodating outliers in meta-analysis. *BMC Medical Research Methodology*, 11(1):19.

Hamza, T. H., Reitsma, J. B., and Stijnen, T. (2008). Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches. *Medical Decision Making*, 28(5):639–649.

Huang, H. (2018). Metatron: meta-analysis for classification data and correction to imperfect reference. R package version 0.1-1. 2014.

Jackson, D. and White, I. R. (2018). When should meta-analysis avoid making hidden normality assumptions? *Biometrical Journal*, 60:1040–1058.

Jose, K. K. and Thomas, M. M. (2014). Multivariate normal-Laplace distribution and processes. *Statistica*, 74(1):27–44.

Kotz, S., Kozubowski, T., and Podgorski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Boston, MA.: Springer Science & Business Media.

Lee, K. J. and Thompson, S. G. (2008). Flexible parametric models for random-effects distributions. *Statistics in Medicine*, 27(3):418–434.

Levine, R. A. and Casella, G. (2001). Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics*, 10(3):422–439.

Levine, R. A. and Fan, J. (2004). An automated (Markov chain) Monte Carlo EM algorithm. *Journal of Statistical Computation and Simulation*, 74(5):349–360.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437):162–170.

Neath, R. C. et al. (2013). On convergence properties of the Monte Carlo EM algorithm. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton*, pages 43–62. Institute of Mathematical Statistics.

Negeri, Z. F. and Beyene, J. (2019). Statistical methods for detecting outlying and influential studies in meta-analysis of diagnostic test accuracy studies. *Statistical methods in medical research. https://doi.org/10.1177/0962280219852747.*

Negeri, Z. F., Shaikh, M., and Beyene, J. (2018). Bivariate Random-Effects Models for Meta-analysis of Diagnostic Test Accuracy Studies Using Arcsine-Based Transformations. *Biometrical Journal*, 60(4):827–844.

Oakes, D. (1999). Direct calculation of the information matrix via the EM. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):479–482.

Paul, M., Riebler, A., Bachmann, L., Rue, H., and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine*, 29(12):1325–1339.

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58(10):982–990.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York, NY.: Springer Science & Business Media.

Robert, C. P. and Casella, G. (2010). *Introducing Monte Carlo Methods with R*. New York, NY.: Springer-Verlag.

Rutter, C. M. and Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, 20(19):2865–2884.

Tanner, M. A. (1993). Tools for statistical inference: observed data and data augmentaion methods. *Lecture Notes in Statistics*, 67.

Vestergaard, M., Macaskill, P., Holt, P., and Menzies, S. (2008). Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. *British Journal of Dermatology*, 159(3):669–676.

Viechtbauer, W. and Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2):112–125.

Wang, D., Zhang, C., and Zhao, X. (2008). Multivariate Laplace filter: a heavy-tailed model for target tracking. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE.

Chapter 5

Summary, conclusions, and future directions

5.1 Summary and conclusions

This thesis was prepared with the aim of contributing statistical methods to the meta-analysis of diagnostic test accuracy (DTA) studies literature by answering three research questions.

In Chapter 2, we proposed a new bivariate random-effects model by introducing the flexible bivariate skew-normal distribution for the random effects. The proposed model performed similarly to the traditional bivariate normal-normal (BNN) model in terms of bias and root mean squared error (RMSE) of the overall logit-transformed sensitivity (Se) and specificity (Sp). However, the proposed model outperformed the standard models in terms of confidence interval (CI) width when the random-effects

are assumed to come from both the proposed model and the standard models. In terms of coverage probability (CP), all models performed similarly when the true Se and Sp are small, however, the bivariate binomial-normal (BBN) model outperformed both the BNN and bivariate normal-skew-normal (BNSN) model as the true Se and Sp get closer to one. The proposed model and the BNN model performed similarly in terms of CP regardless of the simulation settings.

With regards to the between-study covariance matrix components, the proposed BNSN model yielded better bias and RMSE for the between-study variances than the traditional BNN model, especially as the true Se and Sp approach one. However, all the methods yielded similar bias and RMSE for the between-study covariance parameter. Consistent with the simulation study, the proposed model yielded a similar point and interval estimates of the overall Se and Sp with the narrowest CI width when applied to a real dataset. Overall, a flexible bivariate random-effects model for meta-analysis of DTA studies that includes the standard BNN model as a special case, was proposed and shown to have performed better than the widely-used BNN and BBN models with regards to CI width of the overall Se and Sp, and in terms of bias and RMSE of the between-study (co)variances relative to the BNN model.

We give the following recommendations for practitioners. If the focus is obtaining both accurate (less biased) and precise (less variable) estimates of the overall Se, Sp, and between-study variances, we suggest using our proposed BNSN model over the BNN model. Although the BBN model overall performs better concerning the

fixed-effect parameters (Se and Sp), the proposed BNSN model yields precise fixed-effect estimates (i.e., narrower CIs), mostly due to its better performance in terms of the random-effects parameters. This feature of the BNSN model is particularly applicable when Se and Sp are not close to one, and within-study sample sizes are large (at least 100 as in our simulation study) since it has comparable CP to the BBN model in these situations. However, we suggest the BBN model when Se and Sp are close to one regardless of within-study sample sizes even when there is evidence of skewness in the dataset. We recommend the BNSN model if the goal is to construct a prediction interval for a future study or to make accurate random-effects parameter inference by quantifying the amount of heterogeneity, especially when skewness is present.

In Chapter 3, we introduced methods that identify outlying and/or influential studies in a meta-analysis of DTA studies. Contrary to the currently used approaches, the methods we proposed are based on rigorous statistical approaches and, therefore, overcome the subjectivity of the current practice. Moreover, the methods we proposed are flexible and can be generalized to the multivariate meta-analysis or reduced to the univariate meta-analysis in the absence of the between-study covariance matrix. Published real datasets and simulation study are used to validate and compare the proposed methods.

Accordingly, it was demonstrated that outlying and/or influential studies can negatively affect the meta-analysis of DTA studies results and lead to misleading conclusions. In summary, we suggest that the residual-based approach for identifying outlying studies should be used for detecting studies with large Se or Sp as

potentially outlying, whereas, the likelihood ratio test-based method should be used for identifying studies with small Se or Sp as potentially outlying. On the other hand, either of the proposed indexes (SIGMARATIO or VARCOVRATIO) can equally be used to detect influential studies.

Practitioners can use our proposed approaches for objectively detecting outlying and influential studies and conducting a sensitivity analysis by removing those studies from the meta-analysis and studying their impact. On the other hand, our proposed methods can be used to optimally identify the presence of outlying and influential studies and hence signal whether a practitioner needs to use a robust meta-analytic model, such as the one we propose in Chapter 4, to optimally accommodate those outlying or influential studies.

Chapter 4 is a follow-up work to Chapter 3 which raised the need for a flexible model that can accommodate outlying and influential studies in a meta-analysis of DTA studies. Accordingly, we developed a new robust bivariate random-effects meta-analytic model that naturally downweights the effects of outlying and influential studies and does not require removing those outlying or influential studies which may have the potential to influence the meta-analysis results. The marginal distribution of the proposed model is analytically derived, and the Monte Carlo expectation-maximization (MCEM) algorithm was designed to obtain the maximum likelihood estimates (MLE) of the parameters of the proposed model since the proposed model does not have a closed-form likelihood function.

A simulation study and real-world datasets have been used to appraise the proposed model and compare it to the traditional meta-analytic models. We chose the

two real-world datasets since they contain outlying and influential studies both in Se and Sp and only in Sp, which closely aligns with our simulation settings. Although the datasets are important to demonstrate the implementation of the proposed BNL model, we note that they have a similar number of studies. We provide the two datasets in **Appendix IV**.

To conclude, we developed a bivariate random-effects model that accommodates outlying and/or influential studies and is robust to model misspecification, outliers and starting values for the MCEM algorithm. Moreover, it was demonstrated that, when present, outlying and influential studies can significantly alter the results of the meta-analysis of diagnostic test accuracy studies unless appropriate models, such as the one we proposed, are employed to perform the meta-analysis.

For practitioners, we note that the benefit of fitting our proposed BNL model to a DTA data is twofold. On the one hand, our proposed BNL model yields point and interval estimates of the overall Se and Sp that are robust to the presence of outlying or influential studies compared to the standard BBN and BNN models. On the other hand, the BNL model yields similar point and interval estimates of the overall Se and Sp to the standard models when there are no outlying or influential studies in the meta-analysis, therefore, the proposed BNL model can safely be fitted to any DTA data.

5.2 Future directions

The methods we proposed in this thesis have limitations which can be improved in several ways. We discuss in the following paragraphs how these drawbacks of the proposed methods can be improved in a future study.

The proposed BNSN model is constrained to the linear mixed-effects model (LMM) approach of modelling which approximates the within-study distribution of the observed proportions, true positives (TP) and true negatives (TN). Therefore, it is worth a future study to extend the idea of modelling the between-study variation using a flexible distribution to the generalized linear mixed-effects model (GLMM) approach which directly models the within-study variation using the exact Binomial distribution. Additionally, as observed in the simulation study, the between-study covariance matrix was underestimated by the models since all the methods employ the MLE. However, the MLE of the covariance matrices resulted in a poor CP particularly for the BNN and proposed BNSN model. Therefore, it would be more appropriate to make use of the restricted maximum likelihood (REML) approach to estimate the random-effects parameters of the proposed model and study the finite-sample properties of the model in a future study.

To overcome the problem of multiple testing, we presented our proposed methods in Chapter 3 by using the 95-th percentile of the first, second and third order statistics of the appropriate test statistics to classify up to three studies as outlying or influential, following Gumedze and Jackson (2011) and Gumedze et al. (2010). Although the 95-th percentile may sound the nominal cutoff, the 90-th or 99-th percentile could also be used in practice and may result in a different number of

potentially outlying and/or influential studies. Therefore, it is worth a future work to study and come up with an optimal percentile cutoff to correctly identify multiple studies as potentially outlying or influential. In this study, we classified studies as either outlying or non-outlying and as either influential or non-influential. However, instead of dichotomizing the state of the studies, one may be interested to quantify the likelihood of being outlying or influential for each of the studies in the meta-analysis. A future study can be done to achieve that by proposing methods based on a Bayesian approach (Zhang et al., 2015) or a Frequentist approach (Beath, 2014).

The proposed bivariate normal-Laplace (BNL) model can also be improved in many ways. We studied the impact of outlying and/or influential studies on the overall point and interval estimates of sensitivity and specificity. In the future, we aim to extend our method by constructing the summary receiver operating characteristic (SROC) curve, which is another popular approach to studying the performance of diagnostic tests, to assess the impact of those outlying or influential studies on the SROC curve. Another potential extension for our proposed BNL model might look into ways for improving the computational time. Although the expectation-maximization (EM) algorithm is in general slower than other methods of finding MLE, the MCEM algorithm is even slower due to the required Monte Carlo (MC) samples at each iteration of the algorithm. Therefore, future research might study alternative approaches such as numerically approximating the unknown likelihood function of the proposed BNL model to find the MLE of the five parameters of the model. Either the Gauss-Hermite quadrature (Liu and Pierce, 1994) or MC simulation (McCulloch, 1997) can be used to numerically approximate the likelihood of

the proposed BNL model given in equation (4.13).

Finally, it is worth noting that although studies in meta-analysis usually report summary points like sensitivity and specificity at different threshold or cutoff values (Roberts et al., 2015; Zhelev et al., 2015; Wacker et al., 2013; Vouloumanou et al., 2011; Aertgeerts et al., 2004), the methods discussed in this Thesis do not have the capacity to analyze such data by taking into account all the available threshold values. Although efforts have recently been made to propose appropriate methods (meta-analytic methods for individual patient data (IPD)) for such datasets (Riley et al., 2014; Putter et al., 2010; Hamza et al., 2009; Dukic and Gatsonis, 2003), the proposed methods are similar in their assumption about the random-effects to the current traditional methods (the BNN and BBN) for meta-analysis of DTA studies. Therefore, it is worth experimenting whether the ideas proposed in this thesis can be extended to the context of a meta-analysis of IPD.

Appendix IV: The Naked eye examination (NEE) and Mini-mental state examination (MMSE) datasets

Table 5.1: The NEE data

Study	Author	TP	FP	FN	TN
1	Argeziano	46	362	39	898
2	Carli'03	3	40	0	0
3	Bono'02	57	56	9	191
4	Bono'06	10	16	13	167
5	Carli'04	3	44	0	255
6	Stanganelli	37	33	18	3284
7	Cristofolini	28	46	5	141
8	Benelli	40	71	20	270
9	Dummer	15	49	8	699

Table 5.2: The MMSE data

Study	Author	Year	TP	FP	FN	TN
1	Buchhave	2008	56	56	7	28
2	Conde-Sala	2012	19	36	24	30
3	Devanand	2008	9	9	24	83
4	Modrego	2005	22	8	7	16
5	Modrego	2013	26	6	31	42
6	Palmqvist	2012	32	13	20	68
7	Pozueta	2011	32	11	18	44
8	Xu	2002	29	52	18	252

Bibliography

Aertgeerts, B., Buntinx, F., and Kester, A. (2004). The value of the CAGE in screening for alcohol abuse and alcohol dependence in general clinical populations: a diagnostic meta-analysis. *Journal of Clinical Epidemiology*, 57(1):30–39.

Baker, R. and Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health Care Management Science*, 11(2):121–131.

Baker, R. and Jackson, D. (2016). New models for describing outliers in meta-analysis. *Research Synthesis Methods*, 7(3):314–328.

Bashore, T. R., Osman, A., and Heffley III, E. F. (1989). Mental slowing in elderly persons: A cognitive psychophysiological analysis. *Psychology and Aging*, 4(2):235.

Beath, K. J. (2014). A finite mixture method for outlier detection and robustness in meta-analysis. *Research Synthesis Methods*, 5(4):285–293.

Berlin, J. A. and Antman, E. M. (1992). Advantages and limitations of meta-analytic regressions of clinical trials data. *Controlled Clinical Trials*, 13(5):422.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley and Sons.

Chalmers, I., Hedges, L. V., and Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions*, 25(1):12–37.

Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity

with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*, 59(12):1331.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, 4(1):102–118.

de Jesus, J. O., Parker, L., Frangos, A. J., and Nazarian, L. N. (2009). Accuracy of MRI, MR Arthrography, and Ultrasound in the Diagnosis of Rotator Cuff Tears: A Meta-Analysis. *American Journal of Roentgenology*, 192:1701–1707.

Devillé, W. L., Buntinx, F., Bouter, L. M., Montori, V. M., De Vet, H. C., Van der Windt, D. A., and Bezemer, P. D. (2002). Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Medical Research Methodology*, 2(1):9.

Doria, A. S., Moineddin, R., Kellenberger, C. J., and et al. (2006). US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis. *Radiology*, 241(1):83–94.

Dukic, V. and Gatsonis, C. (2003). Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics*, 59(4):936–946.

Fisher, R. A. (1932). *Statistical Methods for Research Workers*. London: Oliver and Boyd, 4th ed edition.

Gatsonis, C. and Paliwal, P. (2006). Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. *American Journal of Roentgenology*, 187(2):271–281.

Gumedze, F. N. and Jackson, D. (2011). A random effects variance shift model for detecting and accommodating outliers in meta-analysis. *BMC Medical Research Methodology*, 11(1):19.

Gumedze, F. N., Welham, S. J., Gogel, B. J., and Thompson, R. (2010). A variance shift model for detection of outliers in the linear mixed model. *Computational Statistics and Data Analysis*, 54:2128–2144.

Hamza, T. H., Arends, L. R., van Houwelingen, H. C., and Stijnen, T. (2009). Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Medical Research Methodology*, 9(1):73.

Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P., and Sterne, J. A. (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, 8(2):239–51.

Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Aanalysis*. New York: Academic Press.

Kriston, L., Hölzel, L., Weiser, A. K., Berner, M. M., and Härter, M. (2008). Meta-analysis: Are 3 Questions Enough to Detect Unhealthy Alcohol Use? *Annals of Internal Medicine*, 149:879–888.

Lee, K. J. and Thompson, S. G. (2008). Flexible parametric models for random-effects distributions. *Statistics in Medicine*, 27(3):418–434.

Liu, Q. and Pierce, D. A. (1994). A note on Gauss—Hermite quadrature. *Biometrika*, 81(3):624–629.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437):162–170.

Moses, L., Shapiro, D., and Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary ROC curve: Data-analytic approaches and some additional considerations. *Statistics in Medicine*, 12:1293–1316.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal*, 3:1243–1246.

Pearson, K. (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, pages 379–410.

Petignat, P., Faltin, D. L., and Bruchim, I. (2007). Are self-collected samples comparable to physician-collected cervical specimens for human papillomavirus DNA testing? A systematic review and meta-analysis. *Gynecologic Oncology*, 105:530–535.

Pormohammad, A., Riahi, S. M., Nasiri, M. J., and et al. (2017). Diagnostic test accuracy of adenosine deaminase for tuberculous meningitis: A systematic review and meta-analysis. *Journal of Infection*, 74:545–554.

Putter, H., Fiocco, M., and Stijnen, T. (2010). Meta-analysis of diagnostic test accuracy studies with multiple thresholds using survival methods. *Biometrical Journal*, 52(1):95–110.

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58(10):982–990.

Riley, R. D., Takwoingi, Y., Trikalinos, T., Guha, A., Biswas, A., Ensor, J., Morris, R. K., and Deeks, J. J. (2014). Meta-Analysis of Test Accuracy Studies with Multiple and Missing Thresholds: A Multivariate-Normal Model. *Journal of Biometrics and Biostatistics*, 5(3):196.

Roberts, E., Ludman, A. J., Dworzynski, K., Al-Mohammad, A., Cowie, M. R., McMurray, J. J., and Mant, J. (2015). The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. *British Medical Journal*, 350:h910.

Rutter, C. M. and Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, 20(19):2865–2884.

Schwarzer, G., Carpenter, J. R., and Rücker, G. (2015). *Meta-Analysis with R*, volume 4724. New York: Springer.

Shannon, H. (2016). A statistical note on Karl Pearson's 1904 meta-analysis. *Journal of the Royal Society of Medicine*, 109(8):310–311.

Singal, A., Volk, M. L., Waljee, and et al. (2009). Meta-analysis: surveillance with

ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Alimentary Pharmacology & Therapeutics*, 30:37–47.

Van Houwelingen, H. C., Arends, L. R., and Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, 21(4):589–624.

Viechtbauer, W. and Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2):112–125.

Vouloumanou, E. K., Plessa, E., Karageorgopoulos, D. E., Mantadakis, E., and Falagas, M. E. (2011). Serum procalcitonin as a diagnostic marker for neonatal sepsis: a systematic review and meta-analysis. *Intensive Care Medicine*, 37(5):747–762.

Wacker, C., Prkno, A., Brunkhorst, F. M., and Schlattmann, P. (2013). Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, 13(5):426–435.

Walter, S. (2002). Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine*, 21(9):1237–1256.

Willis, B. H. and Quigley, M. (2011). Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. *BMC Medical Research Methodology*, 11(1):27.

Zhang, J., Fu, H., and Carlin, B. P. (2015). Detecting outlying trials in network meta-analysis. *Statistics in Medicine*, 34(19):2695–2707.

Zhelev, Z., Hyde, C., Youngman, E., Rogers, M., Fleming, S., Slade, T., Coelho, H., Jones-Hughes, T., and Nikolaou, V. (2015). Diagnostic accuracy of single baseline measurement of Elecsys Troponin T high-sensitive assay for diagnosis of acute myocardial infarction in emergency department: systematic review and meta-analysis. *British Medical Journal*, 350:h15.

Zhou, Y., Wang, C., Zhu, H., Lin, Y., Pan, B., Zhang, X., and et al. (2016). Diagnostic Accuracy of PIK3CA Mutation Detection by Circulating Free DNA in Breast Cancer: A Meta-Analysis of Diagnostic Test Accuracy. *PLoS ONE*, 11(6).