# META-ANALYSIS THROUGH COMBINING CONFIDENCE DISTRIBUTIONS

**BY Guang Yang**

**A dissertation submitted to the**

**Graduate School—New Brunswick**

**Rutgers, The State University of New Jersey**

**in partial fulfillment of the requirements**

**for the degree of**

**Doctor of Philosophy**

**Graduate Program in Statistics and Biostatistics**

**Written under the direction of**

**Minge Xie**

**and approved by**

_____

_____

_____

_____

**New Brunswick, New Jersey**

**October, 2013**

## ABSTRACT OF THE DISSERTATION

# Meta-Analysis Through Combining Confidence Distributions

**by Guang Yang**

**Dissertation Director: Minge Xie**

This dissertation develops a set of new statistical methods for synthesizing joint information of multiple parameters from different sources by combining multivariate normal confidence distributions. These methods support the development of an asymptotic efficient network meta-analysis approach and also several robust multivariate meta-analysis approaches. Both theoretical and numerical results show that the developed methods are superior to the conventional frequentist meta-analysis approach and the commonly used Bayesian methods. They also indicate that the developed approaches can mitigate effectively the undue impact from potential outlying studies.

Meta-analysis generally refers to the process of systemically combining the results from independent but similar studies in support of data-driven decision making. It has been widely used in many fields, including clinical researches, social sciences, among others. Many methods have been developed to combine information effectively and efficiently.

However, there still remain several challenging problems. This dissertation aims to solve two challenging problems that are often seen in meta-analysis.

The first part of this dissertation is on how to efficiently incorporate indirect evidence in the network meta-analysis setting, which aims to strengthen the pairwise direct comparison by borrowing information from indirect comparisons. The developed network meta-analysis approach can efficiently combine all studies from a network of direct and indirect evidence, and, moreover, effectively include studies that compare more than two treatments.

The second part of this dissertation is on how to mitigate effectively the effect of inconsistent or outlying studies in the meta-analysis by developing two robust multivariate meta-analysis approaches. One approach assumes that the number of studies goes to infinity, whereas the other assumes that the number of studies is finite but each study size may go to infinity. These approaches are shown to be robust against the effect of inconsistent or outlying studies, as well as model misspecifications. We present both theoretical and numerical results to show that these two robust approaches achieve high breakdown points and retain relatively high efficiency in comparison with the most efficient approach.

Finally, an R package **gmeta** has been developed to facilitate the use of the unified univariate meta-analysis framework through combining confidence distributions.

# List of Tables

# List of Figures

# Acknowledgements

This dissertation would not have been finished without the guidance from my advisor, assistance from my dissertation committee members, friendship from my friends, and the unwavering support from my family.

First, I would like to express my deepest appreciation to my advisor, Professor Minge Xie. His invaluable guidance has sustained my research ability and interest. This dissertation would not have been possible without his continuous support and encouragement. He does not only inspire me to find right directions to proceed my research projects, but also patiently help me improve my communication and writing skills, not to mention the generous financial support during all these years.

I would also like to express my gratitude to Professor Regina Liu, for her great help on my research projects. I would also like to thank Professor Ying Hung, for taking her precious time and effort to serve on my thesis committee. I would like to thank Dr. Junyuan Wang, who patiently and generously provided me with guidance on thinking like a Biostatistician.

In addition, I would like to thank our graduate director Professor John Kolassa, who has always been eager to help. I also have greatly benefited from the Department of Statistics and Biostatistics of Rutgers University. The excellent faculty, staff and fellow students, and the continuous financial support from the department, have provided an ideal research environment throughout my graduate study.

# Dedication

To my family.

# Table of Contents

# Chapter 1

# Introduction

Meta-analysis is the quantitative statistical method for systematically combining information from different sources, which aims to provide combined inference with improved efficiency or strengthen the conclusion of current study (Normand, 1999; van Houwelingen et al., 2002). This method is widely used in many fields in support of evidence-based decision-making (cf., Sutton et al., 2000; Sutton and Higgins, 2008, and references therein). For a single parameter, a general univariate confidence distribution (CD) combining method is proposed and utilized for univariate meta-analysis by Singh et al. (2005) and Xie et al. (2011). This dissertation develops new multivariate meta-analysis approaches using similar ideas by combining multivariate normal confidence distributions.

Multivariate meta-analysis jointly analyzes multiple parameters. It has been demonstrated that analysis of multiple parameters all together can be beneficial to the combined inference by borrowing information from one parameter to the other (Arends et al., 2003; Arends, 2006). However, such a method is still limitedly used in practice, although it has been advocated for almost 30 years since it came into being (Riley, 2009). As the joint collection of outcomes from multiple endpoints becomes more common, the demand for simple and effective multivariate meta-analysis methods has never been greater (Jackson et al., 2011). Two particular challenges that obstruct the widespread use of multivariate meta-analysis are as follows:

- How to efficiently integrate all the studies in a network of evidence, even when individual studies are heterogenous with partial common parameters of interest; and
- How to set up robust approaches that can effectively withstand the impact of unknown outlying studies.

This dissertation aims to address the two important problems above by developing new multivariate meta-analysis approaches:

- We propose an asymptotic efficient network meta-analysis approach, which can effectively synthesize evidence from heterogenous studies with partial common parameters. This method is useful in clinical researches when the primary interest of a meta-analysis is to compare the effectiveness of two experimental treatments. The available studies may directly compare these two treatments and provide direct evidence, or compare one of the two treatments to placebo and thus provide only indirect evidence (Lumley, 2002; Lu and Ades, 2004). The proposed approach can efficiently integrate both direct and indirect comparisons in a network of evidence, including studies involved more than two treatments, and thus outperforms the traditional pairwise meta-analysis approach which only summarizes direct comparisons. This approach is prior free and can always provide proper inference in terms of confidence intervals with correct coverage rates, whereas the commonly used Bayesian method is sensitive to the choice of prior distributions.
- We propose two robust multivariate meta-analysis approaches, which can resist the impact of unknown outlying studies. One approach is appropriate for meta-analysis of a large number of studies, which relies on asymptotic normality and has an inherent connection to an M-estimation approach. The other approach is useful for meta-analysis of a set of large studies, where outlying studies are down-weighted

or excluded using data-dependent adaptive weights. These robust approaches can provide consistent estimator when there are outlying studies involved in the meta-analysis, whereas the conventional efficient meta-analysis approach and commonly used Bayesian approaches can not. Meanwhile, they maintain relative high efficiency when no outlying study exists. As a result, they provide a means of protection against model misspecification, where outlying studies are not properly modeled. To our knowledge, none of the existing approaches, including the Bayesian approaches, can provide such protection. The robust approaches can also adapt to network meta-analysis by incorporating heterogenous studies with partial common parameters.

These developments are based on combining confidence distributions. A CD uses a sample-dependent distribution function on the parameter space to estimate the unknown parameter. It naturally contains more information than a point or interval estimator, and is thus a more versatile tool for inference (Cox, 2013). This concept is broad and subsumes normalized likelihood function, $p$-value function, and bootstrap distribution, among others, under the same definition (cf., Xie and Singh, 2013, and reference therein).

In addition, the third part of this dissertation develops a computing software to facilitate the use of the unified univariate meta-analysis framework proposed in Xie et al. (2011):

- We develop an R package **gmeta**. The `gmeta()` function uses the same structure to perform all meta-analysis, including $p$-value combinations (cf., Marden, 1991), conventional model-based meta-analysis (cf., Table IV of Normand, 1999), robust meta-analysis under contaminated models (cf, Section 4 of Xie et al., 2011), and the Mantel-Haenszel method, the Peto's method, and two exact methods using Binomial distribution without arbitrary continuity corrections (Tian et al., 2009; Liu et al., 2013), for synthesizing the findings from $2 \times 2$ tables. The associated plot function

can show the individual and combined confidence distributions through extended forest plots.

This dissertation organizes each chapter as a self-contained paper. There is some overlap in some of background materials. Specifically, Chapter 2 develops an asymptotic network meta-analysis approach by combining multivariate normal CD random vectors. Real examples and simulation studies show that the proposed approach is often superior than the traditional pairwise meta-analysis and commonly used Bayesian methods. Chapter 3 introduces a general multivariate normal CD combining method by combining CD functions, which supports the development of two robust multivariate meta-analysis approaches. These approaches have superior performance than the asymptotic efficient meta-analysis approach and Bayesian method, when outlying studies are involved in the meta-analysis. Chapter 4 presents an R package **gmeta**, which provides an all-in-one solution for univariate meta-analysis. Chapter 5 contains some concluding remarks.

# Chapter 2

# A Confidence Distribution Approach for an Efficient Network Meta-Analysis

In this chapter, we address the first problem mentioned in the introduction chapter by developing a network meta-analysis approach that can *efficiently integrate all the studies in a network of evidence, even when individual studies are heterogenous with partial common parameters of interest.*

## 2.1 Introduction

Recent advances in computing and data storage technology have greatly facilitated data gathering from many disparate sources. The demand for efficient methodologies for combining information from independent studies or disparate sources has never been greater. So far, meta-analysis is one of the most, if not the most, commonly used approaches for synthesizing findings from different sources for pairwise comparisons. For example, it is used in medical research for summarizing estimates from a set of randomized controlled trials (RCTs) of the relative efficacy of two treatments (cf. Normand, 1999; Sutton and Higgins, 2008). For more-complicated comparative effectiveness research, where the comparisons involve a network of more than two treatments, several generalizations have been developed for combining information from various sources. A useful survey can be found

in the report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices (Jansen et al., 2011; Hoaglin et al., 2011) and its references. A key advantage of network meta-analysis is that it can perform indirect comparisons among multiple treatments.

We elaborate on network meta-analysis with a general setting and a worked example. In the general setting, the process begins with a systematic research for RCTs that have compared treatments for a particular condition. The trials that satisfy a set of eligibility criteria yield a network of evidence, in which each node represents a treatment and each edge represents a direct comparison in one or more trials. We assume that the network is connected, and we denote the total number of treatments by $p$ and the number of treatments in trial $i$ by $p_i$ ($2 \leq p_i \leq p$). For example, Stettler et al. (2007) assembled data from 37 trials for comparing the performance of three stents in patients with coronary artery disease. Figure 2.1 illustrates the network of the comparisons among the three stents. Each stent is connected to the other two through a number of direct comparisons, and these three stents form a network. The primary objective is to assess the effectiveness of these three stents (more broadly all treatments in the network). The estimates of network meta-analysis yields pairwise comparisons.

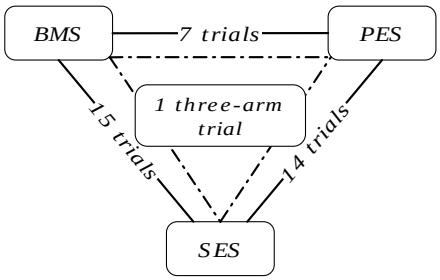


Figure 2.1: Network of comparisons for bare-metal stents (BMS), paclitaxel-eluting stents (PES), and sirolimus-eluting stents (SES) in 37 trials (Stettler et al., 2007)

Several network meta-analysis approaches have been reported in the literature. Lumley

(2002) introduced a model for combining evidence from trials with pairwise comparisons between treatments. Although this method allows borrowing of evidence from indirect comparisons to strengthen the results of direct comparisons, it is somewhat restricted in practice because it requires that each individual trial be a two-arm trial (i.e., compare exactly two treatments). Thus, this method cannot deal with multi-arm trials as in the example of Figure 2.1. Generalizing the method in Smith et al. (1995), Lu and Ades (2004) introduced a network meta-analysis approach using a Bayesian hierarchical model. Although this approach can include multi-arm trials, our simulation studies in Section 4 show that its inferences can be quite sensitive to the choice of priors. More specifically, if the assumptions in the prior distribution does not agree with the underlying true model (the unknown between-trial covariance structure), the resulting credible interval fails to achieve the nominal coverage probability, and, in some cases, its empirical coverage probability can be far below the nominal level.

This paper aims to introduce a new network meta-analysis approach that: i) can efficiently synthesize evidence from a number of independent trials on multiple treatments; ii) can include trials with multiple arms; and iii) does not need to specify priors for parameters of interest or other parameters. The proposed approach is derived from combining multivariate confidence distributions.

To some extent, our proposed CD approach extends of the method developed in Lumley (2002) to include multi-arm trials. Compared with the Bayesian method in Lu and Ades (2004), the proposed CD approach is a pure frequentist approach and it does not require specification of priors. In fact, the proposed CD approach can be viewed as a frequentist counterpart of the Bayesian method of Lu and Ades (2004).

The general idea of combining confidence distributions has been developed in Singh et al. (2005) and Xie et al. (2011). The concept of CD and its utility in statistical inference

have been investigated intensely; see, e.g., Schweder and Hjort (2002) and Singh et al. (2005, 2007). A detailed survey of the recent developments on CD can be found in Xie and Singh (2013). Roughly speaking, a CD bases inferences on a sample-dependent distribution function, rather than a point or an interval, on the parameter space. A CD can be viewed as a frequentist "distribution estimator" of an unknown parameter, as described in Xie and Singh (2013) and Cox (2013). As a distribution function, a CD naturally contains more information than a point or interval estimator, and is thus a more versatile tool for inference. For example, for an odds ratio when the 2x2 table has zero events, point or interval approaches may fail, but the CD approach remains valid, as shown in Liu et al. (2013). CDs have been demonstrated in Singh et al. (2005) and Xie et al. (2011) to be especially useful for combining information on a single parameter. In particular, Xie et al. (2011) showed that the CD combining approach can provide not only a unifying framework for almost all univariate meta-analysis applications, but it can also provide new estimates that can achieve desirable properties such as high efficiency and robustness. Network meta-analysis generally involves multiple parameters, and the information on each parameter may have non-negligible impact on inferences for other parameters. To fully utilize the joint information on multiple parameters, we construct multivariate joint CD functions for the entire set of parameters from each study. The combination of these joint CD functions leads to a novel frequentist approach to network meta-analysis.

Our numerical studies show that the proposed CD approach compares favorably with, and often is superior to, traditional meta-analysis and the hierarchical Bayesian network meta-analysis method proposed by Lu and Ades (2004). Specifically, in comparison with the traditional method, the CD method is more efficient because it uses indirect evidence. In comparison with the Bayesian method, the CD approach is prior-free and can always provide a proper inference (i.e., confidence intervals with correct coverage rates) for treatment

effects, regardless of the between-trial covariance structure. Moreover, our simulation studies show that the performance of the Bayesian approach is sensitive to the choice of prior distributions, which ideally should reflect the true underlying the between-trial covariance structure.

The paper is organized as follows. Section 2 reviews the concept of CD and develops a general method for combining multivariate normal CDs to facilitate network meta-analysis. Section 3 uses two real data examples to illustrate the proposed CD approach in the analysis of a three-treatment network, and to compare it with traditional meta-analysis and the Bayesian network meta-analysis. In Section 4, the results of several simulation studies demonstrate that the proposed CD approach can provide proper inferences. Comparisons with the traditional and Bayesian network meta-analysis approaches are also provided. Moreover, we devise a simple adaptive CD approach to address possible inconsistent (or contradictory) evidence from indirect and direct comparisons. This adaptive approach can alleviate undue influence from indirect comparisons whose evidence contradicts the direct comparisons. Section 5 contains a summary and further remarks.

## 2.2 A CD Approach for Network Meta-Analysis

Assume that the evidence network comprises $k$ independent clinical trials and involves the effects of $p$ treatments, denoted by the vector $\boldsymbol{\theta} \equiv (\theta_1, \cdots, \theta_p)^{\mathbf{T}}$. The individual trials may have studied only a subset of the $p$ treatments. More specifically, the $i$-th trial involves $p_i \leq p$ treatments. If $p_i < p$, the $i$-th trial provides only partial information about $\boldsymbol{\theta}$, in the sense that only the $p_i$-dimensional parameter $\boldsymbol{\theta}_i \equiv \mathbf{A}_i\boldsymbol{\theta}$ is identifiable, where the $p_i \times p$ selection matrix $\mathbf{A}_i$ is obtained by removing from the $p \times p$ identity matrix (or, more generally, any $p \times p$ orthogonal matrix $\mathbf{A}$) the rows that correspond to the omitted

parameters. Throughout this paper, we consider the following multivariate random-effects model for network meta-analysis. It extends the univariate hierarchical random-effects model reviewed in Normand (1999):

$$\boldsymbol{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i), \quad \boldsymbol{\theta}_i|\boldsymbol{\theta}, \mathbf{S} \overset{\text{ind}}{\sim} N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}), \quad i = 1, 2, \ldots, k \tag{2.1}$$

where $\boldsymbol{y}_i$ is the summary statistic from the $i$-th study, $\boldsymbol{\Sigma}_i$ is the covariance matrix of $\boldsymbol{y}_i$, and $\mathbf{S}$ is the covariance matrix of random-effects distribution.

A key question in network meta-analysis is how the information on $\boldsymbol{\theta}_i$ (which may provide only partial information on $\boldsymbol{\theta}$) can be integrated to make efficient inference about $\boldsymbol{\theta}$. Our proposed approach of combining multivariate normal CDs can provide a solution.

Before presenting our CD approach for network meta-analysis, we review the combining CD procedure for the univariate case in Section 2.1 and then extend it to the multivariate case in Section 2.2.

### 2.2.1 Review of CD Approach for Univariate Meta-Analysis

We first consider the special case where the parameter of interest is univariate. Model (2.1) simplifies to model (2)-(3) of Normand (1999); i.e.,

$$y_i|\theta_i, \sigma_i^2 \overset{\text{ind}}{\sim} N(\theta_i, \sigma_i^2), \quad \theta_i|\theta, \tau^2 \overset{\text{ind}}{\sim} N(\theta, \tau^2), \quad i = 1, 2, \ldots, k \tag{2.2}$$

where $\theta_i$ is the study-specific mean (random-effect) and $\theta$ and $\tau^2$ are hyper-parameters for $\theta_i$.

For the univariate case, meta-analysis estimators used in current practice (c.f., Table IV of Normand, 1999) can all be obtained through the unifying framework developed by Xie

et al. (2011) using the CD concept. A CD has been loosely referred to as a distribution function on the parameter space that can represent confidence intervals of all levels for a given parameter of interest. More specifically, the following formal definition is proposed in Schweder and Hjort (2002) and Singh et al. (2005, 2007):

**Definition 2.1** *Suppose* $\Theta$ *is the parameter space of the unknown parameter of interest* $\theta$*, and* $\mathcal{Y}$ *is the sample space corresponding to data* $\mathbf{Y} = \{y_1, \ldots, y_n\}$*. Then a function* $H(\cdot) = H(\mathbf{Y}, \cdot)$ *on* $\mathcal{Y} \times \Theta \to [0, 1]$ *is a confidence distribution (CD) if:*
*(i) For each given* $\mathbf{Y} \in \mathcal{Y}$*,* $H(\cdot)$ *is a continuous cumulative distribution function on* $\Theta$*; and*
*(ii) At the true parameter value* $\theta = \theta_0$*,* $H(\theta_0) = H(\mathbf{Y}, \theta_0)$*, as a function of the sample* $\mathbf{Y}$*, follows the uniform distribution* $U[0, 1]$*.*
*The function* $H(\cdot)$ *is an asymptotic CD (aCD) if the* $U[0, 1]$ *requirement holds only asymptotically and the continuity requirement on* $H(\cdot)$ *is dropped.*

In other words, a confidence distribution is a function defined on both the parameter space and the sample space, satisfying requirements (i) and (ii). Requirement (i) simply says that a CD should be a distribution on the parameter space. Requirement (ii) imposes some restrictions to facilitate desirable frequentist properties such as unbiasedness, consistency and/or efficiency. The CD concept is broad, covering examples from regular parametric (fiducial distribution) to bootstrap distributions, significance functions (also called p-value functions), normalized likelihood functions, and, in some cases, Bayesian priors and posteriors; see, e.g., Singh et al. (2007) and Xie and Singh (2013). A CD can be used to draw various inferences for the unknown parameter. For example, the median/mean of the distribution function $H(\cdot)$ can be used as a point estimator of $\theta$, and the interval $(-\infty, H^{-1}(1-\alpha))$ forms a level $(1-\alpha)$ confidence interval, an immediate consequence of Requirement (ii).

**Example 2.1** *(CDs for univariate normal mean) Let* $\{y_i, i = 1, \ldots, n\}$ *be an iid sample from* $N(\theta, \sigma^2)$ *with mean* $\bar{y}$*. Suppose that the parameter* $\theta$ *is of primary interest. If* $\sigma^2$ *is known, then* $H_\Phi(\theta) = \Phi(\sqrt{n}(\theta - \bar{y})/\sigma)$ *satisfies the two requirements in Definition 2.1, and it is a CD for* $\theta$*. If* $\sigma^2$ *is unknown, one can show that* $H_t(\theta) = F_{t_{n-1}}(\sqrt{n}(\theta - \bar{y})/s)$ *is a CD for* $\theta$*. Here* $s^2$ *is the sample variance, and* $F_{t_{n-1}}$ *is the cumulative distribution function of the student-t distribution with* $(n-1)$ *degrees of freedom. However,* $H_A(\theta) = \Phi(\sqrt{n}(\theta - \bar{y})/s)$ *is only an asymptotic CD for* $\theta$*.*

To combine individual CDs $H_i(\theta) = H_i(\boldsymbol{y}_i, \theta), i = 1, \ldots, k$, Singh et al. (2005) proposed a general recipe that uses a coordinate-wise monotonic function that maps the $k$-dimensional cube $[0, 1]^k$ to the real line. Specifically, a combined CD can be constructed following

$$H^{(c)}(\theta) = G^{(c)}\{g^{(c)}(H_1(\theta), \ldots, H_k(\theta))\}, \tag{2.3}$$

where the function $G^{(c)}$ is defined as $G^{(c)}(t) = \Pr\{g^{(c)}(U_1, \ldots, U_k) \leq t\}$ in which $U_1, \ldots, U_k$ are independent $U[0, 1]$ random variables. Xie et al. (2011) applied this general recipe to meta-analysis, with a special choice of $g^{(c)}$:

$$g^{(c)}(u_1, \ldots, u_k) = \widetilde{w}_1 a_0(u_1) + \cdots + \widetilde{w}_k a_0(u_k), \tag{2.4}$$

where $a_0(\cdot)$ is a given monotonic function and $\widetilde{w}_i \geq 0$, with at least one $\widetilde{w}_i \neq 0$, are generic weights for the combination. Xie et al. (2011) and subsequent research showed that, with suitable choices of $g^{(c)}$, almost all combining methods currently used in meta-analysis can be unified under the framework of Equation (2.3), including p-value combination methods, model-based meta-analysis (fixed-effect and random-effects models), the Mantel-Haenszel method, Peto's method, and also the method in Tian et al. (2009) by combining confidence intervals.

For the special model in (2.2), one can construct $H_i(\theta) = \Phi((\theta - y_i)/(\sigma_i^2 + \tau^2)^{1/2})$ based on the $i$th study and take $a_0(\cdot) = \Phi^{-1}(\cdot)$ and $\widetilde{w}_i = 1/(\sigma_i^2 + \tau^2)^{1/2}$ in (2.4). Here $\tau^2$ is assumed known. If $\tau^2$ is unknown, one can replace it with the DerSimonian and Laird estimator $\widehat{\tau}^2_{\texttt{DL}}$ (DerSimonian and Laird, 1986) or preferably the restricted-maximum-likelihood estimator $\widehat{\tau}^2_{\texttt{REML}}$. Then the combined CD function for $\theta$ is

$$H^{(c)}(\theta) = \Phi\left(\left(\sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \tau^2}\right)^{1/2} (\theta - \widehat{\theta}^{(c)})\right), \tag{2.5}$$

where $\widehat{\theta}^{(c)} = \{\sum_{i=1}^k \frac{y_i}{\sigma_i^2+\tau^2}\}/\{\sum_{i=1}^k \frac{1}{\sigma_i^2+\tau^2}\}$. The combined CD function is normal with mean $\widehat{\theta}^{(c)}$ and variance $s_c^2 = \{\sum_{i=1}^k \frac{1}{\sigma_i^2+\tau^2}\}^{-1}$, which is ready for making point estimates and constructing confidence intervals for the parameter $\theta$.

From Definition 2.1, a CD function $H(\cdot)$ is a cumulative distribution function on the parameter space for each given sample $\mathbf{Y}_n$. Thus, we can construct a random variable $\xi$ defined on $\mathcal{Y} \times \Theta$ such that, conditional on the sample, $\xi$ has the distribution $H(\cdot)$. We call this random variable $\xi$ a *CD random variable* (see, e.g., Singh et al., 2007; Xie and Singh, 2013). Conversely, suppose we have a CD random variable $\xi \in \mathcal{Y} \times \Theta$ whose conditional distribution, conditional on the sample, has a cumulative distribution function $H(\cdot)$. Then $H(\cdot)$ is a CD for the parameter of interest $\theta$.

We can express the normal CD combination (2.5) as a combination of CD random variables. Specifically, for a CD-random variable $\xi_i|y_i \sim H_i(\theta) = \Phi((\theta-y_i)/(\sigma_i^2+\tau^2)^{1/2})$ derived from the $i$-th study, we can define $\xi^{(c)} = \sum_{i=1}^k w_i\xi_i$, where $w_i = 1/(\sigma_i^2+\tau^2)$, and its corresponding combined CD

$$H^{(c)}(\theta) = \Pr(\xi^{(c)} \le \theta|data), \quad \text{for any } \theta \in \Theta. \tag{2.6}$$

It is straightforward to show that the $H^{(c)}(\cdot)$ defined in (2.6) is the same as the one defined

in (2.5).

The concept of CD random variable has been investigated in several recent publications. Xie and Singh (2013) explored the connection of CD random variables with bootstrap estimators when the bootstrap approach applies. Hannig and Xie (2012) discussed the association of a CD random variable with the so-called *belief random set*, a fundamental concept in the Dempster-Shafer theory of belief functions (cf. Dempster, 2008; Martin and Liu, 2013).

### 2.2.2 A General Procedure to Combine Multivariate Normal CDs

Constructing and combining CDs for multi-dimensional parameters is not a straightforward extension of the univariate case. One difficulty is that the cumulative distribution function is not a useful notion in the multivariate case, because (a) the region $F(\mathbf{y}) \leq \alpha$ is not of main interest and (b) the property $F(\boldsymbol{Y}) \overset{L}{=} U[0,1]$ when $\boldsymbol{Y} \overset{L}{=} F$ does not hold in $\Re^p$ (Singh et al., 2007). Research thus far suggests that we either limit our interest to center-outward confidence regions (instead of all Borel sets) in the $p \times 1$ parameter space or use asymptotic normality; see Xie and Singh (2013) and also De Blasi and Schweder (2012). In the present context, it suffices to consider only the multivariate normal CDs because individual CDs are based on asymptotic normality. We use a multivariate normal CD definition proposed in Singh et al. (2007). Intuitively, a distribution function $H(\cdot)$ is a multivariate normal CD for a $p \times 1$ vector $\boldsymbol{\theta}$ if and only if the projected distribution of $H(\cdot)$ on any direction $\boldsymbol{\lambda} \in \Re^p$, $||\boldsymbol{\lambda}||_2 = 1$, is a univariate normal CD for $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$. Here is a formal definition of a multivariate normal CD:

**Definition 2.2** *Let $\boldsymbol{\xi}$ be a random vector on $\Re^p$. For any given $p \times 1$ vector $\boldsymbol{\lambda}$, $||\boldsymbol{\lambda}||_2 = 1$, we denote by $H_{\boldsymbol{\lambda}}(\cdot)$ the conditional distribution of $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\xi}$ given $\mathbf{Y}$. We also denote by $H(\cdot)$ the conditional distribution of $\boldsymbol{\xi}$ given $\mathbf{Y}$. Then we call $H(\cdot)$ the multivariate normal CD*

*(or, asymptotic multivariate normal CD) for a $p \times 1$ parameter vector $\boldsymbol{\theta}$ if and only if, for any given $\boldsymbol{\lambda}$, $H_{\boldsymbol{\lambda}}(\cdot)$ is a univariate normal CD (or asymptotic CD) function for $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$. Also, the random vector $\boldsymbol{\xi}$ is called a CD random vector for $\boldsymbol{\theta}$.*

**Example 2.2** *(CDs for multivariate normal mean) Suppose $\boldsymbol{y}_i, i = 1, \ldots, n$ are identically and independently distributed observations from a multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma}$ is known, then the sample-dependent distribution $N(\bar{\boldsymbol{y}}, \boldsymbol{\Sigma})$ is a multivariate normal CD function for $\boldsymbol{\theta}$, where $\bar{\boldsymbol{y}}$ is the sample mean. If $\boldsymbol{\Sigma}$ is unknown but can be estimated consistently, say by $\widehat{\boldsymbol{\Sigma}}$, then the sample-dependent distribution $N(\bar{\boldsymbol{y}}, \widehat{\boldsymbol{\Sigma}})$ is an asymptotic multivariate normal CD function for $\boldsymbol{\theta}$.*

The CD combination method for the multivariate case cannot be easily specified by following (2.3) and (2.4), especially under the setting of (2.1), where $p_i$ may differ. Instead, we utilize the concept of CD random vector and an extension of (2.6) to propose the following scheme for combining multivariate normal CDs.

**Theorem 2.1** *Let $H_i(\boldsymbol{\theta}_i) \equiv H_i(\mathbf{Y}_i, \boldsymbol{\theta}_i), i = 1 \ldots, k$ are multivariate normal CD functions for the multivariate parameters $\boldsymbol{\theta}_i$ from $k$ independent samples $\mathbf{Y}_i$, where $\boldsymbol{\theta}_i = \mathbf{A}_i\boldsymbol{\theta}$ for the same $p$-dimensional target parameter vector $\boldsymbol{\theta}$. Additionally, let $\boldsymbol{\xi}_i$ be the CD random vector for $\boldsymbol{\theta}_i$. For any $\mathbf{t} \in \Re^p$, we define*

$$H^{(c)}(\mathbf{t}) = \Pr\left\{ \left( \sum_{i=1}^{k} W_i \right)^{-1} \sum_{i=1}^{k} W_i \mathbf{A}_i^{+} \boldsymbol{\xi}_i \leq \mathbf{t} \middle| \mathbf{Y}_1, \ldots, \mathbf{Y}_k \right\}, \tag{2.7}$$

*where $\mathbf{A}_i^{+}$ is the Moore-Penrose pseudo-inverse of $\mathbf{A}_i$. Then $H^{(c)}(\cdot) = H(\mathbf{Y}_1, \ldots, \mathbf{Y}_k; \cdot)$ is a multivariate normal CD for $\boldsymbol{\theta}$ provided the following conditions hold:*

*(1) Each $p \times p$ matrix $W_i$ is positive semi-definite.*

*(2)* $\mathscr{C}(W_i) = V_i$*, where* $\mathscr{C}(W_i)$ *is the column space of* $W_i$ *and* $V_i$ *is the row space of* $\mathbf{A}_i$*.*
*(3)* $V_1 + V_2 + \cdots + V_k = \Re^p$*, where* $V_1 + V_2 + \cdots + V_k \overset{\triangle}{=} \{\sum_{i=1}^k v_i | v_i \in V_i, i = 1, \ldots, k\}$*.*

In Theorem 1, conditions (2) and (3) state that, even if $\text{rank}(\mathbf{A}_i) < p$ for all $i$, so that $\boldsymbol{\theta}$ is not identifiable in any individual study, we can still derive a multivariate normal CD for $\boldsymbol{\theta}$ as long as the treatments are connected in a network.

Recall the multivariate model introduced in (2.1). We first consider the case in which $\boldsymbol{\Sigma}_i$ and $\mathbf{S}$ are known. From Example 2.2, we know that $N(\boldsymbol{y}_i, \boldsymbol{\Sigma}_i + \mathbf{A}_i \mathbf{S} \mathbf{A}_i^{\mathbf{T}})$ is a multivariate normal CD function for $\boldsymbol{\theta}_i$ based on the $i$-th study. Let $\boldsymbol{\xi}_i$ be the corresponding CD random vector for inference on $\boldsymbol{\theta}_i$ and $W_i = \mathbf{A}_i^+ (\boldsymbol{\Sigma}_i + \mathbf{A}_i \mathbf{S} \mathbf{A}_i^{\mathbf{T}})^{-1} \mathbf{A}_i$. It follows that $\left(\sum_{i=1}^k W_i\right)^{-1} \sum_{i=1}^k W_i \mathbf{A}_i^+ \boldsymbol{\xi}_i$ is normally distributed with mean vector $\widehat{\boldsymbol{\theta}}^{(c)} = (\sum_{i=1}^k W_i)^{-1} (\sum_{i=1}^k W_i \mathbf{A}_i^+ \boldsymbol{y}_i)$ and variance $\mathbf{S}_c = (\sum_{i=1}^k W_i)^{-1}$, given the sample. Thus, following the recipe in Equation (2.7), the combined CD for $\boldsymbol{\theta}$ is

$$H^{(c)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\left(\mathbf{S}_c^{-1/2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(c)})\right) \tag{2.8}$$

where $\boldsymbol{\Psi}(\cdot)$ is the cdf of the standard $p \times 1$ multivariate normal distribution function. Conditions (1) and (2) of Theorem 2.1 are satisfied by the specification of $W_i$, and condition (3) is satisfied as long as the comparisons involved in the studies form a connected network. Based on the combined multivariate CD function in (2.8), we can use $\widehat{\boldsymbol{\theta}}^{(c)}$ as a point estimator for $\boldsymbol{\theta}$ with variance $\mathbf{S}_c$. Furthermore, inferences on any linear contrasts $\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\theta}$ of $\boldsymbol{\theta}$ can be obtained from $\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\xi}^{(c)}$, where $\boldsymbol{\xi}^{(c)}$ follows the distribution specified in Equation (2.8).

If $\boldsymbol{\Sigma}_i$ and $\mathbf{S}$ are unknown, we can replace them with the sample estimators $\widehat{\boldsymbol{\Sigma}}_i$ and $\mathbf{S}_{\texttt{REML}}$. Then, as long as these estimators are consistent, the distribution $N(\boldsymbol{y}_i, \widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \boldsymbol{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})$

is asymptotically a multivariate normal CD for $\boldsymbol{\theta}_i$. Here $\widehat{\boldsymbol{\Sigma}}_i$ is the sample covariance matrix, and $\mathbf{S}_{\texttt{REML}}$ is the restricted-maximum-likelihood estimator of the heterogeneity between studies. As a result, the combined CD function (2.8) is an asymptotic multivariate normal CD for $\boldsymbol{\theta}$ with $\boldsymbol{\Sigma}_i$ and $\mathbf{S}$ replaced by $\widehat{\boldsymbol{\Sigma}}_i$ and $\mathbf{S}_{\texttt{REML}}$, respectively. For the estimation of $\mathbf{S}$, Jackson et al. (2010) developed a direct extension of the DerSimonian and Laird estimator of heterogeneity to multivariate case. Hereafter, we denote by $\mathbf{S}_{\texttt{DL}}$ and $\mathbf{S}_{\texttt{REML}}$ respectively the estimator derived from Jackson et al. (2010) and the restricted-maximum-likelihood estimator. We apply and examine both estimators in our numerical study of real examples and simulations in Sections 3 and 4. Further discussions on the performance of the DL and REML estimators for the heterogeneity in univariate random-effects models can be found in Sidik and Jonkman (2007) and Thorlund et al. (2011).

## 2.3 Real Data Examples

In this section, we illustrate the proposed CD approach for network meta-analysis using two real data examples, one on coronary artery disease and the other on cirrhosis. For comparison, we also include the traditional pairwise meta-analysis and the Bayesian hierarchical model.

### 2.3.1 An Example on Coronary Artery Disease (CAD)

Stettler et al. (2007) used data from a network of 37 trials to compare the performance of three types of stent: bare metal stent (BMS), sirolimus-eluting stent (SES), and paclitaxel-eluting stent (PES), in patients with coronary artery disease. Each trial involved at least two of the three treatments; we analyze the data on a negative outcome, whether patients required target lesion revascularisation (TLR) within one year (cf. Figure 2.1). One trial,

TAXUS I, had zero events and is thus excluded from the analysis. Of the remaining 36 trials, listed in Table 2.1, 15 trials compared BMS with SES, 6 trials compared BMS with PES, 14 trials compared SES with PES, and 1 trial compared all three treatments. The network is connected, so simultaneous inference on the treatment effects is possible.

#### 2.3.1.1 A Multivariate Random-Effects Model

We use treatments A, B, and C to denote the three types of stents BMS, SES and PES, respectively. We use $T_i$ to denote the set of treatments compared in the $i$-th trial; for example, $T_i = \{A, C\}$ for TAXUS IV. Further, let $n_{ij}$ and $r_{ij}$ be the number of total patients and number of patients who experienced a TLR in the $i$-th study with treatment $j$. Then with a binary individual responses we would assume

$$r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), \quad i = 1, 2, \ldots, 36, j \in T_i \tag{2.9}$$

where $p_{ij}$ denotes the probability that a patient on treatment $j$ experiences an event in the $i$-th trial.

The target parameter is $\mathbf{p} = (p_A, p_B, p_C)^{\mathbf{T}}$, the overall probability of an event for BMS, SES, and PES, respectively. In practice, one often applies a log transformation to the observed odds of an event. Owing to the rapid convergence to a normal distribution on the log-odds scale, it is customary to consider a general random-effects model for $\boldsymbol{\theta}_i = (\text{logit}(p_{ij}))^{\mathbf{T}}, \forall j \in T_i$ with parameter $\boldsymbol{\theta} = (\text{logit}(p_A), \text{logit}(p_B), \text{logit}(p_C))^{\mathbf{T}}$; cf. DerSimonian and Laird (1986); Normand (1999). Here, $\text{logit}(p) = \log(p/(1-p))$. Specifically, we have

$$\begin{aligned} &\text{level 1:} \quad r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), \quad i = 1, 2, \ldots, 36, j \in T_i \\ &\text{level 2:} \qquad\qquad \boldsymbol{\theta}_i \sim N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}) \end{aligned} \tag{2.10}$$

Table 2.1: CAD Trial Data, Target Lesion Revascularisation at 1 year

| Study | BMS (A) | | SES (B) | | PES (C) | |
|---|---|---|---|---|---|---|
| | $r_{ij}$ | $n_{ij}$ | $r_{ij}$ | $n_{ij}$ | $r_{ij}$ | $n_{ij}$ |
| BASKET | 35 | 281 | 25 | 264 | 25 | 281 |
| C-SIRIUS | 11 | 50 | 2 | 50 | — | — |
| DECODE | 8 | 29 | 5 | 54 | — | — |
| DIABETES | 27 | 80 | 6 | 80 | — | — |
| E-SIRIUS | 44 | 177 | 8 | 175 | — | — |
| Ortolani 2007 | 11 | 52 | 6 | 52 | — | — |
| Pache 2005 | 51 | 250 | 25 | 250 | — | — |
| PRISON II | 20 | 100 | 4 | 100 | — | — |
| RAVEL | 16 | 118 | 1 | 120 | — | — |
| RRISC | 10 | 37 | 6 | 38 | — | — |
| SCANDSTENT | 47 | 159 | 4 | 163 | — | — |
| SCORPIUS | 20 | 95 | 5 | 95 | — | — |
| SESAMI | 19 | 160 | 7 | 160 | — | — |
| SES-SMART | 27 | 128 | 9 | 129 | — | — |
| SIRIUS | 106 | 525 | 26 | 533 | — | — |
| TYPHOON | 45 | 357 | 13 | 355 | — | — |
| HAAMUS-TENT | 9 | 82 | — | — | 3 | 82 |
| PASSION | 23 | 309 | — | — | 16 | 310 |
| TAXUS II | 39 | 269 | — | — | 13 | 260 |
| TAXUS IV | 96 | 652 | — | — | 28 | 662 |
| TAXUS V | 107 | 579 | — | — | 62 | 577 |
| TAXUS VI | 46 | 227 | — | — | 19 | 219 |
| Cervinka 2006 | — | — | 1 | 37 | 2 | 33 |
| CORPAL | — | — | 22 | 331 | 25 | 321 |
| Han 2006 | — | — | 9 | 202 | 11 | 196 |
| ISAR-DESIRE | — | — | 14 | 100 | 22 | 100 |
| ISAR-DIABETES | — | — | 9 | 125 | 15 | 125 |
| ISAR-SMART3 | — | — | 16 | 180 | 29 | 180 |
| LONG DES II | — | — | 6 | 250 | 18 | 250 |
| Petronio 2007 | — | — | 1 | 42 | 1 | 43 |
| PROSIT | — | — | 3 | 116 | 9 | 115 |
| REALITY | — | — | 44 | 684 | 43 | 669 |
| SIRTAX | — | — | 30 | 503 | 54 | 509 |
| SORT OUT II | — | — | 40 | 1065 | 46 | 1033 |
| TAXi | — | — | 4 | 102 | 2 | 100 |
| Zhang 2006 | — | — | 14 | 225 | 16 | 187 |

where $\mathbf{A}_i$ is the selection matrix associated with $T_i$; for example, $\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ if $T_i = \{A, B\}$, $\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ if $T_i = \{A, C\}$, $\mathbf{A}_i = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ if $T_i = \{B, C\}$, and $\mathbf{A}_i = \mathbf{I}_3$ if $T_i = \{A, B, C\}$.

Further, let $y_{ij} = \log\left(\frac{r_{ij}}{n_{ij} - r_{ij}}\right), \widehat{\sigma}_{ij}^2 = \frac{1}{r_{ij}} + \frac{1}{n_{ij} - r_{ij}}$ and $\boldsymbol{y}_i = [y_{ij}, j \in T_i]^{\mathbf{T}}, \widehat{\boldsymbol{\Sigma}}_i = \text{diag}(\widehat{\sigma}_{ij}^2, j \in T_i)$. Then an asymptotically equivalent model is

$$\begin{aligned} &\text{level 1:} \quad \boldsymbol{y}_i | \boldsymbol{\theta}_i \sim N(\boldsymbol{\theta}_i, \widehat{\boldsymbol{\Sigma}}_i), \quad i = 1, 2, \ldots, 36 \\ &\text{level 2:} \qquad\quad \boldsymbol{\theta}_i \sim N(\mathbf{A}_i \boldsymbol{\theta}, \mathbf{A}_i \mathbf{S} \mathbf{A}_i^{\mathbf{T}}). \end{aligned} \tag{2.11}$$

Finally, if that our primary concern is the efficacy of SES vs BMS, the parameter of interest is the log-odds ratio reflecting the relative efficacy of treatment B vs A, that is $\delta_{AB} \equiv \theta_B - \theta_A$. We proceed to compare the results obtained from the proposed CD procedure with those from the traditional pairwise meta-analysis and the Bayesian network meta-analysis.

#### 2.3.1.2 The CD approach

Consider the random-effects model in (2.11). We estimate the covariance matrix $\mathbf{S}$ by the restricted-maximum-likelihood estimator $\mathbf{S}_{\texttt{REML}}$. We can construct a multivariate normal aCD function for $\boldsymbol{\theta}_i$ based on the $i$-th individual study, namely $N(\boldsymbol{y}_i, \widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \mathbf{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})$. We use $\boldsymbol{\xi}_i | \boldsymbol{y}_i \sim N(\boldsymbol{y}_i, \widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \mathbf{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})$ to denote the associated CD random variable and take $W_i = \mathbf{A}_i^{+} (\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \mathbf{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})^{-1} \mathbf{A}_i$. Then, by (2.8), $H^{(c)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\mathbf{S}_c^{-1/2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(c)}))$ is the combined CD for $\boldsymbol{\theta}$, where $\widehat{\boldsymbol{\theta}}^{(c)} = (\sum_{i=1}^k W_i)^{-1} (\sum_{i=1}^k W_i \mathbf{A}_i^{+} \boldsymbol{y}_i)$ and $\mathbf{S}_c = (\sum_{i=1}^k W_i)^{-1}$. Since we have $\mathbf{A}_i^{+} = \mathbf{A}_i^{\mathbf{T}}$ in the current case, we can replace $\mathbf{A}_i^{+}$ with $\mathbf{A}_i^{\mathbf{T}}$ in the above

formulas.

To make inferences for $\delta_{AB} \equiv \theta_B - \theta_A$, we can use the marginal distribution of $\boldsymbol{\lambda}_{AB}^{\mathbf{T}}\boldsymbol{\xi}^{(c)}$ where $\boldsymbol{\lambda}_{AB} = (-1, 1, 0)^{\mathbf{T}}$ and $\boldsymbol{\xi}^{(c)}|data \sim N(\widehat{\boldsymbol{\theta}}^{(c)}, \mathbf{S}_c)$. Therefore, the point estimator $\widehat{\delta}_{AB}$ and its variance based on the CD procedure are

$$\widehat{\delta}_{AB} = \boldsymbol{\lambda}_{AB}^{\mathbf{T}} \left\{ \sum_{i=1}^{k} \mathbf{A}_i^{+} (\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \mathbf{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})^{-1} \mathbf{A}_i \right\}^{-1} \sum_{i=1}^{k} \mathbf{A}_i^{+} (\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \mathbf{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})^{-1} \mathbf{A}_i \mathbf{A}_i^{\mathbf{T}} \boldsymbol{y}_i$$

$$\text{var}(\widehat{\delta}_{AB}) = \boldsymbol{\lambda}_{AB}^{\mathbf{T}} \left\{ \sum_{i=1}^{k} \mathbf{A}_i^{+} (\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \mathbf{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})^{-1} \mathbf{A}_i \right\}^{-1} \boldsymbol{\lambda}_{AB}.$$

In practice, we might also be interested in simultaneous inferences on, say, $q$ linear combinations of $\boldsymbol{\theta}$, e.g., $\mathbf{Q}\boldsymbol{\theta}$ where $\mathbf{Q} \in \Re^{q \times p}$. The Bayesian approach often uses the marginal posterior distribution of $\mathbf{Q}\boldsymbol{\theta}$ as the basis for statistical inference. Similarly, to draw inferences for $\boldsymbol{\theta}$, the proposed CD network meta-analysis approach can use the marginal distribution of $\mathbf{Q}\boldsymbol{\xi}^{(c)}$ given the data. Here $\boldsymbol{\xi}^{(c)}$ is the CD random vector associated with the combined CD function $H^{(c)}(\cdot)$ for $\boldsymbol{\theta}$.

#### 2.3.1.3 Traditional Pairwise Meta-Analysis

A traditional meta-analysis for such a problem uses only the direct evidence, e.g., clinical trials that explicitly compared BMS vs SES; see, e.g., Simmonds and Higgins (2007) and Hoaglin et al. (2011). Let $\widehat{\delta}_{AB,i} = \log\left(\dfrac{r_{iB}(n_{iA} - r_{iA})}{r_{iA}(n_{iB} - r_{iB})}\right)$ for $A, B \in T_i$. A random-effects model (DerSimonian and Laird, 1986) is considered:

$$\begin{aligned} &\text{level 1:} \quad \widehat{\delta}_{AB,i} \sim N(\delta_{AB,i}, \sigma_{AB,i}^2), \quad i \text{ s.t. } A, B \in T_i \\ &\text{level 2:} \qquad\qquad \delta_{AB,i} \sim N(\delta_{AB}, \tau_{AB}^2). \end{aligned} \tag{2.12}$$

An overall estimate of the common log-odds ratio $\delta_{AB}$, based on the direct evidence, is often a weighted average of the estimates $\widehat{\delta}_{AB,i}$ from individual studies (Hardy and Thompson, 1996):

$$\widehat{\delta}_{AB,direct} = \frac{\sum_i w_i \widehat{\delta}_{AB,i}}{\sum_i w_i} \quad \text{with} \quad \text{var}(\widehat{\delta}_{AB}) = \frac{1}{\sum_i w_i}, \tag{2.13}$$

where the weight $w_i$ is often taken as the empirical weight determined by the reciprocal of the variance $\sigma^2_{AB,i}$ adjusted to incorporate the heterogeneity $\tau^2_{AB}$, for example $w_i = 1/(\sigma^2_{AB,i} + \tau^2_{AB})$, as suggested in DerSimonian and Laird (1986).

In practice, when the variance $\sigma^2_{AB,i}$ and the heterogeneity $\tau^2_{AB}$ are unknown, they are often replaced by their corresponding estimates $\widehat{\sigma}^2_{AB,i}$ and $\widehat{\tau}^2_{AB}$, where $\widehat{\sigma}^2_{AB,i} = \frac{1}{r_{iA}} + \frac{1}{n_{iA}-r_{iA}} + \frac{1}{r_{iB}} + \frac{1}{n_{iB}-r_{iB}}$, provided that $r_{ij} \neq 0$ and $r_{ij} \neq n_{ij}$, and $\widehat{\tau}^2_{AB}$ is the REML estimate.

Similarly, we can obtain estimates $\widehat{\delta}_{AC}$ and $\widehat{\delta}_{BC}$ for the pairwise comparisons of BMS vs PES and SES vs PES, respectively, based on the 7 and 15 trials that compared them directly. Then an indirect comparison of BMS vs SES can be obtained by taking

$$\widehat{\delta}_{AB,indirect} = \widehat{\delta}_{AC} - \widehat{\delta}_{BC} \quad \text{and} \quad \text{var}(\widehat{\delta}_{AB}) = \text{var}(\widehat{\delta}_{AC}) + \text{var}(\widehat{\delta}_{BC}). \tag{2.14}$$

We can then combine the $\widehat{\delta}_{AB,direct}$ and $\widehat{\delta}_{AB,indirect}$ to obtain an estimator that integrates the two sources of information, provided that the direct and indirect comparisons are consistent with each other or at least not contradictory. Here is a simple illustration of inconsistent/contradictory evidence: the direct comparison concludes that the effect of treatment X is larger than that of treatment Y, but the indirect comparison concludes the opposite. Some discussion on issues of inconsistent evidence in network meta-analysis can be found in Lumley (2002), Lu and Ades (2006), and Dias et al. (2010).

Although one can always apply the procedure above to combine the direct and indirect

estimates, this procedure splits the three-arm trial into three two-arm trials and uses them for three difference estimates. This is a drawback for traditional pairwise meta-analysis — *Trials with more than two arms cannot be fully incorporated in the meta-analysis unless they are split into multiple two-arm trials. Those two-arm trials are treated as if they were independent; whereas they came from the same trial.* Consequently, such a network meta-analysis often incurs bias and loss of efficiency, as observed in Jansen et al. (2011) and Hoaglin et al. (2011). Taking into account this drawback, we consider $\widehat{\delta}_{AB,direct}$ and $\widehat{\delta}_{AB,indirect}$ as two separate estimators of $\delta_{AB}$ in the analysis in later sections.

We show later that the CD approach can combine the direct and indirect evidence for $\delta_{AB}$ efficiently, provided that the observed evidences from the direct and indirect comparisons are consistent with each other or at least not contradictory.

#### 2.3.1.4 Bayesian Hierarchical Model

Similar to the CD approach, a Bayesian approach can also incorporate all trials. However, the Bayesian approach has to rely on prior distributions, which then impose additional assumptions.

To carry out network meta-analysis on clinical trials with direct and indirect treatment comparisons, Lu and Ades (2004, 2006) proposed the following hierarchical Bayesian model:

$$\begin{aligned}
&\text{level 1:} \quad r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), i = 1, 2, \ldots, 36, j = A, B, C \\
&\text{level 2:} \quad (\delta_{AB,i}, \delta_{AC,i})^{\mathbf{T}}|\boldsymbol{\delta}, \mathbf{C} \sim N(\boldsymbol{\delta}, \mathbf{C}) \quad \perp \quad \mu_i|\mu, \sigma_\mu^2 \sim N(\mu, \sigma_\mu^2) \\
&\text{level 3:} \quad \text{hyper prior distributions for } \boldsymbol{\delta},\ \mathbf{C} \\
&\qquad\quad \text{and parameters in the distribution of } \mu, \sigma_\mu^2 \text{ if necessary}
\end{aligned} \tag{2.15}$$

where

$$\begin{bmatrix} \delta_{AB,i} \\ \delta_{AC,i} \\ \mu_i \end{bmatrix} = \mathbf{T}_{\mathtt{BS}} \begin{bmatrix} \text{logit}(p_{iA}) \\ \text{logit}(p_{iB}) \\ \text{logit}(p_{iC}) \end{bmatrix} \text{ and } \mathbf{T}_{\mathtt{BS}} \stackrel{\triangle}{=} \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

As stated in Lu and Ades (2004), this model extends the one proposed by Smith et al. (1995) to address the issues of incorporating indirect comparisons and to fully incorporate trials with more than two arms.

Specifically, Lu and Ades (2004) considered two sets of prior distributions, Bayesian-HOM prior and Bayesian-HET prior. The first set of prior distributions ("Bayesian-HOM") assumes a homogenous variance for $\delta_{AB,i}$ and $\delta_{AC,i}$:

$$\begin{gathered} \boldsymbol{\delta} \sim N(0, 10^3 \mathbf{I}_2) \\ \mathbf{C} = \sigma^2 \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}, \quad \sigma^{-2} \sim \text{Gamma}(10^{-3}, 10^{-3}) \\ \mu \sim N(0, 10^3), \quad \sigma_\mu^{-2} \sim \text{Gamma}(10^{-3}, 10^{-3}) \end{gathered} \tag{2.16}$$

The second set of prior distributions ("Bayesian-HET") allows heterogenous variances for $\delta_{AB,i}$ and $\delta_{AC,i}$:

$$\begin{gathered} \boldsymbol{\delta} \sim N(0, 10^3 \mathbf{I}_2) \\ \mathbf{C} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \quad \text{where } \rho = 0.5 \\ \sigma_j^2 \sim \text{Gamma}(a, b), a \sim \text{Exp}(0.01), b \sim \text{Gamma}(10^{-3}, 10^{-3}), j = 1, 2 \\ \mu \sim N(0, 10^3), \quad \sigma_\mu^{-2} \sim \text{Gamma}(10^{-3}, 10^{-3}) \end{gathered} \tag{2.17}$$

Except for the different assumptions on the structure of covariance matrix $\mathbf{C}$, both

Bayesian-HOM and Bayesian-HET impose the same noninformative priors on $\boldsymbol{\delta}, \mu$, and $\sigma_\mu^2$. The assumptions of priors are subjective and often difficult to verify. Our numerical studies in Section 4 suggest that the Bayesian approach is sensitive to the choice of priors.

#### 2.3.1.5 Results

We consider the following six methods and compare their inferences on $\delta_{AB}$:

- Traditional-Direct: Traditional frequentist meta-analysis on direct pairwise comparisons.
- Traditional-Indirect: Traditional frequentist meta-analysis on indirect pairwise comparisons.
- Bayesian-HOM: Bayesian network meta-analysis with homogeneous variance structure on $\boldsymbol{\delta}$.
- Bayesian-HET: Bayesian network meta-analysis with heterogeneous variance structure on $\boldsymbol{\delta}$.
- CD[$\mathbf{S}_{\texttt{DL}}$]: The proposed CD procedure with $\mathbf{S}$ estimated by an extension of the DerSimonian and Laird method to the multivariate case (Jackson et al. (2010)).
- CD[$\mathbf{S}_{\texttt{REML}}$]: The proposed CD procedure with $\mathbf{S}$ estimated by maximizing restricted likelihood.

The values of $\widehat{\delta}_{AB}$ and its corresponding 95% confidence interval (CI) or 95% credible interval (CrI) from all six methods are summarized in Table 2.2.

Table 2.2 shows that all six methods yield similar point estimates of $\delta_{AB}$. However, because they use both direct and indirect evidence, the Bayesian methods and the CD methods yield

Table 2.2: Results of meta-analyses on CAD data

| Method | $\widehat{\delta}_{AB}$ | s.d.$(\widehat{\delta}_{AB})$ | 95% CI | Length of 95% CI |
|---|---|---|---|---|
| Traditional-Direct | $-1.3757$ | 0.1672 | $(-1.7035, -1.0479)$ | 0.6556 |
| Traditional-Indirect | $-1.2874$ | 0.5129 | $(-2.2926, -0.2822)$ | 2.0104 |
| Bayesian-HOM | $-1.3681$ | 0.1084 | $(-1.5900, -1.1650)$ | 0.4250 |
| Bayesian-HET | $-1.3770$ | 0.1312 | $(-1.6170, -1.1028)$ | 0.5142 |
| CD$[\mathbf{S}_{\mathtt{DL}}]$ | $-1.2984$ | 0.1174 | $(-1.5285, -1.0683)$ | 0.4602 |
| CD$[\mathbf{S}_{\mathtt{REML}}]$ | $-1.2957$ | 0.1096 | $(-1.5104, -1.0809)$ | 0.4295 |

smaller variance estimates and tighter confidence interval, in comparison with traditional pairwise meta-analysis. Also, the results from indirect comparisons are in line with those obtained from direct comparisons, although less efficient. It seems appropriate to combine the trials with direct and indirect evidence.

### 2.3.2 An Example on Cirrhosis

As another example, we consider the data presented in Pagliaro et al. (1992) and used in Lu and Ades (2004). The authors analyzed 26 trials of non-surgical treatments intended to prevent first bleeding in patients with cirrhosis and esophageal varices who had never bled, in order to assess the effectiveness of three types of treatments: beta-blockers, endoscopic sclerotherapy and non-active treatment (control), denoted by A, B, and C, respectively. Of the 26 trials, 2 trials compared all three treatments, 7 trials compared beta-blockers vs control, and 17 trials compared sclerotherapy vs control. In Table 2.3, for trial $i$ and treatment $j$, $r_{ij}$ is the number of patients who had a first bleeding event and $n_{ij}$ is the total number of patients. Our concern is with the relative performance of the active treatments: beta-blockers vs sclerotherapy. However, the only trials that compared them directly were the two three-arm trials, which were not sufficiently large. In this situation direct evidence is not strong enough, and incorporating indirect evidence is particularly important for

making inferences.

Table 2.3: Cirrhosis data: number of patients who had a first bleeding event.

| Study | Beta-blockers (A) | | Sclerotherapy (B) | | Control (C) | |
|---|---|---|---|---|---|---|
| | $r_{ij}$ | $n_{ij}$ | $r_{ij}$ | $n_{ij}$ | $r_{ij}$ | $n_{ij}$ |
| 1 | 2 | 43 | 9 | 42 | 13 | 41 |
| 2 | 12 | 68 | 13 | 73 | 13 | 72 |
| 3 | 4 | 20 | — | — | 4 | 16 |
| 4 | 20 | 116 | — | — | 30 | 111 |
| 5 | 1 | 30 | — | — | 11 | 49 |
| 6 | 7 | 53 | — | — | 10 | 53 |
| 7 | 18 | 85 | — | — | 31 | 89 |
| 8 | 2 | 51 | — | — | 11 | 51 |
| 9 | 8 | 23 | — | — | 2 | 25 |
| 10 | — | — | 4 | 18 | 0 | 19 |
| 11 | — | — | 3 | 35 | 22 | 36 |
| 12 | — | — | 5 | 56 | 30 | 53 |
| 13 | — | — | 5 | 16 | 6 | 18 |
| 14 | — | — | 3 | 23 | 9 | 22 |
| 15 | — | — | 11 | 49 | 31 | 46 |
| 16 | — | — | 19 | 53 | 9 | 60 |
| 17 | — | — | 17 | 53 | 26 | 60 |
| 18 | — | — | 10 | 71 | 29 | 69 |
| 19 | — | — | 12 | 41 | 14 | 41 |
| 20 | — | — | 0 | 21 | 3 | 20 |
| 21 | — | — | 13 | 33 | 14 | 35 |
| 22 | — | — | 31 | 143 | 23 | 138 |
| 23 | — | — | 20 | 55 | 19 | 51 |
| 24 | — | — | 3 | 13 | 12 | 16 |
| 25 | — | — | 3 | 21 | 5 | 28 |
| 26 | — | — | 6 | 22 | 2 | 24 |

We apply the same six methods as in the CAD data set. The parameter of interest is $\delta_{AB}$, the log-odds ratio of first bleeding for beta-blockers vs sclerotherapy. The results are presented in Table 2.4.

Table 2.4: Results of meta-analysis on cirrhosis data

| Method | $\widehat{\delta}_{AB}$ | s.d.$(\widehat{\delta}_{AB})$ | 95% CI | Length of 95% CI |
|---|---|---|---|---|
| Traditional-Direct | 0.7284 | 0.8439 | $(-0.9256, 2.3824)$ | 3.3080 |
| Traditional-Indirect | $-0.0927$ | 0.8069 | $(-1.6738, 1.4884)$ | 3.1622 |
| Bayesian-HOM | 0.5228 | 0.3171 | $(-0.0969, 1.1461)$ | 1.2430 |
| Bayesian-HET | 0.6466 | 0.3250 | ( 0.0410, 1.3151) | 1.2741 |
| CD[$\mathbf{S}_{\mathtt{DL}}$] | 0.5688 | 0.2588 | ( 0.0617, 1.0761) | 1.0144 |
| CD[$\mathbf{S}_{\mathtt{REML}}$] | 0.6381 | 0.2445 | ( 0.1589, 1.1174) | 0.9585 |

In Table 2.4, we again observe that the Bayesian methods and the CD procedures have substantially lower variance as a result of integrating all treatment comparisons. Therefore, the network-meta-analysis approaches have effectively strengthened the results obtained from direct comparisons by borrowing information from indirect comparisons. Unlike the results in the CAD example, pairwise meta-analysis using only direct comparisons does not achieve significant results, whereas the Bayesian and CD approaches yield significant or almost significant results. However, the validity of combining direct and indirect treatment comparisons should be carefully investigated, the difference between $\widehat{\delta}_{AB,indirect}$ and $\widehat{\delta}_{AB,direct}$ raises concerns about consistency between direct and indirect evidence. The topic of inconsistent evidence is discussed in Higgins et al. (2002, 2003). We also discuss this topic further in Section 4.3 and Section 5.

In these two examples, the CD and Bayesian approaches yield similar results. The confidence intervals derived from the CD approach are only slightly tighter than those derived from the Bayesian approach. However, our simulation studies in the next section show that the Bayesian credible intervals may not achieve the nominal coverage probability, and their empirical coverage probabilities may be far below the nominal level when the assumed prior on the between-trial covariance structure does not agree with the underlying true model. This latter condition is almost impossible to verify in practice. In contrast,

the proposed CD combining approach does not require any prior, and the derived confidence intervals can maintain adequate coverage probability regardless of the between-trial covariance structure.

## 2.4 Simulation Studies

We conducted simulation studies to compare the performance of the proposed CD combining approach with traditional pairwise meta-analysis and the Bayesian method.

### 2.4.1 Simulation Settings

We based our simulation on the structure of the cirrhosis data. Specifically, the evidence network involves three treatments (A, B, and C). The problem of interest is to infer the relative effectiveness of A vs B.

Consider two scenarios, one with 24 trials and the other with 96 trials. In the first scenario, the 24 clinical trials, comprise 1 trial comparing all three treatments, 3 trials comparing A and B, 10 trials comparing treatments A and C, and 10 trials comparing B and C. The number of patients in each arm of each trial is 100, i.e., $n_{ij} = 100, \forall i$ and $j \in T_i$. In the second scenario the number of trials of each type is four times that in the first scenario. The simulation is designed to show the benefit of borrowing strength from indirect evidence when direct evidence (trials directly comparing treatments A and B) is somewhat limited.

Table 2.5: Simulation Settings - Number of Trials $k$ and Patients Involved in Each Group $n_{ij}$

| Total Number of Trials $k$ \Type of Trial | | ABC | AB | AC | BC | $n_{ij}$ |
|---|---|---|---|---|---|---|
| Simulation Scenario 1 | k=24 | 1 | 3 | 10 | 10 | 100 |
| Simulation Scenario 2 | k=96 | 4 | 12 | 40 | 40 | 100 |

We generate the simulated data from the model:

$$r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), \quad p_{ij} = \frac{\exp(\theta_{ij})}{1+\exp(\theta_{ij})}, \quad i = 1, 2, \ldots, 24 \text{ or } 96, j \in T_i$$
$$\boldsymbol{\theta}_i \sim N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}) \tag{2.18}$$

where $\mathbf{A}_i$ consists of the rows of the identity matrix corresponding to the treatments in $T_i$.

We specify the true value of $\boldsymbol{\theta} = (-1.82, -1.21, -0.80)^{\mathbf{T}}$ as the values are close to those estimated from the cirrhosis data. It follows that the probabilities of observing an event in treatment A, B, and C are $\mathbf{p} = (0.14, 0.23, 0.31)^{\mathbf{T}}$. For the covariance matrix $\mathbf{S}$, we consider three cases:

Case 1:

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \iff \mathbf{B} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix};$$

Case 2:

$$\mathbf{S} = \begin{bmatrix} 2.5736 & -1.2868 & 1.7132 \\ -1.2868 & 4.8528 & -0.5660 \\ 1.7132 & -0.5660 & 1.8528 \end{bmatrix} \iff \mathbf{B} = \begin{bmatrix} 10 & 1.5811 & 0 \\ 1.5811 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

Case 3:

$$\mathbf{S} = \begin{bmatrix} 3.1070 & 0.4314 & 1.2358 \\ 0.4314 & 0.7557 & 0.4693 \\ 1.2358 & 0.4693 & 0.8645 \end{bmatrix} \iff \boldsymbol{B} = \begin{bmatrix} 3.0000 & 1.9092 & -1.0392 \\ 1.9092 & 1.5000 & -0.7348 \\ -1.0392 & -0.7348 & 1.0000 \end{bmatrix},$$

where $\mathbf{B} = \text{cov}(\delta_{AB,i}, \delta_{AC,i}, \mu_i) = \mathbf{T}_{\mathtt{BS}}\mathbf{S}\mathbf{T}_{\mathtt{BS}}^{\mathbf{T}}$, and $\delta_{AB,i}, \delta_{AC,i}, \mu_i$ and $\mathbf{T}_{\mathtt{BS}}$ are defined as in model (2.15). Here " $\iff$ " indicates the one-to-one correspondence between the covariance matrix $\mathbf{S}$ in model (2.18) and the covariance matrix $\mathbf{B}$ in the Bayesian models.

In Case 1, $\mathbf{S}$ is set to an identity matrix to ensure that the true model (2.18) meets the assumptions of Bayesian-HOM in Section 3.1.4, and is thus equivalent to the case of (2.16) with $\sigma^2 = 2$. Similarly, the covariance matrix $\mathbf{S}$ in Case 2 allows the true model (2.18) to meet the assumptions of Bayesian-HET, and is thus equivalent to the case of $\sigma_1^2 = 10, \sigma_2^2 = 1$ and $\rho = 0.5$ in (2.17). As suggested in Joseph et al. (1997), we further extend the model to incorporate correlations between $\delta_{AB,i}, \delta_{AC,i}$ and $\mu_i$, instead of assuming independence. Therefore, in Case 3, the covariance matrix $\mathbf{S}$ is specified to give an arbitrary covariance structure such that $\mathbf{B}$ fails to meet the assumptions of either Bayesian-HOM or Bayesian-HET.

In summary, we consider a total of six ($= 2 \times 3$) settings in our simulation study: 24 and 96 trials each with three specifications of the covariance matrix $\mathbf{S}$.

### 2.4.2 Results

We consider and compare the performance of a total of nine approaches. They include the six methods listed in Section 3.1.5: Traditional-Direct and Traditional-Indirect, Bayesian-HOM and Bayesian-HET, and CD[$\mathbf{S}_{\texttt{DL}}$] and CD[$\mathbf{S}_{\texttt{REML}}$]. Additionally, we include three other CD approaches: two semi-Bayesian approaches, CD[$\mathbf{S}_{\texttt{BHOM}}$] and CD[$\mathbf{S}_{\texttt{BHET}}$], in which the covariance matrix $\mathbf{S}$ is estimated by the Bayesian method with prior in (2.16) and (2.17), respectively, and CD[$\mathbf{S}_{\texttt{TRUE}}$], which uses the true covariance matrix $\mathbf{S}$. The CD[$\mathbf{S}_{\texttt{TRUE}}$] method allows us to separate the effect of estimating the mean alone and study the potential impacts on estimation of the mean when different approaches are used to estimate $\mathbf{S}$. Thus, the nine methods are:

- Traditional frequentist methods:

  - Traditional-Direct: Traditional frequentist meta-analysis of direct pairwise comparisons.
  - Traditional-Indirect: Traditional frequentist meta-analysis via indirect pairwise comparisons.
- Bayesian methods:
  - Bayesian-HOM: Bayesian network meta-analysis with homogenous variance structure on $\boldsymbol{\delta}$.
  - Bayesian-HET: Bayesian network meta-analysis with heterogenous variance structure on $\boldsymbol{\delta}$.
- CD methods:
  - CD[$\mathbf{S}_{\texttt{DL}}$]: $\mathbf{S}$ estimated by $\mathbf{S}_{\texttt{DL}}$.
  - CD[$\mathbf{S}_{\texttt{REML}}$]: $\mathbf{S}$ estimated by $\mathbf{S}_{\texttt{REML}}$.
  - CD[$\mathbf{S}_{\texttt{BHOM}}$]: $\mathbf{S}$ estimated by $\mathbf{S}_{\texttt{BHOM}}$.
  - CD[$\mathbf{S}_{\texttt{BHET}}$]: $\mathbf{S}$ estimated by $\mathbf{S}_{\texttt{BHET}}$.
  - CD[$\mathbf{S}_{\texttt{TRUE}}$]: using the known true $\mathbf{S}$.

In simulation Scenario 1 Case 1, for example, we generate data according to the model specified in (2.18), and then apply each method to estimate $\delta_{AB}$ and calculate the corresponding 95% confidence (credible) interval. We repeat this process 1000 times. For each method, we report the mean and standard deviation of the 1000 $\widehat{\delta}_{AB}$ and the percentage of times (coverage) that the 1000 95% CIs cover the true $\delta_{AB} = 0.6070$ and the average interval length. The results for Scenarios 1 and 2 with Case 1 ($\mathbf{S} = \mathbf{I}_{3\times 3}$) are presented in Table 2.6. Similarly, the results for Case 2 and Case 3 are presented in Tables 2.7 and 2.8. It is straightforward to verify that the chance that no trial has zero events in the entire

1000 replications is at least 99.97%. Thus the zero events issue is not considered in the simulation study.

Table 2.6: Summary of results of simulation studies - Case 1

| Method | $\widehat{\delta}_{AB}$ | s.d.$(\widehat{\delta}_{AB})$ | 95% CI coverage | Average Length of 95% CI |
|---|---|---|---|---|
| Scenario 1 - Small Number of Trials $k = 24$ | | | | |
| Traditional-Direct | 0.5952 | 0.7167 | 0.867 | 2.7041 |
| Traditional-Indirect | 0.5913 | 0.6312 | 0.941 | 2.5225 |
| Bayesian-HOM | 0.5796 | 0.4097 | 0.937 | 1.5704 |
| Bayesian-HET | 0.5736 | 0.4104 | 0.938 | 1.5712 |
| CD[$\mathbf{S}_{\texttt{REML}}$] | 0.5677 | 0.4057 | 0.897 | 1.3766 |
| CD[$\mathbf{S}_{\texttt{TRUE}}$] | 0.5732 | 0.3850 | 0.955 | 1.5554 |
| CD[$\mathbf{S}_{\texttt{DL}}$] | 0.5718 | 0.4195 | 0.862 | 1.2550 |
| CD[$\mathbf{S}_{\texttt{BHOM}}$] | 0.5719 | 0.3925 | 0.940 | 1.5337 |
| CD[$\mathbf{S}_{\texttt{BHET}}$] | 0.5714 | 0.3927 | 0.943 | 1.5225 |
| Scenario 2 - Large Number of Trials $k = 96$ | | | | |
| Traditional-Direct | 0.5843 | 0.3658 | 0.927 | 1.3950 |
| Traditional-Indirect | 0.6104 | 0.3118 | 0.962 | 1.2681 |
| Bayesian-HOM | 0.6126 | 0.2016 | 0.948 | 0.7663 |
| Bayesian-HET | 0.6126 | 0.2016 | 0.943 | 0.7701 |
| CD[$\mathbf{S}_{\texttt{REML}}$] | 0.5780 | 0.1915 | 0.936 | 0.7242 |
| CD[$\mathbf{S}_{\texttt{TRUE}}$] | 0.5856 | 0.1900 | 0.966 | 0.7777 |
| CD[$\mathbf{S}_{\texttt{DL}}$] | 0.5762 | 0.1932 | 0.904 | 0.6536 |
| CD[$\mathbf{S}_{\texttt{BHOM}}$] | 0.5852 | 0.1920 | 0.959 | 0.7716 |
| CD[$\mathbf{S}_{\texttt{BHET}}$] | 0.5852 | 0.1918 | 0.954 | 0.7680 |

From the results in Tables 2.6, 2.7 and 2.8, it is evident that the traditional pairwise meta-analysis is much less efficient than the CD network meta-analysis approaches. Specifically, compared with the results from the CD[$\mathbf{S}_{\texttt{REML}}$] method, the lengths of 95% CIs obtained from traditional meta-analysis methods are much greater, even though the probabilities of covering the true value are comparable. This suggests that, when the parameter of interest is a vector, information on one parameter may be potentially useful for inferences on other parameters. Thus, mixed treatment comparisons should be considered in our settings.

Table 2.7: Summary of results of simulation studies - Case 2

| Method | $\widehat{\delta}_{AB}$ | s.d.$(\widehat{\delta}_{AB})$ | 95% CI coverage | Average Length of 95% CI |
|---|---|---|---|---|
| Scenario 1 - Small Number of Trials $k = 24$ | | | | |
| Traditional-Direct | 0.6176 | 1.4759 | 0.849 | 5.4220 |
| Traditional-Indirect | 0.5905 | 0.8818 | 0.937 | 3.4753 |
| Bayesian-HOM | 0.6095 | 0.7450 | 0.887 | 2.4177 |
| Bayesian-HET | 0.5706 | 0.7360 | 0.913 | 2.6355 |
| CD[$\mathbf{S}_{\texttt{REML}}$] | 0.5793 | 0.6922 | 0.916 | 2.5426 |
| CD[$\mathbf{S}_{\texttt{TRUE}}$] | 0.5820 | 0.6865 | 0.973 | 2.9649 |
| CD[$\mathbf{S}_{\texttt{DL}}$] | 0.6165 | 0.7289 | 0.811 | 2.0011 |
| CD[$\mathbf{S}_{\texttt{BHOM}}$] | 0.6323 | 0.7030 | 0.901 | 2.3815 |
| CD[$\mathbf{S}_{\texttt{BHET}}$] | 0.6044 | 0.6930 | 0.906 | 2.4856 |
| Scenario 2 - Large Number of Trials $k = 96$ | | | | |
| Traditional-Direct | 0.6433 | 0.7431 | 0.924 | 2.8474 |
| Traditional-Indirect | 0.6287 | 0.4279 | 0.951 | 1.7643 |
| Bayesian-HOM | 0.6852 | 0.3540 | 0.899 | 1.1858 |
| Bayesian-HET | 0.6454 | 0.3436 | 0.960 | 1.3952 |
| CD[$\mathbf{S}_{\texttt{REML}}$] | 0.6200 | 0.3226 | 0.959 | 1.3164 |
| CD[$\mathbf{S}_{\texttt{TRUE}}$] | 0.6261 | 0.3254 | 0.980 | 1.4823 |
| CD[$\mathbf{S}_{\texttt{DL}}$] | 0.6455 | 0.3227 | 0.864 | 0.9721 |
| CD[$\mathbf{S}_{\texttt{BHOM}}$] | 0.6636 | 0.3324 | 0.933 | 1.2085 |
| CD[$\mathbf{S}_{\texttt{BHET}}$] | 0.6279 | 0.3256 | 0.968 | 1.3876 |

Table 2.8: Summary of results of simulation studies - Case 3

| Method | $\widehat{\delta}_{AB}$ | s.d.$(\widehat{\delta}_{AB})$ | 95% CI coverage | Average Length of 95% CI |
|---|---|---|---|---|
| Scenario 1 - Small Number of Trials $k = 24$ | | | | |
| Traditional-Direct | 0.4706 | 0.8260 | 0.868 | 3.0721 |
| Traditional-Indirect | 0.4250 | 0.4582 | 0.915 | 1.8193 |
| Bayesian-HOM | 0.4135 | 0.4400 | 0.855 | 1.4116 |
| Bayesian-HET | 0.4065 | 0.4388 | 0.853 | 1.4186 |
| CD[$\mathbf{S}_{\texttt{REML}}$] | 0.4834 | 0.4201 | 0.892 | 1.4924 |
| CD[$\mathbf{S}_{\texttt{TRUE}}$] | 0.5010 | 0.4058 | 0.953 | 1.7241 |
| CD[$\mathbf{S}_{\texttt{DL}}$] | 0.3957 | 0.4510 | 0.787 | 1.2756 |
| CD[$\mathbf{S}_{\texttt{BHOM}}$] | 0.3750 | 0.4169 | 0.855 | 1.3811 |
| CD[$\mathbf{S}_{\texttt{BHET}}$] | 0.3753 | 0.4141 | 0.852 | 1.3824 |
| Scenario 2 - Large Number of Trials $k = 96$ | | | | |
| Traditional-Direct | 0.4823 | 0.4132 | 0.912 | 1.5936 |
| Traditional-Indirect | 0.4472 | 0.2250 | 0.896 | 0.9051 |
| Bayesian-HOM | 0.4603 | 0.2131 | 0.807 | 0.6828 |
| Bayesian-HET | 0.4589 | 0.2097 | 0.822 | 0.6996 |
| CD[$\mathbf{S}_{\texttt{REML}}$] | 0.5057 | 0.1943 | 0.919 | 0.7724 |
| CD[$\mathbf{S}_{\texttt{TRUE}}$] | 0.5261 | 0.1978 | 0.949 | 0.8620 |
| CD[$\mathbf{S}_{\texttt{DL}}$] | 0.4435 | 0.2029 | 0.749 | 0.6242 |
| CD[$\mathbf{S}_{\texttt{BHOM}}$] | 0.3959 | 0.2027 | 0.754 | 0.6954 |
| CD[$\mathbf{S}_{\texttt{BHET}}$] | 0.3950 | 0.2002 | 0.759 | 0.7042 |

Consider the probability that the nominal 95% CI covers the true $\delta_{AB}$ as one criterion for assessing the performance of each meta-analysis method. It is evident from the simulation study that the results of the Bayesian methods are sensitive to the specifications of their prior distributions. Specifically, Bayesian-HOM fails to achieve appropriate coverage in Cases 2 and 3 (e.g., 89% and 90% in Table 2.7 and 86% and 81% in Table 2.8), regardless whether the number of studies is small or large. Similarly, Bayesian-HET fails to provide satisfactory coverage in the Case 3 (85% and 82% in Table 2.8) when its assumption on prior cannot cover the true model. In summary, both Bayesian methods are able to estimate $\delta_{AB}$ properly only if their prior assumptions cover the underlying true covariance model, and they fail to do so when their prior assumptions are not compatible with the underlying true covariance model. So the Bayesian procedures are vulnerable to their assumptions on priors, and we should make as few assumptions as possible when specifying priors.

In examining the results of the CD procedures, we first observe that CD[$\mathbf{S}_{\texttt{TRUE}}$] achieves desirable coverage rates in all cases ($95\% - 98\%$ in Tables 2.6, 2.7, and 2.8). Therefore, the performance of the CD procedure is satisfactory for combining information on $\boldsymbol{\theta}$. However, the performance of the CD procedure is strongly affected by the quality of estimating the covariance matrix $\mathbf{S}$. To help establish a practical guideline, we compare the quality of estimates based on the extended DL method $\mathbf{S}_{\texttt{DL}}$ and the REML method $\mathbf{S}_{\texttt{REML}}$. Specifically, we plug in the corresponding estimates in the process of constructing and combining individual CDs, and again we study the performance of estimates $\widehat{\delta}_{AB}$ and the corresponding 95% CIs. The performance of CD[$\mathbf{S}_{\texttt{REML}}$] is reasonable in all settings, i.e., close to the nominal 95% coverage (see, e.g., $92\% - 96\%$ in Tables 2.6, 2.7, and 2.8) as long as the number of studies is sufficiently large. Further, the coverage rate of CD[$\mathbf{S}_{\texttt{REML}}$] improves from $89\% - 92\%$ to $92\% - 96\%$ as the number of studies increases from 24 to 96. On the other hand, the coverage rate of CD[$\mathbf{S}_{\texttt{DL}}$] is relatively low, around $79\% - 86\%$, when the

sample size is small. Moreover, the performance of CD[$\mathbf{S}_{\texttt{DL}}$] does not always improve as the number of studies increases. For example, the coverage rate of CD[$\mathbf{S}_{\texttt{DL}}$] drops from 78.7% to 74.9% in Table 2.8. Thus, the REML method is preferable to the extended DL method for estimating the covariance matrix $\mathbf{S}$. This observation is consistent with the shortcomings of the DL method reported in univariate random-effects models by Emerson et al. (1993). Between the REML and DL methods, we recommend the CD procedure with $\mathbf{S}_{\texttt{REML}}$ for network meta-analysis when $\mathbf{S}$ is unknown.

Finally, the results for the semi-Bayesian CD procedures appear to be similar to the results for the corresponding Bayesian procedures. Specifically, the performance of CD[$\mathbf{S}_{\texttt{BHOM}}$] is in line with Bayesian-HOM. It achieves appropriate coverage in Case 1 (94% and 96% in Table 2.6), but fails in Cases 2 and 3 (90% and 93% in Table 2.7 and 86% and 75% in Table 2.8), regardless of the number of studies $k = 24$ or 96. Similarly, the results for CD[$\mathbf{S}_{\texttt{BHET}}$] are in line with Bayesian-HET. It provides satisfactory coverage in Cases 1 and 2 (94% and 95% in Table 2.6 and 91% and 97% in Table 2.7), but fails Case 3 (85% and 76% in Table 2.8). Once again, the CD procedure is sensitive to the quality of estimation of $\mathbf{S}$. Also, the confidence distribution $H^{(c)}(\cdot)$ in (2.8)is an asymptotic CD that is more suitable for making inferences on $\boldsymbol{\theta}$ when $k \to \infty$, under which both the mean vector $\boldsymbol{\theta}$ and the between-trials covariance matrix $\mathbf{S}$ can be estimated consistently.

### 2.4.3 A CD Approach with Adaptive Weights

As we observed from in Section 4.2, the overall findings for a network can be quite unreliable when indirect evidence and direct evidence inconsistent. In this section, an adaptive weighting system improves resistance to the impact of inconsistent indirect comparisons by down-weighting the trials that contribute to the inconsistent evidence. Here, the degree

of inconsistency from an indirect comparison is measured by how the trials in the indirect comparison deviate from the overall outcome for the direct comparison. The precise formulation of this measure, which we loosely call "distance," is given after Model (2.19). Taking into account this distance, the CD combining process can still use indirect comparisons that provide outcomes consistent with those from the direct comparisons, but it can also reduce the impact of inconsistent indirect comparisons. We demonstrate this property through the following simulation studies.

We consider the model (2.18) used in Scenario 1 in Section 4.1, with two modifications. First, we increase the total number of trials from 24 to 33 so that three trials, instead of one trial, compare treatments A, B, and C, and ten trials, instead of three trials, directly compare treatments A and B. We still have ten trials comparing treatments A and C and ten trials comparing treatments B and C. Thus, for inferences on $\delta_{AB}$, we have 13 direct comparisons and 20 trials with information on the indirect comparison. Second, the trials containing information on the direct comparison are consistent, but some of the remaining 20 trials containing information on the indirect comparison may be biased. Specifically, we consider the following model to generate the simulation data:

$$r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), \quad p_{ij} = \frac{\exp(\theta_{ij})}{1+\exp(\theta_{ij})}, \quad i = 1, 2, \ldots, 33, \quad j \in T_i$$
$$\boldsymbol{\theta}_i \sim (1-\epsilon)N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}) + \epsilon N(\mathbf{A}_i(\boldsymbol{\theta}-\boldsymbol{\eta}_i), \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}})$$

where

$$\epsilon = 0 \text{ and } \boldsymbol{\eta}_i = \mathbf{0} \text{ for } i \text{ s.t. } T_i = \{A, B, C\} \text{ or } T_i = \{A, B\}$$
$$\epsilon = 0.4 \text{ and } \boldsymbol{\eta}_i = \begin{cases} (\eta_{A,i}, 0, 0)^{\mathbf{T}} & \text{for } i \text{ s.t. } T_i = \{A, C\} \\ (0, \eta_{B,i}, 0)^{\mathbf{T}} & \text{for } i \text{ s.t. } T_i = \{B, C\} \end{cases} \tag{2.19}$$

Here, the values of $\eta_{A,i}$ and $\eta_{B,i}$ are fixed numbers simulated from $N(2, 4)$.

Model (2.19) indicates that all trials that compare both treatments A and B directly have the same underlying true parameter $\boldsymbol{\theta}$, whereas some trials involving A only or B only may have different underlying true parameters. If we are to include the trials that provide the indirect comparison in our analysis, it would be desirable to exclude or down-weight those trials. In this case, we devise the following notion of distance $d_i$,

$$d_i = \begin{cases} \dfrac{(\widehat{\delta}_{AC,i} - \text{median}_{l\ s.t.\ T_l=\{B,C\}} \widehat{\delta}_{BC,l}) - \widehat{\delta}_{AB,direct}}{\sqrt{\text{var}(\widehat{\delta}_{AB,direct})}} & \text{for } i \text{ s.t. } T_i = \{A, C\} \\ \dfrac{(\text{median}_{l\ s.t.\ T_l=\{A,C\}} \widehat{\delta}_{AC,l} - \widehat{\delta}_{BC,i}) - \widehat{\delta}_{AB,direct}}{\sqrt{\text{var}(\widehat{\delta}_{AB,direct})}} & \text{for } i \text{ s.t. } T_i = \{B, C\}, \end{cases}$$

where $\widehat{\delta}_{AB,direct}$ and $\text{var}(\widehat{\delta}_{AB,direct})$ are obtained from Equation (2.13). Heuristically, $d_i$ for each indirect comparison trial measures its deviation from the overall outcome given by all direct comparison trials. For example, we could consider including only the studies with distance $|d_i| \leq 1$ in the meta-analysis. In other words, we would define $w_i^*$ as

$$w_i^* = \begin{cases} 1 & \text{if } |d_i| \leq 1 \\ 0 & \text{if } |d_i| > 1, \end{cases}$$

and use $w_i^*$ in the method CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted. Specifically, we set $W_i = w_i^* \times \mathbf{A}_i^+(\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \mathbf{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})^{-1} \mathbf{A}_i$, and take the cdf of the random vector in (2.7) as the combined multivariate normal CD. We show that in this way the combined CD is able to exclude those inconsistent indirect trials – trials with large $d_i$. There are many other choices of adaptive weights. For convenience, we use here the simple, though somewhat restrictive, $|d_i| \leq 1$ to remove inconsistent studies from combination. A detailed discussion of choices of adaptive weights and their applications to combining CDs can be found in Xie et al. (2011).

In a further simulation study (Case 4), we consider two settings. In Setting 1, we generate

the simulated data using model (2.18), in which all studies have the same underlying true parameter value, but modify it to have 33 trials with the same composition of trials as model (2.19). In Setting 2, the simulated data are generated from model (2.19). In this case, some trials used in the indirect comparison have a different underlying true parameter value. In both settings, three trials compare all three treatments, ten trials compare treatments A and B, ten trials A and C, and ten trials B and C. The number of patients involved in each arm of each study is 100. We apply CD[$\mathbf{S}_{\texttt{REML}}$], CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted, and CD[$\mathbf{S}_{\texttt{TRUE}}$] to the simulated data sets. We repeat the entire process 1000 times and report the results in Table 2.9.

Table 2.9: Summary of results of simulation studies - Case 4

| Method | $\widehat{\delta}_{AB}$ | s.d.($\widehat{\delta}_{AB}$) | 95% CI coverage | Average Length of 95% CI |
|---|---|---|---|---|
| Setting 1 - 33 Trials without Inconsistent Indirect Trials | | | | |
| CD[$\mathbf{S}_{\texttt{REML}}$] | 0.5733 | 0.2984 | 0.9200 | 1.1122 |
| CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted | 0.5780 | 0.3705 | 0.9230 | 1.4078 |
| CD[$\mathbf{S}_{\texttt{TRUE}}$] | 0.5818 | 0.2955 | 0.9520 | 1.2139 |
| Setting 2 - 33Trials with Inconsistent Indirect Trials | | | | |
| CD[$\mathbf{S}_{\texttt{REML}}$] | 1.1425 | 0.3932 | 0.7190 | 1.4808 |
| CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted | 0.6479 | 0.3934 | 0.9770 | 1.9963 |
| CD[$\mathbf{S}_{\texttt{TRUE}}$] | 1.1001 | 0.3367 | 0.6260 | 1.2250 |

All three methods are able to achieve appropriate coverage rate ($92\% - 95\%$ in Setting 1) if all trial outcomes are consistent with one another. However, in Setting 2, with inconsistent indirect trials, only CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted provides appropriate inference on $\delta_{AB}$. In particular, the estimate $\widehat{\delta}_{AB} = 0.6479$ by CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted is not far from the true $\delta_{AB} = 0.6070$, and its 95% CI has a coverage rate of 97.7%. Therefore, with carefully designed study-specific weights, the CD procedure is able to provide some resistance to the impact of inconsistent indirect trials mistakenly included in the meta-analysis.

## 2.5 Concluding Remarks

In this paper, we have proposed a frequentist method for network meta-analysis by combining multivariate normal confidence distributions (CDs) associated with individual studies. This proposed CD approach can perform indirect comparisons in a network of mixed treatment comparisons, and it can use the findings from indirect comparisons efficiently to enhance the overall inference of the entire network. The CD approach can also be modified by using an adaptive weighting scheme to reduce the effect of indirect comparisons whose findings contradict those from the direct comparisons. Overall, the proposed CD approach can effectively and efficiently integrate direct and indirect information from disparate sources. In fact, the CD approach can estimate consistently and efficiently the parameters of interest as well as the between-trials covariance matrix when the number of studies goes to infinity. Through simulation studies, we have also demonstrated that the CD approach generally outperforms traditional pairwise meta-analysis and the Bayesian hierarchical model. In conclusion, the CD approach is highly competitive for network meta-analysis.

In comparing the approaches on the CAD data in Section 3.1, we excluded the TAXUS I trial to avoid addressing the issue of zero events there. In traditional pairwise meta-analysis, one customarily adds 0.5 to zero events. This correction is arbitrary and introduces bias in the inferences. By removing zero-event trials from the analysis, one would lose the information they contain. For example, for TAXUS I, zero event is a favorable outcome for both BMS and PES. This loss can cause concerns as well, especially if the zero-event trials constitute a sizable portion of the data. For an exact inference method involving zero events, the approach of combining significance functions proposed in Liu et al. (2013) can avoid the shortcomings of the earlier approaches.

In network meta-analysis, it is important to assess the consistency of the evidence from all trials in the network. However, such assessment is often difficult. One reason is that designs often differ between the trials yielding direct comparisons and the trials leading to indirect comparisons. Furthermore, it is practically impossible to distinguish between inconsistency and heterogeneity of random effects. See Higgins et al. (2002, 2003) for further discussion of this topic.

Although our examples involve clinical trials in medical studies, we emphasize that the proposed CD approach can be applied broadly for any multiple comparison studies in many other domains. For example, to establish ratings for a list of restaurants based on a survey of customer ratings, customers would be able to provide data only on the restaurants that they have patronized. The CD approach could be applied by constructing and combining CDs based on the ratings given to those restaurants by a group of customers.

## 2.6 Appendix

**Lemma 1** Suppose $W_i, i = 1, \ldots, k$ are $p \times p$ positive semi-definite symmetric matrices and $V_i$ is the column space of $W_i$. Let $V = V_1 + V_2 + \cdots + V_k \stackrel{\triangle}{=} \{\sum_{i=1}^{k} v_i | v_i \in V_i, i = 1, \ldots, k\}$. Then $\sum_{i=1}^{k} W_i$ is positive definite provided that $V = \Re^p$.

**Proof of Lemma 1:**

It is a direct result that $\sum_{i=1}^{k} W_i$ is positive semi-definite. Suppose there exists a $p \times 1$ vector $\mathbf{v} \neq 0$ such that $\mathbf{v^T}(\sum_{i=1}^{k} W_i)\mathbf{v} = 0$. Then, for any fixed $i$, we have $\mathbf{v^T} W_i \mathbf{v} = 0$, which implies that $W_i^{1/2}\mathbf{v} = 0$. It follows that $\mathbf{v} \in \text{kernel}(W_i^{1/2})$, and immediately $\mathbf{v} \in \text{kernel}(W_i)$ since $W_i$ is symmetric. Thus $\mathbf{v} \perp V_i$. Since $i$ is arbitrary, we conclude that $\mathbf{v} \perp V = \Re^p$ and $\mathbf{v}$ has to be 0, which contradicts the assumption that $\mathbf{v} \neq 0$.

**Proof of Theorem 2.1:**

Let $\boldsymbol{\xi}^{(c)} = (\sum_{i=1}^{k} W_i)^{-1} \sum_{i=1}^{k} W_i \mathbf{A}_i^{+} \boldsymbol{\xi}_i$ and $H^{(c)}(\mathbf{t}) = \Pr\{\boldsymbol{\xi}^{(c)} \leq \mathbf{t} | \mathbf{Y}_1, \ldots, \mathbf{Y}_k\}$. We need to show that $H^{(c)}(\cdot) = H(\mathbf{Y}_1, \ldots, \mathbf{Y}_k; \cdot)$ is a multivariate normal CD for $\boldsymbol{\theta}$. Define $H_{\boldsymbol{\lambda}}(t) = \Pr\{\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\xi}^{(c)} \leq t | \mathbf{Y}_1, \ldots, \mathbf{Y}_k\}$ for any given vector $\boldsymbol{\lambda}$ satisfying $||\boldsymbol{\lambda}||_2 = 1$. By Definition 2.2, it suffices to show that $H_{\boldsymbol{\lambda}}(t)$ is a univariate normal CD function for $\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\theta}$.

To do so, we first note that $H_{\boldsymbol{\lambda}}(t)$ goes from 0 to 1 monotonically as $t$ goes from $-\infty$ to $\infty$. Thus, $H_{\boldsymbol{\lambda}}(t)$ is a cdf. Second, we note that $\boldsymbol{\xi}_i$, defined by $\boldsymbol{\xi}_i | \boldsymbol{Y}_i = \boldsymbol{y}_i \sim N(\boldsymbol{y}_i, \mathrm{var}(\boldsymbol{Y}_i))$, is a CD random vector for $\boldsymbol{\theta}_i$, and furthermore, $\mathbf{A}_i^{+} \boldsymbol{\xi}_i$ is a CD random vector for $\boldsymbol{\theta}$ in the sense that the distribution function of $\boldsymbol{\eta}^{\mathbf{T}} \mathbf{A}_i^{+} \boldsymbol{\xi}_i$ is a CD for $\boldsymbol{\eta}^{\mathbf{T}} \boldsymbol{\theta}$ for any $\boldsymbol{\eta} \in V_i$. Since $(\sum_{i=1}^{k} W_i)^{-1}$ exists by Lemma 1, we consider the conditional distribution of $\left(W_i (\sum_{i=1}^{k} W_i)^{-1} \boldsymbol{\lambda}\right)^{\mathbf{T}} \mathbf{A}_i^{+} \boldsymbol{\xi}_i$ given $\mathbf{Y}_i$. Clearly, it is a univariate normal CD for $\left(W_i (\sum_{i=1}^{k} W_i)^{-1} \boldsymbol{\lambda}\right)^{\mathbf{T}} \boldsymbol{\theta}$, because $W_i (\sum_{i=1}^{k} W_i)^{-1} \boldsymbol{\lambda} \in V_i$. Therefore, it is straightforward to show that, at the true parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\Pr\{H_{\boldsymbol{\lambda}}(\mathbf{Y}_1, \ldots, \mathbf{Y}_k) \leq s\} = \Pr\left\{ \Phi\left( \frac{\sum_{i=1}^{k} \left(W_i (\sum_{i=1}^{k} W_i)^{-1} \boldsymbol{\lambda}\right)^{\mathbf{T}} \mathbf{A}_i^{+} \mathbf{Y}_i - \boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\theta}_0}{\sqrt{\sum_{i=1}^{k} \sigma_i^2}} \right) \leq s \right\} = s$$

where $\sigma_i^2 = \mathrm{var}\left(\left(W_i (\sum_{i=1}^{k} W_i)^{-1} \boldsymbol{\lambda}\right)^{\mathbf{T}} \mathbf{A}_i^{+} \boldsymbol{\xi}_i\right)$. Thus, we have established that, at the true $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and as a function of the sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_k$, $H_{\boldsymbol{\lambda}}(\mathbf{Y}_1, \ldots, \mathbf{Y}_k)$ follows the uniform distribution $U[0, 1]$. This completes the proof.

# Chapter 3

# Combining Multivariate Normal Confidence Distributions and its Application to Multivariate Meta-Analysis

In this chapter, we propose a general method for combining multivariate normal confidence distributions. The proposed methodology can support the development of efficient and robust multivariate meta-analysis approaches. The robust approaches can *effectively mitigate the undue impact of potential outlying studies* and thus address the second problem mentioned in the introduction chapter.

## 3.1 Introduction

Meta-analysis is a statistical method that aims to combine information from different sources. It is widely applied in support of making decisions in many fields, including education, marketing, medical research and etc., e.g., see the 281 references in Sutton and Higgins (2008)'s review of recent developments of meta-analysis. Most meta-analysis methods are developed for combining inference of a single parameter. As a result, such methods do not use the information of correlations between the parameters, if there is any. Multivariate meta-analysis instead simultaneously combines inferences of all parameters. It can often strength the inference by borrowing information from other related parameters. Along with the advanced technology, collection of high dimensional data is much easier

than before, and the request for jointly collecting outcomes of multiple endpoints evolves as a common practice. As the sequel, the demand for simple and effective multivariate meta-analysis method has never been greater.

Multivariate meta-analysis method can be traced back to Hedges and Olkin (1985). It is recently reviewed by van Houwelingen et al. (2002) and Arends (2006). Although it has been advocated for almost 30 years from it came into being, the method is still limitedly used in the realm of practical evidence synthesis (Riley, 2009). The lack of awareness of the benefit brought by multivariate meta-analysis, and the lack of effective approaches obstruct its widely use (Jackson et al., 2011). One particular challenge is that the crucial assumption – all studies involved in the meta-analysis must have the same underlying parameters or hyper-parameters, must be satisfied. A violation of such an assumption would lead to biased or even invalid combined estimates. This issue is aggravated in the multivariate case in the sense that if one parameter in one study has different underlying parameter, then the combined estimates for all parameters might be impacted. However, diagnosis of such violations is difficult, especially in the multivariate case. Therefore, it is desirable to have a robust approach that can tolerate such violations to some extent.

This paper proposes a general method for multivariate meta-analysis. This method is broad in the sense that it incorporates the model-based multivariate meta-analysis approach as a special case. Such an approach provides asymptotically efficient estimator under the standard multivariate fixed-effect or random-effects models (van Houwelingen et al., 2002; Arends, 2006). The proposed method can also support developing new robust multivariate meta-analysis approaches. These robust approaches can provide combined estimators that are resistant to the impact of the potential outlying studies, and as a result, can provide a means of protection against model misspecification. To our knowledge, no existing multivariate meta-analysis approach, including Bayesian methods, can make consistent inference

on the target parameters, in the existence of unknown outlying studies.

The first robust approach applies to meta-analysis of a large number of studies. It relies on asymptotic normality, and has an inherent connection to an M-estimation approach. The combined estimator maintains high relative efficiency, e.g., it attains $\sqrt{3/\pi} \approx 97.72\%$ efficiency asymptotically in both fixed-effect and random-effects models when there are no outlying studies. The second approach applies to meta-analysis of a set of large studies. It uses data-dependent adaptive weights to down-weight or exclude studies contained little or misleading information about the parameter of interest. The combined estimator has an oracle property under the fixed-effect model, and as a result, it is asymptotically efficient as the individual studies sample sizes goes to infinity.

Another challenge in the multivariate meta-analysis is that often only the summary statistics of each endpoints are available but not their correlations. Thus, it is worth knowing that the proposed method can include studies with misspecified correlations, and provide consistent estimator with minor conditions (Liu, 2012). An important subject of multivariate meta-analysis is network meta-analysis, which lays particular emphasis on improving the inference of one parameter by borrowing information from other related parameters. The recent development on network meta-analysis is controversial on attributing the disagreement between the direct and indirect evidence to heterogeneity or inconsistency, cf., the discussion in Jansen et al. (2011); Hoaglin et al. (2011); Li et al. (2011) and the references therein. The proposed method can adapt to apply to network meta-analysis, by incorporating studies with information about only a subset of target parameters. In particular, the robust approaches can adaptively down-weight studies provided potentially inconsistent information in the combining process, and thus can avoid to make decisions on the inclusion or exclusion of such studies in the evidence collecting process. Such decisions are often ambiguous due to the lack of information. Therefore, the robust approaches

have an advantage in network meta-analysis.

The proposed method is developed by combining multivariate normal confidence distributions (CDs). Loosely speaking, a CD uses a sample-dependent distribution function on the parameter space that can represent all level confidence intervals for making inference on the parameter of interest. The concept of CD can be rooted back to Bayes (1763) and Fisher (1930), and gains renewed interest in recent years (Schweder and Hjort, 2002; Singh et al., 2005, 2007). A comprehensive review of recent developments on CDs, including a renewal of definition from pure frequentist viewpoint, methods of constructing CDs, and using CDs for making inference, can be found in Xie and Singh (2013) and the followed discussions. As a distribution estimator, a CD inherently contains more information, such as skewness, than a point or interval estimator (Xie and Singh, 2013; Cox, 2013). Therefore, it is useful for making inference based on one study, and it is also an ideal tool for combining the information from different studies.

The application of using CD for combining information from independent sources is proposed by Singh et al. (2005). This CD combining method is shown to be useful for meta-analysis. For the single parameter case, a CD combining method, which unifies almost all univariate meta-analysis approaches under the same framework, and develops new robust meta-analysis approaches, is proposed by Xie et al. (2011). The proposed method in this paper is in essential a generalization of that CD combining method to the multivariate case. However, this generalization is not a simple extension, since the property used in the univariate combining framework, mapping $\mathbf{U}[0,1] \to \Re^p$ by a single monotonic function when $p = 1$, is not meaningful when $p \geq 2$. To circumstance such difficulty, a method of combining CD random vectors, rather than combining CD functions, is proposed by Yang et al. (2013a). Although such a method can provide efficient multivariate meta-analysis

approach under the standard fixed-effect or random-effects model, it is not ready for developing robust multivariate meta-analysis approaches. Thus, we present a method that directly combines multivariate normal CD functions as in the univariate case. It indeed generalizes the CD combining method proposed by Xie et al. (2011) to multivariate case. As a result, all derived meta-analysis approaches, including the robust approaches, can be correspondingly developed under the new framework. It also includes the method of combining CD random vectors as a special case.

This article proceeds as follows. Section 2 reviews and explores the concept of multivariate normal CD, and then proposes a general method that combines multivariate normal CDs. Section 3 discusses on the application of the general combining method to multivariate meta-analysis. It shows the properties of different approaches based on combining joint CDs, e.g., efficiency and robustness. It indicates that the newly developed robust multivariate meta-analysis methods can provide consistent estimator for the target parameters, whereas the conventional frequentist and Bayesian methods fail, when outlying studies are inadvertently included in the meta-analysis. Furthermore, the proposed general combining method can incorporate studies with missing endpoints, which are often the case in the network meta-analysis scenarios. Section 4 presents some numerical results and Section 5 concludes with a summary and some further discussions.

## 3.2 Multivariate Normal CDs and their Combinations

### 3.2.1 Multivariate Normal CD

A CD uses a sample-dependent distribution function to estimate the unknown parameter. It is a distribution estimator like the Bayesian posterior, however, it is a pure frequentist concept without any Bayesian reasoning involved. In principle, any sample dependent

distribution function, which can provide all levels of confidence intervals, can be used as a CD. An attractive aspect of using such a distribution estimator is that it contains wealthy information for making almost all types of inference, including point estimates, confidence interval and $p$-values for hypothesis testing problems. In fact, such inferences can be drawn from a CD in a same way as they are drawn from a Bayesian posterior distribution. The CD can subsume a broad range of frequentist concepts. For example, it has been shown that normalized likelihood function, $p$-value function, and bootstrap distribution can all be viewed as confidence distributions (cf., Xie and Singh, 2013, and reference therein).

More formally, a univariate CD is defined as following. For a single parameter $\theta \in \Theta$, suppose $\mathcal{Y}$ is the sample space corresponding to data $\boldsymbol{Y} = \{y_1, \ldots, y_n\}$, then a confidence distribution (CD) is defined as a function $H(\cdot) \equiv H(\boldsymbol{Y}, \cdot)$ on $\mathcal{Y} \times \Theta \to [0, 1]$ such that: i) $H(\cdot)$ is a cumulative distribution function on $\Theta$ for each given $\mathbf{Y} \in \mathcal{Y}$; and ii) at the true parameter value $\theta = \theta_0$, $H(\theta_0) = H(\boldsymbol{Y}, \theta_0)$, as a function of the sample $\boldsymbol{Y}$, follows the uniform distribution $\mathbf{U}[0, 1]$ (Schweder and Hjort, 2002; Singh et al., 2005). The second requirement imposes restrictions to facilitate desirable frequentist properties such as unbiasedness, consistency and/or efficiency. The function $H(\cdot)$ is an asymptotic CD (aCD) if the $\mathbf{U}[0, 1]$ requirement holds only asymptotically. A concomitant concept is *CD random variable*. Intuitively, a CD random variable is a random variable $\xi$ defined on $\mathcal{Y} \times \Theta$ such that it has the distribution $H(\cdot)$ given the sample $\boldsymbol{Y}$, where $H(\cdot)$ is the CD for $\theta$. Given any CD function $H(\cdot)$, one can construct a CD random variable associated with $H(\cdot)$. On the other hand, given any CD random variable $\xi \in \mathcal{Y} \times \Theta$ whose conditional distribution is $H(\cdot)$ (conditional on the sample $\boldsymbol{Y}$), then $H(\cdot)$ is a CD for $\theta$, the parameter of interest (Singh et al., 2007). A CD is useful for making inference. For example, let $H(\cdot) = H(\cdot, \mathbf{Y})$ denote a CD, obtained based on the sample $\mathbf{Y}$, for the parameter $\theta$. Then, the mean/median/mode of the distribution estimator $H(\cdot)$ provides point estimators for

$\theta$, the interval $(-\infty, H^{-1}(\alpha))$ is a $100(1-\alpha)\%$ level one-sided confidence interval (CI) for $\theta$, and the tail mass $H(b)$ provides a $p$-value for the one-sided hypothesis test $K_0 : \theta \leq b$ versus $K_1 : \theta > b$, for any given $b$ (Xie and Singh, 2013).

Extend the CD concept to multivariate case is not straightforward, since the property $F(\boldsymbol{y}) \overset{L}{=} \mathbf{U}[0,1]$ when $\boldsymbol{Y} \overset{L}{=} F$ does not hold in $\Re^p$ for $p \geq 2$ (Singh et al., 2007). The development of joint CD function, as a result, either concentrates the interest within center-outwards confidence regions (rather than all Borel sets) in the parameter space, or makes use of asymptotic normality (De Blasi and Schweder, 2012; Xie and Singh, 2013). In this paper, we consider the multivariate normal CD, which assumes asymptotic normality, during our development of the general combining method. It is sufficient to cover all our later discussions in the sense that almost every CD considered for combining in practice is asymptotically normal (Hannig and Xie, 2012). The definition of multivariate normal CD relies on *CD random variable*, or *CD random vector* in the multivariate case. Loosely speaking, for a $p \times 1$ vector parameter $\boldsymbol{\theta}$, a distribution function $H(\cdot)$ is a multivariate normal CD for $\boldsymbol{\theta}$ if and only if the projected distribution of $H(\cdot)$ on any direction $\boldsymbol{\lambda} \in \Re^p$ is a univariate normal CD for $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$. More formally, the following definition is proposed and utilized in Yang et al. (2013a).

**Definition 3.1** *Let $\boldsymbol{\xi}$ be a random vector on $\Re^p$. We denote by $H(\cdot)$ the conditional distribution of $\boldsymbol{\xi}$ given $\boldsymbol{Y}$. For any given $p \times 1$ vector $\boldsymbol{\lambda}$, we also denote by $H_{\boldsymbol{\lambda}}(\cdot)$ the conditional distribution of $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\xi}$ given $\boldsymbol{Y}$. Then we call $H(\cdot)$ the multivariate normal CD (or, asymptotic multivariate normal CD) for a $p \times 1$ parameter vector $\boldsymbol{\theta}$ if and only if, for any given $\boldsymbol{\lambda}$, $H_{\boldsymbol{\lambda}}(\cdot)$ is a univariate normal CD (or asymptotic CD) function for $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$. Also, the random vector $\boldsymbol{\xi}$ is called a CD random vector for $\boldsymbol{\theta}$.*

**Example 3.1** *(Multivariate Normal Mean) Suppose $\boldsymbol{y}_i, i = 1, \ldots, n$ are identically and*

*independently distributed observations from a multivariate normal distribution* $N(\boldsymbol{\theta}, \mathbf{S})$. *Then, the function*

$$H_{(n)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\mathbf{S}_{(n)}^{-1/2}(\boldsymbol{\theta} - \bar{\boldsymbol{y}}_{(n)})) \tag{3.1}$$

*satisfies requirements in Definition 3.1 and thus is a multivariate normal CD function for* $\boldsymbol{\theta}$. *Here,* $\bar{\boldsymbol{y}}_{(n)}$ *is the sample mean,* $\mathbf{S}_{(n)}$=$\mathbf{S}/n$, *and* $\boldsymbol{\Psi}(\cdot)$ *is the cumulative distribution function of the standard* $p \times 1$ *multivariate normal distribution function. If* $\mathbf{S}$ *is unknown but can be estimated consistently, say by* $\widehat{\mathbf{S}}$, *then the sample-dependent distribution (3.1) with* $\mathbf{S}_{(n)} = \widehat{\mathbf{S}}/n$ *is an asymptotic multivariate normal CD function for* $\boldsymbol{\theta}$. *Further, the random vector* $\boldsymbol{\xi}_{(n)}$ *such that* $\boldsymbol{\xi}_{(n)}|\boldsymbol{y}_i, i = 1, \ldots, n \sim N(\bar{\boldsymbol{y}}_{(n)}, \mathbf{S}_{(n)})$ *is the CD random vector for* $\boldsymbol{\theta}$ *associated with the CD function* $H_{(n)}(\boldsymbol{\theta})$.

The multivariate normal CD $H_{(n)}(\boldsymbol{\theta})$ can be used to make inference for any linear combination of $\boldsymbol{\theta}$, say $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$, for any given $\boldsymbol{\lambda} \in \Re^p$. Define $H_{(n),\boldsymbol{\lambda}}(t) = \Phi((\boldsymbol{\lambda}^{\mathbf{T}}\mathbf{S}_{(n)}\boldsymbol{\lambda})^{-1/2}(t - \boldsymbol{\lambda}^{\mathbf{T}}\bar{\boldsymbol{y}}_{(n)}))$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard univariate normal distribution function. Based on the Definition 3.1, $H_{(n),\boldsymbol{\lambda}}(t)$ is a univariate normal CD, and can be used to make inference, for the parameter $t = \boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$.

The multivariate normal CD function can also be used to make joint inference for the parameter vector $\boldsymbol{\theta}$. Specifically, we have the following theorem regarding the properties of point estimator, confidence region and hypothesis testing based on the multivariate normal CD.

**Theorem 3.1** *Let* $H_{(n)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\mathbf{S}_{(n)}^{-1/2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{(n)}))$ *be a sample-dependent multivariate normal CD function for* $\boldsymbol{\theta} \in \Re^p$. *Then,*

*(i) Point estimator:*

*If* $\sigma^2_{min}$, *the smallest eigenvalue of* $\mathbf{S}_{(n)}$, *such that* $\sigma^2_{min} \propto 1/n \to 0$ *as the sample size*

*$n \to \infty$, then the centerpoint $\widehat{\boldsymbol{\theta}}_{(n)}$ is a consistent estimator for $\boldsymbol{\theta}$.*

*(ii) Hypothesis testing:*

*For the test $K_0 : \boldsymbol{\theta}_0 \in B$ versus $K_1 : \boldsymbol{\theta}_0 \in B^c$, where $\boldsymbol{\theta}_0$ is the true parameter value, $B = \{\boldsymbol{\theta} : \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{(n)} \in \mathcal{B}\}$ and $\mathcal{B}$ is the Borel set in $\Re^p$, asymptotically $\Pr\{\boldsymbol{\theta}_0 \in B^c\} = \int_{B^c} dH_{(n)}(\boldsymbol{\theta}) = H_{(n)}(B^c)$ is the corresponding p-value of the test.*

*(iii) Confidence region:*

*The set $\{\boldsymbol{\theta} : (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{(n)})^{\mathbf{T}} \mathbf{S}_{(n)}^{-1} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{(n)}) \leq q_{1-\alpha}(\chi_p^2)\}$ provides a level $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$, where $q_{1-\alpha}(\chi_p^2)$ denotes $(1-\alpha)$-th quantile of the chi-square distribution with $p$ degree of freedom. More generally, for any given non-singular non-random $p_A \times p$ matrix $\mathbf{A}$, the set $\{\boldsymbol{\eta} : (\boldsymbol{\eta} - \mathbf{A}\widehat{\boldsymbol{\theta}}_{(n)})^{\mathbf{T}} (\mathbf{A}\mathbf{S}_{(n)}\mathbf{A}^{\mathbf{T}})^{-1} (\boldsymbol{\eta} - \mathbf{A}\widehat{\boldsymbol{\theta}}_{(n)}) \leq q_{1-\alpha}(\chi_{p_A}^2)\}$ provides a level $100(1-\alpha)\%$ confidence region for $\mathbf{A}\boldsymbol{\theta}$.*

A proof of Theorem 3.1 can be found in the Appendix. It provides several means of using $H_{(n)}(\boldsymbol{\theta})$ to make inference for $\boldsymbol{\theta}$. However, in the context of meta-analysis, we need to combine the information for $\boldsymbol{\theta}$ from different sources, e.g. combine $H_{(n_1),1}(\boldsymbol{\theta}), \ldots, H_{(n_k),k}(\boldsymbol{\theta})$, before making any inference. The proposed general CD combining method uses an equivalent expression of a multivariate normal CD, as a set of $p$ univariate normal CDs .

### 3.2.2 Decomposition of Multivariate Normal CD

The Cramér-Wold theorem (Cramér and Wold, 1936) allows a multivariate normal CD to be decomposed as an equivalent set of $p$ univariate normal CDs, without losing any information. Specifically, given $H(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\mathbf{S}^{-1/2}(\boldsymbol{\theta} - \boldsymbol{y}))$, for any basis $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p]$,

define

$$\mathbf{H}^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) = \begin{bmatrix} H_1(\boldsymbol{\lambda}_1^{\mathbf{T}}\boldsymbol{\theta}) \\ \vdots \\ H_p(\boldsymbol{\lambda}_p^{\mathbf{T}}\boldsymbol{\theta}) \end{bmatrix}, \tag{3.2}$$

where $H_j(t) = \Phi((\boldsymbol{\lambda}_j^{\mathbf{T}}\mathbf{S}\boldsymbol{\lambda}_j)^{-1/2}(t - \boldsymbol{\lambda}_j^{\mathbf{T}}\boldsymbol{y}))$ is the projected univariate normal CD for parameter $\boldsymbol{\lambda}_j^{\mathbf{T}}\boldsymbol{\theta}$ based on $H(\boldsymbol{\theta})$. If taking $\boldsymbol{\Lambda}$ such that $\boldsymbol{\Lambda} = \mathbf{D}\mathbf{S}^{-1/2}$ for some diagonal matrix $\mathbf{D}$, then

$$H(\boldsymbol{\theta}) = H_1(\boldsymbol{\lambda}_1^{\mathbf{T}}\boldsymbol{\theta}) \cdot H_2(\boldsymbol{\lambda}_2^{\mathbf{T}}\boldsymbol{\theta}) \cdots H_p(\boldsymbol{\lambda}_p^{\mathbf{T}}\boldsymbol{\theta}) = \det(\mathrm{diag}(\mathbf{H}^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}))). \tag{3.3}$$

Thus, $\mathbf{H}^{\boldsymbol{\Lambda}}(\boldsymbol{\theta})$ reserves all information in $H(\boldsymbol{\theta})$. Further, consider CD random variables $\xi_j|\boldsymbol{Y} \sim H_j(\cdot)$, then the conditional distribution (conditional on sample data $\boldsymbol{Y}$) of $\xi_{j_1}$ and $\xi_{j_2}$ are independent since $\boldsymbol{\Lambda}\mathbf{S}\boldsymbol{\Lambda}^{\mathbf{T}} = \mathbf{D}^2$.

Thus, we conclude:

**Lemma 3.1** *Let* $H(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\mathbf{S}^{-1/2}(\boldsymbol{\theta} - \boldsymbol{y}))$ *be a multivariate normal CD functions for* $\boldsymbol{\theta} \in \Re^p$. *Take* $\boldsymbol{\Lambda} = \mathbf{S}^{-1/2}$, *denote* $\boldsymbol{\Lambda} \overset{\triangle}{=} [\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p]^{\mathbf{T}}$ *and consider the vector* $\mathbf{H}^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) = \begin{bmatrix} H_1(\boldsymbol{\lambda}_1^{\mathbf{T}}\boldsymbol{\theta}) \\ \vdots \\ H_p(\boldsymbol{\lambda}_p^{\mathbf{T}}\boldsymbol{\theta}) \end{bmatrix}$, *then: i)* $H(\boldsymbol{\theta}) = \det(\mathrm{diag}(\mathbf{H}^{\boldsymbol{\Lambda}}(\boldsymbol{\theta})))$*; and ii)* $H_j(\cdot)$ *are orthogonal to each other, in equivalent, the CD random variable* $\xi_j|\boldsymbol{Y} = \boldsymbol{y} \sim H_j(\cdot)$ *are independent to each other.*

Essentially, Lemma 3.1 indicates that the multivariate CD, $H(\boldsymbol{\theta})$, and the set of univariate CDs, $\mathbf{H}^{\boldsymbol{\Lambda}}(\boldsymbol{\theta})$, can be derived from one another without any other information.

### 3.2.3 A General Method for Combining Multivariate Normal CDs

The use of CD for combining information from independent sources is proposed by Singh et al. (2005), in light of the wealth information CD contained. In the univariate case, suppose we have $k$ independent CDs $H_i(\theta), i = 1, \ldots, k$ regarding to the same single parameter $\theta$, then the function

$$H^{(c)}(\theta) = G^{(c)}\{g^{(c)}(H_1(\theta), \ldots, H_k(\theta))\}, \tag{3.4}$$

is a combined CD function for $\theta$, where $G^{(c)}(t) = \Pr\{g^{(c)}(U_1, \ldots, U_k) \leq t\}$, and $U_1, \ldots, U_k$ are independent $U[0, 1]$ random variables.

Specifically, Xie et al. (2011) showed that a special family of $g^{(c)}(\cdot)$ such that

$$g^{(c)}(u_1, \ldots, u_k) = w_1 a_0(u_1) + \cdots + w_k a_0(u_k), \tag{3.5}$$

can unify all commonly used univariate meta-analysis methods under the same framework (3.4). Here, $a_0(\cdot)$ is any given monotonic function and $w_i \geq 0$, with at least one $w_i \neq 0$, are generic weights for the combination. Methods unified under the equation (3.5) includes the $p$-value combination methods (cf., Marden, 1991), model-based meta-analysis (cf., Table IV of Normand, 1999), the Mantel-Haenszel method, Peto's method, and the method proposed by Tian et al. (2009) which combines confidence intervals.

To extend the combining method to multivariate case, the difficulty resides in applying $a(\cdot)$ to $H_i(\boldsymbol{\theta})$ to get a vector for combination. A direct application of $a(\cdot)$ to $H_i(\boldsymbol{\theta})$ would give a set of vectors, instead of a particular vector, when the parameter $\boldsymbol{\theta}$ is a vector. And therefore, it would be meaningless for combination. The proposed method circumvents this difficulty using the results in Section 2.2, by expressing a $p$-dimensional multivariate CD as $p$ univariate CDs, and applying $a(\cdot)$ to such $p$ univariate CDs individually. Specifically, denote

$\mathbf{a}_0[\cdot]$ as element-wise applying of $a_0(\cdot)$ to a vector, e.g., $\mathbf{a}_0[\mathbf{H}_i^{\mathbf{\Lambda}}(\boldsymbol{\theta})] = \begin{bmatrix} a_0(H_{i1}(\boldsymbol{\lambda}_{i1}^{\mathbf{T}}\boldsymbol{\theta})) \\ \vdots \\ a_0(H_{ip}(\boldsymbol{\lambda}_{ip}^{\mathbf{T}}\boldsymbol{\theta})) \end{bmatrix}$.

Then,

**Theorem 3.2** *Let $H_i(\boldsymbol{\theta}) = \mathbf{\Psi}(\mathbf{S}_i^{-1/2}(\boldsymbol{\theta} - \boldsymbol{y}_i)), i = 1, \ldots, k$ be the multivariate normal CD functions for the same multivariate parameter $\boldsymbol{\theta}$ from $k$ independent studies. For any monotonic function $a_0(\cdot)$ with first derivative, denote by*

$$H^{(c)}(\boldsymbol{\theta}) = G^{(c)}\left\{\sum_{i=1}^{k} \mathbf{W}_i^{1/2}\mathbf{a}_0[\mathbf{H}_i^{\mathbf{\Lambda}}(\boldsymbol{\theta})]\right\}, \tag{3.6}$$

*then $H^{(c)}(\boldsymbol{\theta})$ is a (asymptotic) multivariate normal CD for $\boldsymbol{\theta}$ given the generic weights matrices $\mathbf{W}_i$ are $p \times p$ positive-definite. Here, $\mathbf{\Lambda}_i = \mathbf{S}_i^{-1/2}$ and $G^{(c)}(\cdot)$ is the cumulative distribution function of $\sum_{i=1}^{k} \mathbf{W}_i^{1/2}\mathbf{a}_0[\mathbf{U}_i]$, where $\mathbf{U}_i = [U_{i1}, \ldots, U_{ip}]^{\mathbf{T}}$, and $U_{ij}$ are independent $\mathbf{U}[0, 1]$ random variables.*

A proof the Theorem 3.2 can be found in the Appendix. The proposed combining method (3.6) reduces to (3.4) with $g^{(c)}(\cdot)$ in (3.5) when $p = 1$. It also covers the method of combining CD random vectors proposed by Yang et al. (2013a), by taking $a_0(\cdot) = \Phi^{-1}(\cdot)$. It can yield new robust multivariate meta-analysis approaches by taking $a_0(t) = t$ or using adaptive weights, which are difficult to derive through combining CD random vectors. In the next section, we develop different multivariate meta-analysis approaches with distinct properties, by choosing different $a(\cdot)$ functions when applying the general method in Theorem 3.2 to different models.

## 3.3 Multivariate Meta-analysis by Combining CDs

In this section, we show that our proposed general method for combining multivariate CDs can yield a variety of new methods for multivariate meta-analysis. Specifically, we show in Section 3.1 an efficient combining method for standard multivariate random-effects model. We further develop in Section 3.2 robust combining methods for situations where the study population is contaminated. We present in Section 3.3 an extension of the combining method that can incorporate studies with missing endpoints.

### 3.3.1 Efficient Combination Method

We consider the multivariate random-effects model (3.7) (van Houwelingen et al., 2002; Arends, 2006), which can be viewed as a natural extension of the univariate random-effects model considered in Normand (1999). Suppose

$$\text{Model 3.1:} \qquad \boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}_i, \mathbf{S}_i), \quad \boldsymbol{\theta}_i|\boldsymbol{\theta}, \boldsymbol{\Sigma} \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}, \boldsymbol{\Sigma}), \quad i = 1, 2, \ldots, k \tag{3.7}$$

where $\boldsymbol{y}_i$ is the summary statistic from the $i$-th study, $\mathbf{S}_i$ is the covariance matrix of $\boldsymbol{y}_i$, $\boldsymbol{\theta}_i$ is the study-specific mean (random-effects), and $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are hyper-parameters for $\boldsymbol{\theta}_i$. This model also covers the fixed-effect model, which is equivalent to take $\boldsymbol{\Sigma} \equiv \mathbf{O}$.

Then, under above setting, we can show that $H_i(\boldsymbol{\theta}) = \boldsymbol{\Psi}((\mathbf{S}_i + \widehat{\boldsymbol{\Sigma}})^{-1/2}(\boldsymbol{\theta} - \boldsymbol{y}_i))$ is an aCD for $\boldsymbol{\theta}$, where $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimate of $\boldsymbol{\Sigma}$. Apply Theorem 3.2 to Model 3.1, and take $a_0(\cdot) = \Phi^{-1}(\cdot)$ and $\mathbf{W}_i = (\mathbf{S}_i + \widehat{\boldsymbol{\Sigma}})^{-1}$ in (3.6), then

$$\text{Method 3.1:} \qquad H_{\mathtt{E}}^{(c)}(\boldsymbol{\theta}) \;=\; \boldsymbol{\Psi}\left(\mathbf{S}_{\mathtt{c},\mathtt{E}}^{-1/2}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\mathtt{E}}^{(c)}\right)\right) \tag{3.8}$$

is a combined CD for $\boldsymbol{\theta}$, where $\widehat{\boldsymbol{\theta}}_{\mathtt{E}}^{(c)} = (\sum_{i=1}^{k} \mathbf{W}_i)^{-1} \sum_{i=1}^{k} \mathbf{W}_i \boldsymbol{y}_i$ can be used as a combined point estimator, with variance $\mathbf{S}_{\mathtt{c,E}} = (\sum_{i=1}^{k} \mathbf{W}_i)^{-1}$. This combined CD $H_{\mathtt{E}}^{(c)}(\cdot)$ is equivalent to the one given in Yang et al. (2013a), which uses CD random vector as the combining vehicle.

The traditional multivariate meta-analysis combines point estimators $\boldsymbol{y}_i$ from individual studies, and uses $\widehat{\boldsymbol{\theta}}_{\mathtt{pt}} = \left(\sum_{i=1}^{k} (\mathbf{S}_i + \widehat{\boldsymbol{\Sigma}}_{\mathtt{REML}})^{-1}\right)^{-1} \left(\sum_{i=1}^{k} (\mathbf{S}_i + \widehat{\boldsymbol{\Sigma}}_{\mathtt{REML}})^{-1} \boldsymbol{y}_i\right)$ as a combined point estimator for $\boldsymbol{\theta}$, where $\widehat{\boldsymbol{\Sigma}}_{\mathtt{REML}}$ is the restricted-maximum-likelihood (REML) estimator for $\boldsymbol{\Sigma}$ (Jennrich and Schluchter, 1986; Jackson et al., 2011). This $\widehat{\boldsymbol{\theta}}_{\mathtt{pt}}$ is the maximum likelihood estimator (MLE) for $\boldsymbol{\theta}$, which is asymptotically normal with mean $\boldsymbol{\theta}$ and covariance matrix $\mathbf{S}_{\mathtt{pt}} = \mathrm{var}(\widehat{\boldsymbol{\theta}}_{\mathtt{pt}}) = \left(\sum_{i=1}^{k} (\mathbf{S}_i + \widehat{\boldsymbol{\Sigma}}_{\mathtt{REML}})^{-1}\right)^{-1}$. It is often referred as the most efficient point estimator since its variance $\mathbf{S}_{\mathtt{pt}}$ achieves the Cramér-Rao lower bound when $\boldsymbol{\Sigma}$ known. Evidently, $\widehat{\boldsymbol{\theta}}_{\mathtt{E}}^{(c)}$ is equivalent to $\widehat{\boldsymbol{\theta}}_{\mathtt{pt}}$ by replacing $\widehat{\boldsymbol{\Sigma}}$ with $\widehat{\boldsymbol{\Sigma}}_{\mathtt{REML}}$. Thus, the combined CD $H_{\mathtt{E}}^{(c)}(\cdot)$ can lead to efficient inference.

### 3.3.2 Robust Combination Methods

The proposed general combining method in Theorem 3.2 can also support the development of robust meta-analysis methods. One great concern in meta-analysis is that the studies involved might not have the same underlying true parameters or hyper-parameters value. This concern is more severe in multivariate case because even only one study has one different parameter can affect the estimation of all parameters. Thus, there is a great need to develop robust multivariate meta-analysis methods that can provide protections on failing to satisfy such an assumption. The proposed robust method achieved such robustness by limiting the impact of outlying studies. Specifically, Section 3.3.2.1 considers the setting of combining a large number of studies, e.g., $k$ goes to infinity, and Section 3.3.2.2

considers the setting large studies, e.g., $n_i$ goes to infinity.

#### 3.3.2.1 Robust Meta-Analysis of a Large Number of Studies

As an extension of Model 3.1, we borrow ideas from Huber (1964) and consider the following contaminated random-effects model:

$$\text{Model 3.2:} \qquad \begin{array}{c} \boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}_i, \mathbf{S}_i), \\ \boldsymbol{\theta}_i|\boldsymbol{\theta}, \boldsymbol{\Sigma} \overset{\text{ind}}{\sim} (1-\epsilon)N(\boldsymbol{\theta}, \boldsymbol{\Sigma}) + \epsilon D_\epsilon(\boldsymbol{\theta}), \quad i = 1, \ldots, k \end{array} \tag{3.9}$$

where $\epsilon < 1/2$ and $D_\epsilon(\boldsymbol{\theta})$ is any contaminating distribution.

In above model, the study-specific means $\boldsymbol{\theta}_i$ come from a mixed model, where the majority comes from a normal distribution, and the rest comes from a contaminating distribution $D_\epsilon(\boldsymbol{\theta})$. Let $\boldsymbol{\theta}_0$ be the population mean of the uncontaminated population $N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, and $\boldsymbol{\theta}_*$ be the population mean of the contaminated population $(1-\epsilon)N(\boldsymbol{\theta}, \boldsymbol{\Sigma}) + \epsilon D_\epsilon(\boldsymbol{\theta})$. Note that $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_*$ if $\epsilon = 0$, or the contaminated distribution $D_\epsilon(\boldsymbol{\theta})$ is symmetric around $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}_0$ for any $\boldsymbol{\lambda}$, e.g., $D_\epsilon(\boldsymbol{\theta}) = N(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_1)$.

Let $H_i(\boldsymbol{\theta})$ for $\boldsymbol{\theta}_*$ be the individual joint CD functions based on the sample from $i$-th study. Apply Theorem 3.2 to Model 3.2, and take $a_0(t) = t$, then

$$\text{Method 3.2a:} \qquad H_{\mathtt{R}}^{(c)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\left\{\left(\frac{1}{12}\sum_{i=1}^{k}\mathbf{W}_i\right)^{-1/2}\sum_{i=1}^{k}\mathbf{W}_i^{1/2}\left(\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) - \frac{1}{2}\mathbf{1}\right)\right\} \tag{3.10}$$

is a combined aCD for $\boldsymbol{\theta}_*$, where $\mathbf{1}$ is $p\times 1$ vector with all elements equal to 1. This combined aCD $H_{\mathtt{R}}^{(c)}(\cdot)$ relies on the fact that the summation of $k$ random vectors, whose elements are independent $\mathbf{U}[0,1]$-distributed random variables, can approximate a multivariate normal distribution well even when $k$ is quite small.

Method 3.2a provides a robust combining approach by taking $a_0(t) = t$ to limit the impact of individual CDs on the combined CD, e.g., CDs based on studies came from $D_\epsilon(\boldsymbol{\theta})$. The idea is like using sample median, instead of sample mean, as robust estimator for the center point of the population. However, it is often in practice that Model 3.2 is misspecified as Model 3.1, where the contamination component is not considered. If so and if there is indeed some outlying studies, then $H_i(\boldsymbol{\theta})$ based on the misspecified model is no longer a CD for $\boldsymbol{\theta}_*$. Instead, it is a CD-like function, a function that follows Definition 3.1 without the $\mathbf{U}[0,1]$ distribution assumption (Xie et al., 2011). Using CD-like function for combining would bring an issue to the covariance matrix estimation in $H_{\mathtt{R}}^{(c)}(\cdot)$. Therefore, we propose to consider

$$\text{Method 3.2b:} \qquad \widetilde{H}_{\mathtt{R}}^{(c)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\left\{\left(\sum_{i=1}^{k}\mathbf{W}_i^{1/2}\mathbf{D}_i\mathbf{W}_i^{1/2}\right)^{-1/2}\sum_{i=1}^{k}\mathbf{W}_i^{1/2}\left(\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) - \frac{1}{2}\mathbf{1}\right)\right\} \tag{3.11}$$

as a combined aCD for $\boldsymbol{\theta}_*$, where $\mathbf{D}_i = \text{diag}\left(\begin{bmatrix} (H_{i1}(\boldsymbol{\lambda}_{i1}^{\mathbf{T}}\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}) - 1/2)^2 \\ \vdots \\ (H_{ip}(\boldsymbol{\lambda}_{ip}^{\mathbf{T}}\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}) - 1/2)^2 \end{bmatrix}\right)$. As a result, Method 3.2b extends the robust method to combine CD-like functions, by evaluating the covariance matrix through $\mathbf{D}_i$ instead of $\frac{1}{12}\mathbf{I}$. In addition, it provides a means of protection against model misspecification.

The asymptotically efficient point estimator under the uncontaminated model (3.7), $\widehat{\boldsymbol{\theta}}_{\mathtt{E}}^{(c)}$, is a weighed average of observed sample means. Thus, it lacks robustness, because its breakdown point equals to zero in the limit. On the other hand, the centerpoint of $\widetilde{H}_{\mathtt{R}}^{(c)}(\cdot)$, $\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}$, as another point estimator for $\boldsymbol{\theta}_*$, equals to the solution of an M-estimating equations:

$$\sum_{i=1}^{k}\mathbf{W}_i^{1/2}\left(\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) - \frac{1}{2}\mathbf{1}\right) = \mathbf{0}. \tag{3.12}$$

As a result, we have the following theorem regarding the robustness of $\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}$.

**Theorem 3.3** *Let $v_{i(j)}^2$ be the $j$-th largest singular value of the generic weight matrix $\mathbf{W}_i$, then (1) As $k \to \infty$, $\widehat{\boldsymbol{\theta}}_R^{(c)}$ is a consistent estimator of $\boldsymbol{\theta}_*$.*
*(2) The breakdown point of $\widehat{\boldsymbol{\theta}}_R^{(c)}$ is*

$$\min_{1\le t\le k}\left\{t:\sum_{i=1}^{t} v_{i(1)} \ge \sum_{i=t+1}^{k} v_{i(p)}\right\}/k. \tag{3.13}$$

*Therefore, the breakdown point achieves 1/2 if the weights $\mathbf{W}_i$ are all taken as $\mathbf{W}_i = \mathbf{I}$, where $\mathbf{I}$ is the $p \times p$ identity matrix.*

A proof of the Theorem 3.3 can be found in Appendix. It indicates that Method 3.2b is asymptotically equivalent to an M-estimation method. As a result, the combined CD $\widetilde{H}_{\mathtt{R}}^{(c)}(\cdot)$ can lead to robust inference.

Method 3.2b also maintains relative high efficiency. When no contamination, Model 3.1 is the true model and $H_{\mathtt{E}}^{(c)}(\cdot)$ gives asymptotically efficient inference for $\boldsymbol{\theta}_0$. It can be shown that the length of CI based on $\widetilde{H}_{\mathtt{R}}^{(c)}(\cdot)$ is relatively $\sqrt{\pi/3} \approx 1.0233$ wider than those based on $H_{\mathtt{E}}^{(c)}(\cdot)$, for any given $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}_0$ at the same confidence level. Thus, Method 3.2b achieves a relative efficiency of $\sqrt{3/\pi} \approx 0.9772$ when there is no outlying studies, as shown in the following theorem.

**Theorem 3.4** *Let the weight matrix $\mathbf{W}_i = (\mathbf{S}_i + \widehat{\boldsymbol{\Sigma}})^{-1}$, and $\sigma_{i(j)}^2$ be the $j$-th largest singular value of $(\mathbf{S}_i + \widehat{\boldsymbol{\Sigma}})$. If $\epsilon = 0$, and $\sigma_{i(j)}^2 \propto 1/n_i \to 0$ at the same rate $n$, then, as $n \to \infty$ and $k \to \infty$, the asymptotic relative efficiency of $\widetilde{H}_R^{(c)}(\cdot)$ compared to $H_E^{(c)}(\cdot)$ is $\sqrt{3/\pi} \approx 0.9772$.*

A proof of the Theorem 3.4 can be found in Appendix.

It is important to be aware that the combined CD function $\widetilde{H}_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$ has the correct variance estimates, but not the $H_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$, under misspecified models. As a result, $\widetilde{H}_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$, rather than $H_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$, provides protection against model misspecification, when outlying studies are unawarely included in the meta-analysis. If all inputs $H_i(\boldsymbol{\theta})$ are indeed CD functions, then $\mathbf{D}_i \to \frac{1}{12}\mathbf{I}$, and $\widetilde{H}_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$ is asymptotically the same as $H_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$. On the other hand, under the contaminated model with small studies, the construction of a CD or aCD function for each study might be a difficult task. Thus, the extension of the method to combine CD-like functions is also practical useful.

#### 3.3.2.2 Robust Meta-Analysis of a Set of Large Studies

Section 3.3.2.1 develops a robust method using asymptotic normality which requires the number of studies $k$ goes to infinity. To expand our development to cover meta-analysis of limited studies, we develop in this subsection a robust method which assumes $k$ is bounded. This robust method assumes fixed-effect model with outlying studies, which can be viewed as a special case of the random-effects model specified in Model 3.2. It leads to a combined CD function for the parameter of interest, rather than an aCD function as in Method 3.2b.

Specifically, suppose $k$ studies each has sample sizes $n_1, \ldots, n_k$, and without loss of generality, we assume $n_i$ goes to infinity at the same rate $n$. Let $\boldsymbol{\theta}$ be the parameter of interest, and the underlying true value of $\boldsymbol{\theta}$ for the $i$-th study be $\boldsymbol{\theta}_i$, which may not be the same across all $k$ studies, and the studies with different values are unknown. A special example of this setting is the following fixed-effect model:

$$\text{Model 3.3:} \quad \begin{array}{c} \boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}_i, \mathbf{S}_i), \\ \mathcal{I}_0 = \{i : \boldsymbol{\theta}_i = \boldsymbol{\theta}_0\}, \quad |\mathcal{I}_0| \geq k/2, \quad i = 1, \ldots, k \end{array} \tag{3.14}$$

where $\boldsymbol{\theta}_0$ is the true parameter value of the majority of the studies for meta-analysis, and $|\mathcal{I}_0|$ is the size of the set $\mathcal{I}_0$. Model 3.3 can be viewed as a special case of Model 3.2 with $\boldsymbol{\Sigma} = \mathbf{O}$. It reduces to the conventional fixed-effect model when $\boldsymbol{\theta}_i = \boldsymbol{\theta}_0$ for all studies.

The reason we restricted the model to fixed-effect model is the development of the robust method requires $H_i(\boldsymbol{\theta})$ satisfies:

*Condition (A).* For any fixed $\gamma$, $0 < \gamma < \frac{1}{2}$, and $\boldsymbol{\lambda} \in \Re^p$, $L_{i,\boldsymbol{\lambda}} = H^{-1}_{i,\boldsymbol{\lambda}}(1-\gamma) - H^{-1}_{i,\boldsymbol{\lambda}}(\gamma) \to 0$, in probability, as $n \to \infty$.

This is a generalized version of the condition specified in Xie et al. (2011) Section 4. Intuitively, it requires $\widehat{\boldsymbol{\theta}}_i \to \boldsymbol{\theta}_i$ as $n \to \infty$, where $\widehat{\boldsymbol{\theta}}_i$ is the centerpoint of $H_i(\boldsymbol{\theta})$.

The key idea in developing the robust approach is utilizing data-dependent adaptive weights to down-weight or exclude studies contained little or misleading information about the parameter of interest, and therefore only keep studies contained correct information for combining. Specifically, we present the Method 3.3 as a two steps algorithm:

*Method 3.3 - Algorithm:*

*Step 1: For a fixed $i$, let $\widetilde{w}_{i(j)}, j = 1, \ldots, k$ be a set of adaptive weights satisfies*

$$\lim_{n\to\infty} \widetilde{w}_{i(j)} = \begin{cases} 1 & \text{if} \quad \boldsymbol{\theta}_j = \boldsymbol{\theta}_i \\ 0 & \text{if} \quad \boldsymbol{\theta}_j \neq \boldsymbol{\theta}_i \end{cases}, \tag{3.15}$$

*then combine $H_1(\boldsymbol{\theta}), \ldots, H_k(\boldsymbol{\theta})$ using Theorem 2 by taking $a_0(\cdot) = \Phi^{-1}(\cdot)$ and $W_j = \widetilde{w}_{i(j)}\mathbf{S}_j^{-1}$, and denote the combined CD by $H_i^{(c)}(\boldsymbol{\theta})$. Repeat this step for $i = 1, \ldots, k$.*

*Step 2: Let $\widehat{\boldsymbol{\theta}}_i^{(c)}$ be the centerpoint of $H_i^{(c)}(\boldsymbol{\theta})$, and define depth$(\widehat{\boldsymbol{\theta}}_i^{(c)})$ to be the depth of $\widehat{\boldsymbol{\theta}}_i^{(c)}$ with respect to the set $\mathcal{I}_c = \{\widehat{\boldsymbol{\theta}}_1^{(c)}, \ldots, \widehat{\boldsymbol{\theta}}_k^{(c)}\}$. Let $\mathcal{I}_{d,max} = \{i : depth(\widehat{\boldsymbol{\theta}}_i^{(c)}) = \max depth(\widehat{\boldsymbol{\theta}}_i^{(c)})\}$.*

*If* $|\mathcal{I}_{d,max}| = 1$*, then denote* $H^{(c)}_{i_{max}}(\boldsymbol{\theta})$ *as* $H^{(c)}_{R,adpt}(\boldsymbol{\theta})$*, where* $i_{max} = \text{argmax}\, depth(\boldsymbol{\theta}^{(c)}_i)$*. And,* $H^{(c)}_{R,adpt}(\boldsymbol{\theta})$ *is a combined CD for* $\boldsymbol{\theta}_0$*.*

*If* $|\mathcal{I}_{d,max}| > 1$*, then combine* $H^{(c)}_l(\boldsymbol{\theta})$ *for* $l \in \mathcal{I}_{d,max}$ *using Method 3.1, and the combined CD* $H^{(c)}_{R,adpt}(\boldsymbol{\theta})$ *is a combined CD for* $\boldsymbol{\theta}_0$*.*

Here, the data depth can be calculated using the Tukey's method (Rousseeuw and Ruts, 1996; Rousseeuw and Struyf, 1998) or the Liu's method (Liu, 1990; Liu et al., 1999).

Intuitively, Step 1 uses the CD-combining Method 3.1 to combine $H_1(\boldsymbol{\theta}), \ldots, H_k(\boldsymbol{\theta})$, while adjusting weight matrix by the adaptive weight $\widetilde{w}_{i(j)}$. The combined CD $H^{(c)}_i(\boldsymbol{\theta})$ is a combined CD function for $\boldsymbol{\theta}_i$, if $\widetilde{w}_{i(j)}$ satisfies (3.15). For example, the adaptive weights are often inversely related to the "distance" of $\boldsymbol{\theta}_j$ to $\boldsymbol{\theta}_i$,

$$\widetilde{w}_{i(j)} = \boldsymbol{\Psi}(\mathbf{S}_i^{-1/2}(\widehat{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\theta}}_i))/\boldsymbol{\Psi}(0). \tag{3.16}$$

And, Step 2 identifies the CD(s) corresponding to the median(s) of $\widehat{\boldsymbol{\theta}}^{(c)}_i$, combines them if not unique, and uses it as the combined CD for $\boldsymbol{\theta}_0$. Because $\boldsymbol{\theta}_0$ is the majority of $\boldsymbol{\theta}_i$, the median of $\widehat{\boldsymbol{\theta}}^{(c)}_i$ is a consistent estimator for $\boldsymbol{\theta}_0$, and as a result, the correspondent CD is a CD for $\boldsymbol{\theta}_0$.

Let $H^{(c)}_0(\boldsymbol{\theta})$ be the combined CD function, in the ideal case, using Method 3.1 and only the studies in the set of $\mathcal{I}_0$, assuming the membership of $\mathcal{I}_0$ is known. Let $\widehat{\boldsymbol{\theta}}^{(c)}_0$ denote the centerpoint of $H^{(c)}_0(\boldsymbol{\theta})$, then $\widehat{\boldsymbol{\theta}}^{(c)}_0$ is an asymptotic efficient estimator of $\boldsymbol{\theta}_0$. The following theorem shows that the centerpoint of $H^{(c)}_{\mathtt{R,adpt}}(\boldsymbol{\theta})$, $\widehat{\boldsymbol{\theta}}^{(c)}_{\mathtt{R,adpt}}$, is a consistent, robust, and also an asymptotic efficient estimator of $\boldsymbol{\theta}_0$. In fact, $\widehat{\boldsymbol{\theta}}^{(c)}_{\mathtt{R,adpt}}$ achieves an oracle property, which is a stronger result than $\widehat{\boldsymbol{\theta}}^{(c)}_{\mathtt{R}}$ based on Method 3.2b, under the Model 3.3.

**Theorem 3.5** *Under the Model 3.3, let* $H^{(c)}_{R,adpt}(\boldsymbol{\theta})$ *be the combined CD using Method 3.3,*

*and denote* $\widehat{\boldsymbol{\theta}}^{(c)}_{R,\,adpt}$ *as the centerpoint of* $H^{(c)}_{R,\,adpt}(\boldsymbol{\theta})$*, then as* $n \to \infty$*:*

*(1) The estimator* $\widehat{\boldsymbol{\theta}}^{(c)}_{R,\,adpt} \to \boldsymbol{\theta}_0$*, in probability.*

*(2) The breakdown point of the estimator* $\widehat{\boldsymbol{\theta}}^{(c)}_{R,\,adpt}$ *is* $1/2$*.*

*(3) The estimator* $\widehat{\boldsymbol{\theta}}^{(c)}_{R,\,adpt}$ *is asymptotically efficient, since* $n^{1/2}||\widehat{\boldsymbol{\theta}}^{(c)}_{R,\,adpt} - \widehat{\boldsymbol{\theta}}^{(c)}_{0}||_2 \to 0$ *as* $n \to \infty$*.*

The proof of part (1) and (2) is the same as the proof of Theorem 1 in Xie et al. (2011), and follows the fact that $|\mathcal{I}_0| \geq k/2$. The proof of part (3) is the same as the proof of Theorem 2 in Xie et al. (2011), which relies on the fact that $\widetilde{w}_{i(j)}$ goes to zero at rate $o(n^{\alpha})$ for any $\alpha < 0$ if $j \notin \mathcal{I}_i$, and $|\widetilde{w}_{i(j)} - 1| = O(k \cdot n^{-1/2}(\log n)^2)$ if $j \in \mathcal{I}_i$, where $\mathcal{I}_i = \{l : \boldsymbol{\theta}_l = \boldsymbol{\theta}_i\}$.

### 3.3.3 Extension to Incorporate Studies with Missing Endpoints

The previous developments assumes all studies involved in meta-analysis having the parameter of interest in the same dimension. This assumption might be too restrictive. Studies for meta-analysis are often undergone independently, and as a result, often have different designs and outcomes (Sutton and Higgins, 2008). Thus, some parameters might not be identifiable in some of the studies. For example, studies in network meta-analysis in general only have partial common parameters. Specifically, suppose $k$ studies are performed for investigating the effects of $p$ treatments, denoted by the vector $\boldsymbol{\theta} = (\theta_{\mathtt{T1}}, \ldots, \theta_{\mathtt{Tp}})^{\mathbf{T}}$. An individual study might investigate only a subset of such $p$ treatments. For example, the $i$-th study involves $p_i < p$ treatments, and only the $p_i$-dimensional parameter $\boldsymbol{\theta}_i = (\theta_{\mathtt{T1}}, \ldots, \theta_{\mathtt{Tp_i}})^{\mathbf{T}}$ is indefinable. Therefore, it only provides partial information about $\boldsymbol{\theta}$. Theorem 3.6 shows that the proposed general combining method can adapt to include studies where only a subset of the parameters is identifiable.

**Theorem 3.6** *Let* $H_i(\boldsymbol{\theta}_i) \equiv \boldsymbol{\Psi}(\mathbf{S}_i^{-1/2}(\boldsymbol{\theta}_i - \boldsymbol{y}_i)), i = 1, \ldots, k$ *be multivariate normal CD functions for the multivariate parameter* $\boldsymbol{\theta}_i$ *from* $k$ *independent sample* $\boldsymbol{y}_i$*, where* $\boldsymbol{\theta}_i = \mathbf{A}_i\boldsymbol{\theta}$ *for the same parameter vector* $\boldsymbol{\theta}$*. Here, the* $p_i \times p$ *selection matrix* $\mathbf{A}_i$ *is obtained by removing from the* $p \times p$ *identity matrix (or, more generally, any* $p \times p$ *orthonormal matrix* $\mathbf{A}$*) the rows that correspond to the omitted parameters. For any monotonic function* $a_0(\cdot)$*, denote by*

$$H^{(c)}(\boldsymbol{\theta}) = G^{(c)}\left\{\sum_{i=1}^{k} \mathbf{W}_i^{1/2}\mathbf{A}_i^{+}\mathbf{a}_0[\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}_i)]\right\}, \tag{3.17}$$

*then* $H^{(c)}(\boldsymbol{\theta})$ *is a (asymptotic) multivariate normal CD for* $\boldsymbol{\theta}$ *if:*

*(1) Each* $p \times p$ *matrix* $\mathbf{W}_i$ *is positive semi-definite.*

*(2) The column space* $\mathcal{C}(\mathbf{W}_i)$ *of* $\mathbf{W}_i$ *is identical to the rows space* $\mathcal{R}(\mathbf{A}_i)$ *of* $\mathbf{A}_i$*.*

*(3)* $\mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_k = \Re^p$ *where* $\mathbf{V}_i = \mathcal{R}(\mathbf{A}_i)$*, where* $\mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_k \stackrel{\triangle}{=} \{\sum_{i=1}^{k} v_i | v_i \in \mathbf{V}_i, i = 1, \ldots, k\}$*.*

*Here,* $\mathbf{A}_i^{+}$ *denotes the Moore-Penrose pseudo inverse of* $\mathbf{A}_i$*, and* $G^{(c)}(\cdot)$ *is the cumulative distribution function of* $\sum_{i=1}^{k} \mathbf{W}_i^{1/2}\mathbf{A}_i^{+}\mathbf{a}_0[\mathbf{U}_i]$*, where* $\mathbf{U}_i = [U_{i1}, \ldots, U_{ip_i}]^{\mathbf{T}}$*, and* $U_{ij}$ *are independent* $\mathbf{U}[0,1]$ *random variables.*

As a result, Model 3.1 is generalized to Model 3.4:

$$\text{Model 3.4:} \qquad \boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}_i, \mathbf{S}_i), \quad \boldsymbol{\theta}_i|\boldsymbol{\theta}, \boldsymbol{\Sigma} \overset{\text{ind}}{\sim} N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\boldsymbol{\Sigma}\mathbf{A}_i^{\mathbf{T}}), \quad i = 1, 2, \ldots, k \tag{3.18}$$

The efficient CD combining approaches developed in Method 3.1 is adapted to

$$\text{Method 3.4:} \qquad H_{\mathtt{E}}^{(c)}(\boldsymbol{\theta}) \;=\; \boldsymbol{\Psi}\left((\sum_{i=1}^{k}\mathbf{W}_i)^{1/2}\left(\boldsymbol{\theta} - (\sum_{i=1}^{k}\mathbf{W}_i)^{-1}\sum_{i=1}^{k}\mathbf{W}_i\mathbf{A}_i^{+}\boldsymbol{y}_i\right)\right) \tag{3.19}$$

And correspondingly, Model 3.2 is generalized to Model 3.5:

$$\text{Model 3.5:} \quad \begin{array}{c} \boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}_i, \mathbf{S}_i), \quad \boldsymbol{\theta}_i = \mathbf{A}_i\boldsymbol{\theta}_{\mathtt{i,unobs}}, \\ \boldsymbol{\theta}_{\mathtt{i,unobs}}|\boldsymbol{\theta}, \boldsymbol{\Sigma} \overset{\text{ind}}{\sim} (1-\epsilon)N(\boldsymbol{\theta}, \boldsymbol{\Sigma}) + \epsilon D_\epsilon(\boldsymbol{\theta}), \quad i = 1, \ldots, k \end{array} \tag{3.20}$$

The robust CD combining approaches developed in Method 3.2b is adapted to

$$\text{Method 3.5:} \quad \widetilde{H}_{\mathtt{R}}^{(c)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\left\{(\sum_{i=1}^{k}\mathbf{W}_i^{1/2}\mathbf{A}_i^{+}\mathbf{D}_i\mathbf{A}_i\mathbf{W}_i^{1/2})^{-1/2}\sum_{i=1}^{k}\mathbf{W}_i^{1/2}\mathbf{A}_i^{+}\left(\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) - \frac{1}{2}\mathbf{1}_i\right)\right\} \tag{3.21}$$

where $\mathbf{D}_i = \text{diag}\left(\begin{bmatrix} (H_{i1}(\boldsymbol{\lambda}_{i1}^{\mathbf{T}}\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}) - 1/2)^2 \\ \vdots \\ (H_{ip_i}(\boldsymbol{\lambda}_{ip_i}^{\mathbf{T}}\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}) - 1/2)^2 \end{bmatrix}\right)$, and $\mathbf{1}_i$ is $p_i \times 1$ vector with all 1.

The robustness and relative efficiency results in Theorem 3.3 and 3.4 still holds for Method 3.5.

At last, Model 3.3 is generalized to Model 3.6:

$$\text{Model 3.6:} \quad \begin{array}{c} \boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}_i, \mathbf{S}_i), \quad \boldsymbol{\theta}_i = \mathbf{A}_i\boldsymbol{\theta}_{\mathtt{i,unobs}}, \\ \mathcal{I}_0 = \{i : \boldsymbol{\theta}_{\mathtt{i,unobs}} = \boldsymbol{\theta}_0\}, |\{i : i \in \mathcal{I}_0 \text{ and } l \in T_i\}| \geq k/2, \forall l, \quad i = 1, \ldots, k \end{array} \tag{3.22}$$

where $T_i$ denotes the set of indices of the parameters observed in the $i$-th study.

The algorithm of Method 3.3 stays the same except that the adaptive weights requires

$$\lim_{n\to\infty} \widetilde{w}_{i(j)} = \begin{cases} 1 & \text{if} \quad \boldsymbol{\theta}_{\mathtt{j,unobs}} = \boldsymbol{\theta}_{\mathtt{i,unobs}} \\ 0 & \text{if} \quad \boldsymbol{\theta}_{\mathtt{j,unobs}} \neq \boldsymbol{\theta}_{\mathtt{i,unobs}} \end{cases}, \tag{3.23}$$

and a particular choice could be

$$\widetilde{w}_{i(j)} = \{\boldsymbol{\Psi}(\mathbf{S}_i^{-1/2}(\mathbf{A}_i\widehat{\boldsymbol{\theta}}_{\mathtt{j(mdn)}} - \widehat{\boldsymbol{\theta}}_i)) \times \prod_{l \notin T_i} \Phi((\widehat{\theta}_{\mathtt{j(mdn),l}} - \widehat{\theta}_{\mathtt{mdn,l}})/\sigma_{\mathtt{mdn,l}})\}/\boldsymbol{\Psi}(0). \quad (3.24)$$

where $\widehat{\theta}_{\mathtt{mdn,l}} = \text{median}\{\widehat{\theta}_{i,l}, l \in T_i\}$ is the median of the observed $l$-th parameter value, $\sigma_{\mathtt{mdn,l}}$ is the standard deviation corresponding to $\widehat{\theta}_{\mathtt{mdn,l}}$, and $\widehat{\boldsymbol{\theta}}_{\mathtt{j(mdn)}}$ is obtained by plugin $\widehat{\theta}_{\mathtt{mdn,l}}$ to $\widehat{\boldsymbol{\theta}}_{\mathtt{j}}$ if the $l$-th parameter is not observed in the $j$-th study, e.g.,

$$\widehat{\theta}_{\mathtt{j(mdn),l}} = \begin{cases} \widehat{\theta}_{\mathtt{j,l}} & \text{if} \quad l \in T_j \\ \widehat{\theta}_{\mathtt{mdn,l}} & \text{if} \quad l \notin T_j \end{cases}. \quad (3.25)$$

## 3.4 Numerical Examples

This section illustrates the proposed multivariate meta-analysis methods through three real data examples, namely, the ulcer, lidocaine, and coronary artery disease (CAD) data set. Section 3.4.1 uses the ulcer data to compare the performance of the efficient CD approach (Method 3.1), the robust CD approach (Method 3.2b) and a commonly used Bayesian method. These methods obtain results similar to the previous publications. The data is further modified to include outlying studies for meta-analysis. Only the robust CD approach can resist to the impact of such outlying studies, whereas both efficient CD approach and the Bayesian method fail when outlying studies exist. Section 3.4.2 uses the similar thought to illustrate the other robust approach (Method 3.3) using lidocaine data. Section 3.4.3 applied all above approaches to the CAD data to show the proposed meta-analysis method can incorporate studies with partial information regarding to the target parameter vector.

### 3.4.1 Example 1 - Meta-Analysis of a Large Number of Studies

The first example shows meta-analysis of a large number of studies using the ulcer data set. The data set is a collection of 41 randomized clinical trials, regarding the superiority of a novel surgical treatment for reducing recurrent bleeding on stomach ulcer patients, conducted from 1980 to 1989 (Sacks et al., 1990; Efron, 1996). Table 3.1 lists the number of patients had recurrent bleeding (an adverse event) after the treatment ($r_{ij}$) and the total number of patients had stomach ulcers undergone such treatment ($n_{ij}$) from these 41 studies. This data set was analyzed by Efron (1996) and Xie et al. (2011) using the empirical Bayes method and the univariate CD combining methods, respectively. The parameter of interest in their analysis is the log-odds ratio $\delta$ in favor of the new treatment. Here, we instead consider the vector of log odds $\boldsymbol{\theta} = (\theta_A, \theta_B)^{\mathbf{T}}$ in order to show the advantage of jointly analyzing multiple endpoints, and the log odds ratio can be obtained by $\delta = \theta_A - \theta_B$.

Given the binary type of the response for an individual patient, it often assumes $r_{ij}|p_{ij} \sim$ Binomial$(n_{ij}, p_{ij})$, where $p_{ij}$ denote the probability that a patient experiencing recurrent bleeding in the $i$-th study with $j$-th treatment, $i = 1, \ldots, 41, j = A, B$. After re-parameterizations, it is customary to consider an asymptotic equivalent model as specified in Model 3.1 with $k = 41$:

$$\begin{aligned} \boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i &\sim N(\boldsymbol{\theta}_i, \mathbf{S}_i), \\ \boldsymbol{\theta}_i|\boldsymbol{\theta}, \boldsymbol{\Sigma} &\sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}), \quad i = 1, 2, \ldots, 41 \end{aligned} \tag{3.26}$$

where $\boldsymbol{y}_i = [y_{ij}, j = A, B]^{\mathbf{T}}, \boldsymbol{\theta}_i = [\theta_{ij}, j = A, B]^{\mathbf{T}}$, and $\mathbf{S}_i = \text{diag}(\widehat{\sigma}^2_{ij}, j = A, B)$, with $y_{ij} = \log\left(r_{ij}/(n_{ij} - r_{ij})\right), \theta_{ij} = \log\left(p_{ij}/(1 - p_{ij})\right)$, and $\widehat{\sigma}^2_{ij} = 1/r_{ij} + 1/(n_{ij} - r_{ij})$. Here, we add

0.5 to the entries of zeros, and subtract 0.5 from $r_{ij}$ if $r_{ij} = n_{ij}$, in order to obtain meaningful sample estimates $\boldsymbol{y}_i$, as suggested in Efron (1993) and others. Further, we assume the within study covariance matrices $\mathbf{S}_i$ are diagonal matrices, reflecting the independence between observed outcomes from treatment groups due to randomization (van Houwelingen et al., 2002).

Therefore, our proposed efficient CD approach (CD[Efficient], Method 3.1, (3.8)) and robust CD approach (CD[Robust-M Estimation], Method 3.2b, (3.11)) can be applied to the ulcer data set straightforwardly. For comparison reason, we also consider a Bayesian meta-analysis approach, which assumes Model (3.26) with the following none informative conjugate priors on $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$.

$$\begin{aligned} \text{prior:} \qquad & \boldsymbol{\theta} \sim N(0, 10^3\mathbf{I}), \\ & \boldsymbol{\Sigma} \sim \sigma^2 \text{IW}(\mathbf{I}, 2), \quad \sigma^2 \sim \text{IG}(0.01, 0.01) \end{aligned} \tag{3.27}$$

where IW represents inverse Wishart distribution, and IG represents inverse gamma distribution.

The 2nd column of Table 3.2 presents the results of all three approaches with the point estimators of $\boldsymbol{\theta}$ and $\delta$ and associated 95% CIs. Such results show that the new surgical treatment is significantly better than the traditional treatment in reducing adverse event (recurrent bleeding) rates, which is in agree with Efron (1996) and Xie et al. (2011). Further, we notice that the 95% CIs for $\delta$ in the 2nd column of Table 3.2 are shorter than the corresponding 95% CIs in left half of Table 3 in Xie et al. (2011), where the meta-analysis are performed on the single parameter $\delta$ by combining univariate CDs $H_{\delta,i}(\delta)$. It implies a slight improvement on efficiency of estimating $\delta$ by jointly analyzing $\boldsymbol{\theta} = (\theta_A, \theta_B)^{\mathbf{T}}$.

Figure 3.1 draws the combined CD function for $\boldsymbol{\theta}$ using contour plots, and draws the

derived CD function for $\delta$ using CD density plot. It is obvious that the variability of $\widehat{\theta}_A$ is much less than $\widehat{\theta}_B$. Thus, the new treatment not only reduces the event rate but also achieves a more stable performance. Such an observation can only be discovered under the joint analysis of $\theta_A$ and $\theta_B$.

The original data set contains studies retained relative consistent information with the majorities in favor the new treatment. To demonstrate the impact of gross outlying studies, we devise a "contaminated" data set by artificially creating 6 inconsistent studies. Such studies are created by imitating mistakenly writing 1 on the ten's digit place of the observed log odds that causes the inference to be favorable on the old treatment. Specifically, we increase the original log-odds by 10 if the log-odds in the new treatment group is greater or equal to zero (study 6, 11, 33), and decrease the original log-odds by 10 if the log-odds in the old treatment group is less than $-2$ (study 3, 5, 41). These modified log odds are far out of bound since the observed log odds in original data set range from $-4.2195$ to 4.2195.

We apply the same three approaches and presents their results on the last column of Table 3.2. It is obvious that the inference based on the CD[Efficient] method and Bayesian method are significantly impact by these outlying studies, whereas the inference based on CD[Robust-M Estimation] method is not. For example, the $\widehat{\delta}$ based on CD[Efficient] move from $-0.93$ to 0.09 using the original and contaminated data set, and those based on Bayesian method move from $-0.91$ to 0.06. The superiority conclusion of the new treatment is no longer valid when applying the CD[Efficient] and Bayesian method to the contaminated data set. The estimated overall log odd ratios even flip their signs. On the other hand, the $\widehat{\delta}$ based on CD[Robust-M Estimation] slightly move from $-0.91$ to $-0.78$. Thus, the CD[Robust-M Estimation] method limits the impact of outlying studies, and holds on the superiority conclusion.

Likewise, Figure 3.2 draws the contour plots for $\boldsymbol{\theta}$ and density plot for $\delta$ based on the combined CD functions. It is clear that the combined CD based on CD[Efficient] moves significantly compared with Figure 3.1, whereas the combined CD based on CD[Robust-M Estimation] stays still.

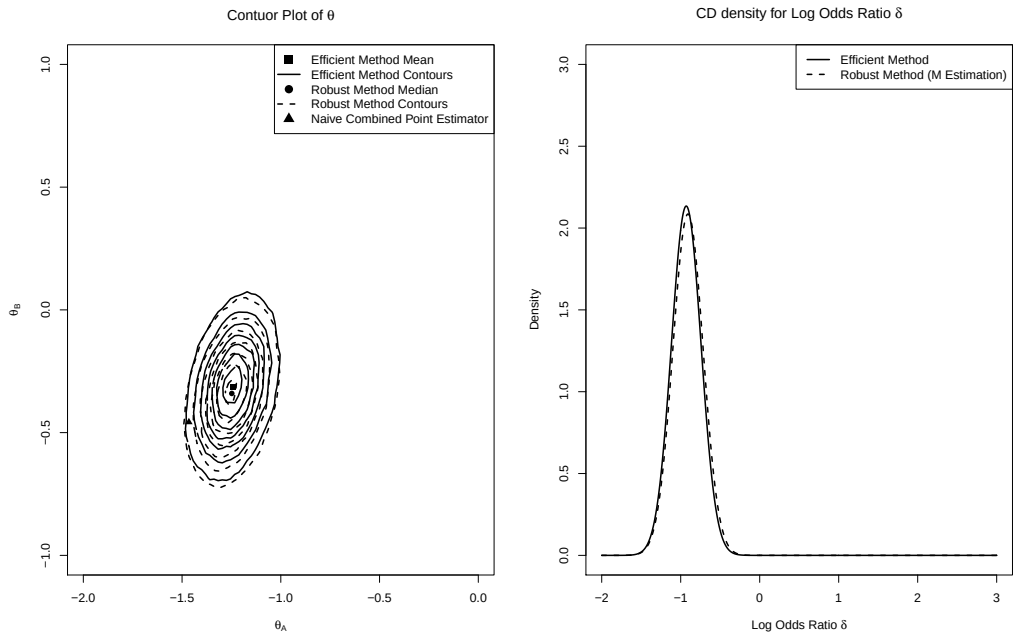


Figure 3.1: Efficient and robust multivariate meta-analysis on the original data set.

### 3.4.2 Example 2 - Meta-Analysis of a Set of Large Studies

The second example shows meta-analysis of a set of large studies using the lidocaine data set. The data set in Table 3.3 contains mortality events after a heart attack using intravenous lidocaine and a control treatment (Hine et al., 1989). It includes 6 studies with sample sizes from 83 to 300, which are relatively large. We consider a fixed-effect model where the parameter of interest are event probabilities $\boldsymbol{\theta} = (p_T, p_C)^{\mathbf{T}}$, and risk difference $\delta = p_T - p_C$ as in line with Normand (1999) and Xie et al. (2011). Here, $p_T$ and $p_C$

Table 3.1: Stomach Ulcer Data Set

| Trial Index | New Treatment (A) | | Old Treatment (B) | |
|---|---|---|---|---|
| | $r_{ij}$ | $n_{ij}$ | $r_{ij}$ | $n_{ij}$ |
| 1 | 7 | 15 | 11 | 13 |
| 2 | 8 | 19 | 8 | 16 |
| 3 | 5 | 34 | 4 | 39 |
| 4 | 7 | 36 | 4 | 31 |
| 5 | 3 | 12 | 0 | 12 |
| 6 | 4 | 7 | 4 | 4 |
| 7 | 4 | 17 | 13 | 24 |
| 8 | 1 | 16 | 13 | 16 |
| 9 | 3 | 14 | 7 | 22 |
| 10 | 2 | 38 | 12 | 32 |
| 11 | 6 | 12 | 8 | 8 |
| 12 | 2 | 7 | 7 | 9 |
| 13 | 9 | 21 | 7 | 24 |
| 14 | 7 | 21 | 5 | 25 |
| 15 | 3 | 25 | 11 | 32 |
| 16 | 4 | 11 | 6 | 10 |
| 17 | 2 | 10 | 8 | 10 |
| 18 | 1 | 31 | 4 | 27 |
| 19 | 4 | 28 | 15 | 31 |
| 20 | 7 | 43 | 16 | 43 |
| 21 | 6 | 40 | 13 | 21 |
| 22 | 4 | 18 | 5 | 39 |
| 23 | 14 | 68 | 13 | 74 |
| 24 | 6 | 21 | 8 | 21 |
| 25 | 0 | 6 | 6 | 6 |
| 26 | 1 | 10 | 5 | 15 |
| 27 | 5 | 17 | 5 | 15 |
| 28 | 0 | 10 | 12 | 14 |
| 29 | 0 | 22 | 8 | 24 |
| 30 | 2 | 18 | 10 | 21 |
| 31 | 1 | 15 | 7 | 13 |
| 32 | 8 | 24 | 15 | 27 |
| 33 | 6 | 12 | 7 | 9 |
| 34 | 0 | 20 | 5 | 23 |
| 35 | 4 | 17 | 2 | 16 |
| 36 | 10 | 40 | 12 | 20 |
| 37 | 3 | 16 | 2 | 16 |
| 38 | 4 | 34 | 5 | 19 |
| 39 | 7 | 38 | 15 | 37 |
| 40 | 0 | 34 | 34 | 34 |
| 41 | 0 | 9 | 0 | 16 |

Table 3.2: Meta-analysis of results for log-odds and log-odds-ratio of the stomach ulcers example

| | Original Data Set | Contaminated Date Set |
|---|---|---|
| CD[Efficient] | | |
| $\widehat{\theta}_A$ (95%CI) | $-1.2433(-1.4710, -1.0156)$ | $-0.5484(-0.7762, -0.3207)$ |
| $\widehat{\theta}_B$ (95%CI) | $-0.3132(-0.6788, \ 0.0523)$ | $-0.6389(-1.0045, -0.2734)$ |
| $\widehat{\delta}$ (95%CI) | $-0.9301(-1.2964, -0.5637)$ | $0.0905(-0.2759, \ 0.4568)$ |
| CD[Robust-M Estimation] | | |
| $\widehat{\theta}_A$ (95%CI) | $-1.2482(-1.4812, -1.0152)$ | $-1.1803(-1.4133, -0.9473)$ |
| $\widehat{\theta}_B$ (95%CI) | $-0.3404(-0.7145, \ 0.0337)$ | $-0.4024(-0.7765, -0.0283)$ |
| $\widehat{\delta}$ (95%CI) | $-0.9078(-1.2827, -0.5329)$ | $-0.7779(-1.1528, -0.4030)$ |
| Bayesian Model | | |
| $\widehat{\theta}_A$ (95%CI) | $-1.2406(-1.5080, -1.0000)$ | $-0.8202(-1.8560, \ 0.1999)$ |
| $\widehat{\theta}_B$ (95%CI) | $-0.3258(-0.6683, \ 0.0409)$ | $-0.8760(-1.9440, \ 0.2333)$ |
| $\widehat{\delta}$ (95%CI) | $-0.9148(-1.3100, -0.5458)$ | $0.0558(-1.3030, \ 1.3621)$ |

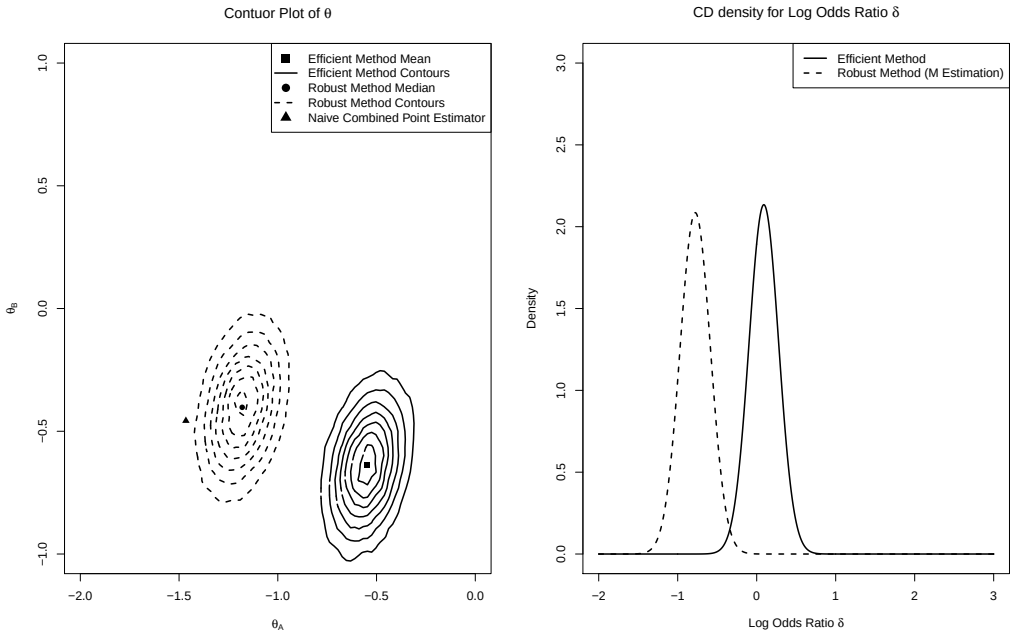


Figure 3.2: Efficient and robust multivariate meta-analysis on the contaminated data set.

denote the probability of mortality events in the lidocaine group and the control group, respectively. Specifically,

$$\boldsymbol{y}_i|\boldsymbol{\theta}, \mathbf{S}_i \sim N(\boldsymbol{\theta}, \mathbf{S}_i), \quad i = 1, \ldots, 6 \tag{3.28}$$

where $\boldsymbol{y}_i = [y_{ij}, j = T, C]^{\mathbf{T}}$, $y_{ij} = r_{ij}/n_{ij}$, and $\mathbf{S}_i = \text{diag}(\widehat{\sigma}^2_{ij}, j = T, C)$. $\widehat{\sigma}^2_{ij} = y_{ij}(1 - y_{ij})/n_{ij}$.

We apply the efficient CD approach (CD[Efficient], Method 3.1), the robust CD approach (CD[Robust-Adaptive Weights], Method 3.3), and a Bayesian approach assumes non-informative prior $N(0, 10^3\mathbf{I}_2)$ on $\boldsymbol{\theta}$. The results are presented on the 2nd column of Table 3.4. The results are close to each other, and in agree with the significance of increased mortality rate of using lidocaine. Figure 3.3 plots the $\widehat{\boldsymbol{\theta}}_i^{(c)}$ from the Step 1 of Method 3.3 in open circles, marks the median $\widehat{\boldsymbol{\theta}}^{(c)}_{\texttt{R,adpt}}$ using filled circle, and draws $\widehat{\boldsymbol{\theta}}_E^{(c)}$ using filled square. The $\widehat{\boldsymbol{\theta}}_E^{(c)}$ is instead the region confined by $\widehat{\boldsymbol{\theta}}_i^{(c)}$s. It also plots CD densities for $\delta$. The CD densities based on the efficient and robust method are very close to each other.

We further illustrate the impact of potential outliers by creating an outlying study. Suppose one erroneously uses 21, instead of 1, as the number mortality event, when calculating the point estimate of the first study. We repeat the analysis on the contaminated data set, and present the results on the 3rd column of Table 3.4. The efficient CD approach and the Bayesian approach change a lot, and the significance conclusion does no longer hold. On the other hand, the robust CD approach instead appears to change only slightly, and the significance conclusion still holds. A comparison of Figure 3.3 and Figure 3.4 shows that the $\widehat{\boldsymbol{\theta}}_E^{(c)}$ is pulled out of the place where $\widehat{\boldsymbol{\theta}}_i^{(c)}$s gathered by the outlying study. The CD densities plot shows the impact of injecting the outlying study for inference $\delta$. The robust method is slightly change, whereas the efficient method is largely change with its center

moved from positive to negative.

Table 3.3: Lidocaine Set

| Trial Index | Treatment (T) | | Control (C) | |
|---|---|---|---|---|
| | $r_{ij}$ | $n_{ij}$ | $r_{ij}$ | $n_{ij}$ |
| 1 | 2 | 39 | 1 | 43 |
| 2 | 4 | 44 | 4 | 44 |
| 3 | 6 | 107 | 4 | 110 |
| 4 | 7 | 103 | 5 | 100 |
| 5 | 7 | 110 | 3 | 106 |
| 6 | 11 | 154 | 4 | 146 |

Table 3.4: Meta-analysis of results for log-odds and log-odds-ratio of the lidocaine example

| | Original Data Set | Contaminated Date Set |
|---|---|---|
| CD[Efficient] | | |
| $\widehat{\theta}_T$ (95%CI) | $0.0652(0.0447, 0.0857)$ | $0.0652(0.0447, 0.0857)$ |
| $\widehat{\theta}_C$ (95%CI) | $0.0336(0.0186, 0.0486)$ | $0.0854(0.0703, 0.1004)$ |
| $\widehat{\delta}$ (95%CI) | $0.0316(0.0062, 0.0570)$ | $-0.0201(-0.0456, 0.0053)$ |
| CD[Robust-Adaptive Weights] | | |
| $\widehat{\theta}_T$ (95%CI) | $0.0636(0.0414, 0.0857)$ | $0.0659(0.0425, 0.0894)$ |
| $\widehat{\theta}_C$ (95%CI) | $0.0310(0.0152, 0.0469)$ | $0.0349(0.0174, 0.0524)$ |
| $\widehat{\delta}$ (95%CI) | $0.0325(0.0053, 0.0598)$ | $0.0311(0.0018, 0.0604)$ |
| Bayesian Model | | |
| $\widehat{\theta}_T$ (95%CI) | $0.0648(0.0449, 0.0848)$ | $0.0648(0.0449, 0.0848)$ |
| $\widehat{\theta}_C$ (95%CI) | $0.0337(0.0186, 0.0488)$ | $0.0855(0.0704, 0.1005)$ |
| $\widehat{\delta}$ (95%CI) | $0.0311(0.0063, 0.0556)$ | $-0.0206(-0.0455, 0.0039)$ |

### 3.4.3 Example 3 - Meta-Analysis of Studies with Missing Endpoints

The last example shows network meta-analysis using the CAD data set, which gathers 37 clinical trials that investigated the performance of three types of stent: bare-metal stents (BMS), sirolimus-eluting stents (SES) and paclitaxel-eluting stents (PES), in patients with coronary artery disease (CAD), denoted by treatment $A$, $B$ and $C$, respectively (Stettler

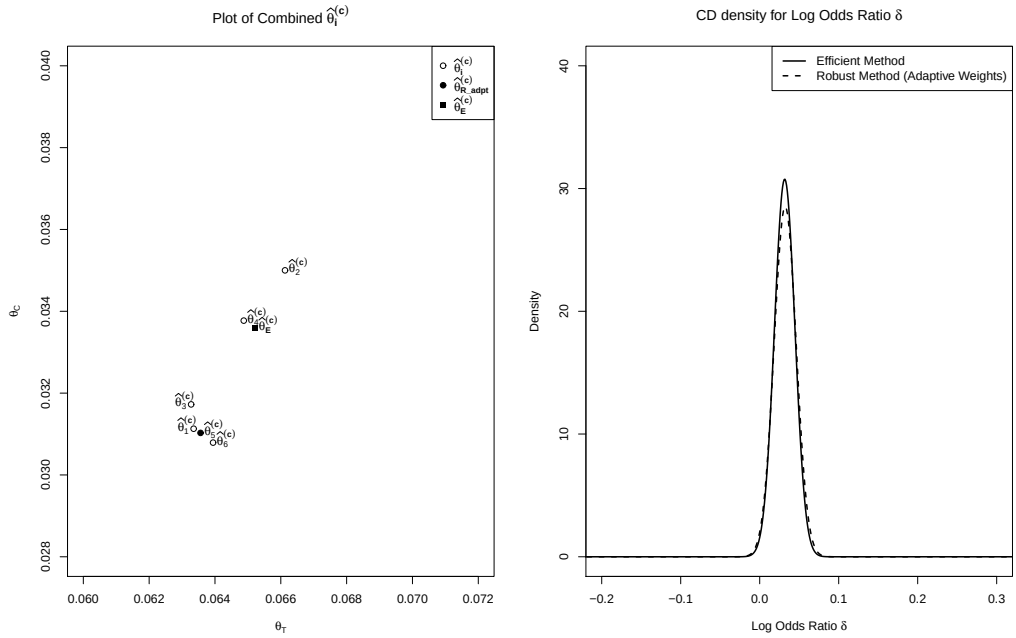


Figure 3.3: Efficient and robust multivariate meta-analysis on the original data set.

et al., 2007). Each row of Table 3.5 represents one study, which compared the rate of target lesion revascularisations (TLRs) at one year after surgeries using BMS, SES and PES. The outcomes of some stents are missing in some individual studies. The parameter of interest is the vector of log odds $\boldsymbol{\theta} = (\theta_A, \theta_B, \theta_C)^{\mathbf{T}}$ and all pairwise comparisons $\delta_{j_1 j_2} = \theta_{j_2} - \theta_{j_1}, j_1 = A, B, C, j_2 = A, B, C, j_1 \neq j_2$. This example intends to show that the proposed combining methods are adapt to incorporate studies with missing endpoints.

Let us consider the multivariate random-effects model that incorporates missing arms using Model 3.4 with $k = 37$:

$$\begin{aligned} \boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i &\sim N(\boldsymbol{\theta}_i, \mathbf{S}_i), \\ \boldsymbol{\theta}_i|\boldsymbol{\theta}, \boldsymbol{\Sigma} &\sim N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\boldsymbol{\Sigma}\mathbf{A}_i^{\mathbf{T}}), \quad i = 1, 2, \ldots, 37. \end{aligned} \tag{3.29}$$

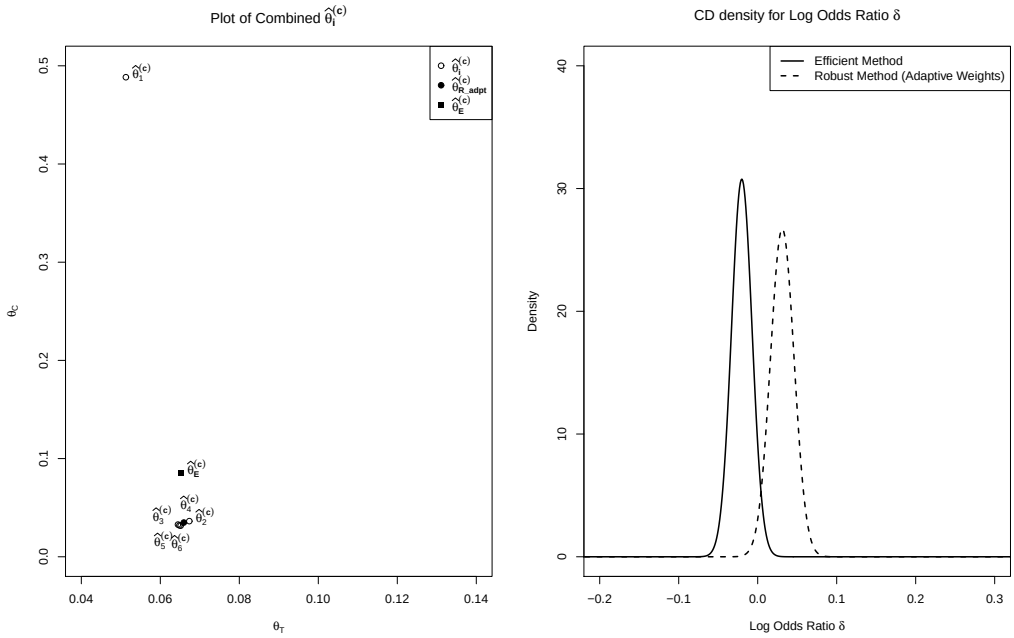


Figure 3.4: Efficient and robust multivariate meta-analysis on the contaminated data set.

where $\boldsymbol{y}_i = [y_{ij}, j \in T_i]^{\mathbf{T}}, \boldsymbol{\theta}_i = [\theta_{ij}, j \in T_i]^{\mathbf{T}}$, and $\mathbf{S}_i = \text{diag}(\widehat{\sigma}_{ij}^2, j \in T_i)$, with $y_{ij} = \log(r_{ij}/(n_{ij} - r_{ij}))$, $\theta_{ij} = \log(p_{ij}/(1 - p_{ij}))$, and $\widehat{\sigma}_{ij}^2 = 1/r_{ij} + 1/(n_{ij} - r_{ij})$. Here, $\mathbf{A}_i$ is the selection matrix associated with $T_i$, the set of treatments involved in the $i$-th study. For example, $T_i = \{A, C\}$ for study TAXUS-I, and $\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ if $T_i = \{A, B\}$, $\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ if $T_i = \{A, C\}$, $\mathbf{A}_i = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ if $T_i = \{B, C\}$, and $\mathbf{A}_i = \mathbf{I}_3$ if $T_i = \{A, B, C\}$.

The proposed efficient CD approach (CD[Efficient], Method 3.4), and robust CD approaches (CD[Robust-M Estimation], Method 3.5, and CD[Robust-Adaptive Weights], Method 3.6) can incorporate studies with missing arms and apply to the above model. It

is worth to note that the practical performance of Method 3.6 is good under the random-effect model, though the conclusions in Theorem 3.5 only holds under the fixed-effect model. For comparison reason, we also consider the following Bayesian approach assumed none informative conjugate prior for both $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$:

$$\begin{aligned} &\boldsymbol{y}_i|\boldsymbol{\theta}_i, \mathbf{S}_i \sim N(\boldsymbol{\theta}_i, \mathbf{S}_i), \quad \boldsymbol{\theta}_i = \mathbf{A}_i\boldsymbol{\theta}_{\mathtt{i,unobs}}, \\ &\boldsymbol{\theta}_{\mathtt{i,unobs}}|\boldsymbol{\theta}, \boldsymbol{\Sigma} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}), \quad i = 1, \ldots, 37, \\ \text{prior:} \quad &\\ &\boldsymbol{\theta} \sim N(0, 10^3\mathbf{I}), \\ &\boldsymbol{\Sigma} \sim \sigma^2 \mathrm{IW}(\mathbf{I}, 3), \quad \sigma^2 \sim \mathrm{IG}(0.01, 0.01) \end{aligned} \tag{3.30}$$

The 2nd column of Table 3.6 presents the results of all three approaches with the point estimators of $\boldsymbol{\theta}$ and the log odds ratios $\delta_{j_1 j_2}$ and associated 95% CIs. Such results show that SES and PES are superior than BMS in terms of reducing TLRs at one year, which is in line with the results reported by Hoaglin et al. (2011) and Yang et al. (2013a).

As in the previous example, we create a "contaminated" data set for CAD study. Specifically, we modify only one small study, TAXUS-I. This intends to show that, if not handled carefully, a single small outlying study can dominate all other large studies in meta-analysis and lead to counterintuitive results. The study TAXUS-I compared BMS vs PES. It reported three events in BMS group, zero event in PES group, and had only 30 patients enrolled in each group. Suppose that a transcription error made by mistakenly interchanged the number of event reported in the two groups in TAXUS-I. This transcription error is prone since the two groups have the same total number of patients. The error leads the TAXUS-I observed a counter example to the majority of the studies, as the common medical practice and our previous analysis on the original data set indicated that PES is superior to BMS for reducing TLRs at one year for CAD patients. Yet it does not change

the overall conclusion because study TAXUS-I is only one small study among all 37 studies involved in the meta-analysis (results not shown).

However, if instead of imputing 0.5 to the zero event, suppose an extreme small value, say $10^{-176}$ is imputed for zero. As a result, the observed log odds magnified by a factor of 100, decreasing from $\log(0.5/(30-0.5)) = -4.0775$ to $-407.75$, which is far away from the log odds values obtained from the other 22 studies involved BMS. Therefore, we create 'contaminated" data set that includes one small study that is strongly in favor of the performance of BMS. Intuitively, impute 0.5 or any other values should not change the result very much due to the relative small sample size. Thus, it is a surprise that the analysis reveals this single study can dominate all other large studies in meta-analysis.

Apply the same three approaches to the contaminated data set, the results are presented on the last column of Table 3.6. It is obvious that the results of the CD[Efficient] method and Bayesian method are dominated by the modified TAXUS-I, and conclude that BMS is the stents with least TLRs at one year. On the other hand, the CD robust methods (CD[Robust-M Estimation] and CD[Robust-Adaptive Weights]) holds on to the original conclusion regarding to the superiority of SES and PES over BMS, by limiting the impact of this single small extreme study TAXUS-I.

In this section, we illustrate the performance of the proposed multivariate meta-analysis approaches by comparing the results of the efficient CD approach, the robust CD approaches, and a corresponding Bayesian approach on three real data examples from the literatures. The studies involved in these data sets have been shown to be consistent to each other by previous analysis, and consequently, the efficient CD approach, robust CD approaches and the Bayesian approach are expected to obtain comparable results. These original data sets are further manipulated to create artificially "contaminated" data sets by introducing some outlying studies. These outlying studies have summary statistics which are far away

Table 3.5: CAD Trial Data Set, Target Lesion Revascularisation at 1 Year

| Study | BMS (A) | | SES (B) | | PES (C) | |
|---|---|---|---|---|---|---|
| | $r_{ij}$ | $n_{ij}$ | $r_{ij}$ | $n_{ij}$ | $r_{ij}$ | $n_{ij}$ |
| BASKET | 35 | 281 | 25 | 264 | 25 | 281 |
| C-SIRIUS | 11 | 50 | 2 | 50 | — | — |
| DECODE | 8 | 29 | 5 | 54 | — | — |
| DIABETES | 27 | 80 | 6 | 80 | — | — |
| E-SIRIUS | 44 | 177 | 8 | 175 | — | — |
| Ortolani 2007 | 11 | 52 | 6 | 52 | — | — |
| Pache 2005 | 51 | 250 | 25 | 250 | — | — |
| PRISON II | 20 | 100 | 4 | 100 | — | — |
| RAVEL | 16 | 118 | 1 | 120 | — | — |
| RRISC | 10 | 37 | 6 | 38 | — | — |
| SCANDSTENT | 47 | 159 | 4 | 163 | — | — |
| SCORPIUS | 20 | 95 | 5 | 95 | — | — |
| SESAMI | 19 | 160 | 7 | 160 | — | — |
| SES-SMART | 27 | 128 | 9 | 129 | — | — |
| SIRIUS | 106 | 525 | 26 | 533 | — | — |
| TYPHOON | 45 | 357 | 13 | 355 | — | — |
| HAAMUS-TENT | 9 | 82 | — | — | 3 | 82 |
| PASSION | 23 | 309 | — | — | 16 | 310 |
| TAXUS I | 3 | 30 | — | — | 0 | 30 |
| TAXUS II | 39 | 269 | — | — | 13 | 260 |
| TAXUS IV | 96 | 652 | — | — | 28 | 662 |
| TAXUS V | 107 | 579 | — | — | 62 | 577 |
| TAXUS VI | 46 | 227 | — | — | 19 | 219 |
| Cervinka 2006 | — | — | 1 | 37 | 2 | 33 |
| CORPAL | — | — | 22 | 331 | 25 | 321 |
| Han 2006 | — | — | 9 | 202 | 11 | 196 |
| ISAR-DESIRE | — | — | 14 | 100 | 22 | 100 |
| ISAR-DIABETES | — | — | 9 | 125 | 15 | 125 |
| ISAR-SMART3 | — | — | 16 | 180 | 29 | 180 |
| LONG DES II | — | — | 6 | 250 | 18 | 250 |
| Petronio 2007 | — | — | 1 | 42 | 1 | 43 |
| PROSIT | — | — | 3 | 116 | 9 | 115 |
| REALITY | — | — | 44 | 684 | 43 | 669 |
| SIRTAX | — | — | 30 | 503 | 54 | 509 |
| SORT OUT II | — | — | 40 | 1065 | 46 | 1033 |
| TAXi | — | — | 4 | 102 | 2 | 100 |
| Zhang 2006 | — | — | 14 | 225 | 16 | 187 |

Table 3.6: Meta-analysis of results for log-odds of the treatments in CAD data set

| | Original Data Set | Contaminated Date Set |
|---|---|---|
| CD[Efficient] | | |
| $\widehat{\theta}_A$ (95%CI) | $-1.4978(-1.6784, -1.3173)$ | $-2.9203(-3.1008, -2.7397)$ |
| $\widehat{\theta}_B$ (95%CI) | $-2.7864(-2.9646, -2.6083)$ | $-2.9888(-3.1670, -2.8107)$ |
| $\widehat{\theta}_C$ (95%CI) | $-2.4365(-2.6453, -2.2276)$ | $-2.5592(-2.7680, -2.3503)$ |
| $\widehat{\delta}_{AB}$ (95%CI) | $-1.2886(-1.5026, -1.0746)$ | $-0.0686(-0.2826, \ 0.1454)$ |
| $\widehat{\delta}_{AC}$ (95%CI) | $-0.9386(-1.1606, -0.7166)$ | $0.3611(\ 0.1391, \ 0.5831)$ |
| $\widehat{\delta}_{BC}$ (95%CI) | $0.3500(\ 0.1639, \ 0.5360)$ | $0.4297(\ 0.2436, \ 0.6157)$ |
| CD[Robust-M Estimation] | | |
| $\widehat{\theta}_A$ (95%CI) | $-1.4889(-1.6737, -1.3042)$ | $-1.4850(-1.6697, -1.3002)$ |
| $\widehat{\theta}_B$ (95%CI) | $-2.7944(-2.9768, -2.6121)$ | $-2.7870(-2.9694, -2.6047)$ |
| $\widehat{\theta}_C$ (95%CI) | $-2.4478(-2.6615, -2.2341)$ | $-2.4317(-2.6454, -2.2180)$ |
| $\widehat{\delta}_{AB}$ (95%CI) | $-1.3055(-1.5246, -1.0865)$ | $-1.3020(-1.5211, -1.0830)$ |
| $\widehat{\delta}_{AC}$ (95%CI) | $-0.9589(-1.1860, -0.7317)$ | $-0.9467(-1.1739, -0.7195)$ |
| $\widehat{\delta}_{BC}$ (95%CI) | $0.3467(\ 0.1563, \ 0.5371)$ | $0.3553(\ 0.1649, \ 0.5457)$ |
| CD[Robust-Adaptive Weights] | | |
| $\widehat{\theta}_A$ (95%CI) | $-1.4125(-1.6987, -1.1264)$ | $-1.4126(-1.6987, -1.1265)$ |
| $\widehat{\theta}_B$ (95%CI) | $-2.7709(-3.0469, -2.4949)$ | $-2.7709(-3.0469, -2.4949)$ |
| $\widehat{\theta}_C$ (95%CI) | $-2.4448(-2.7713, -2.1184)$ | $-2.4448(-2.7713, -2.1184)$ |
| $\widehat{\delta}_{AB}$ (95%CI) | $-1.3584(-1.7302, -0.9865)$ | $-1.3583(-1.7302, -0.9864)$ |
| $\widehat{\delta}_{AC}$ (95%CI) | $-1.0323(-1.4303, -0.6342)$ | $-1.0322(-1.4303, -0.6342)$ |
| $\widehat{\delta}_{BC}$ (95%CI) | $0.3261(\ 0.0608, \ 0.5913)$ | $0.3261(\ 0.0608, \ 0.5913)$ |
| Bayesian Model | | |
| $\widehat{\theta}_A$ (95%CI) | $-1.5004(-1.6910, -1.3140)$ | $-18.3014(-52.4248, 16.2610)$ |
| $\widehat{\theta}_B$ (95%CI) | $-2.7919(-2.9940, -2.6070)$ | $-2.7793(-3.0030, -2.5770)$ |
| $\widehat{\theta}_C$ (95%CI) | $-2.4457(-2.6600, -2.2310)$ | $-2.4936(-2.7300, -2.2720)$ |
| $\widehat{\delta}_{AB}$ (95%CI) | $-1.2915(-1.5280, -1.0640)$ | $15.5222(-19.0913, 49.5858)$ |
| $\widehat{\delta}_{AC}$ (95%CI) | $-0.9453(-1.2000, -0.7150)$ | $15.8079(-18.7015, 50.1003)$ |
| $\widehat{\delta}_{BC}$ (95%CI) | $0.34620(\ 0.1425, \ 0.5619)$ | $0.2857(\ 0.0308, \ 0.5456)$ |

from the original majorities. In these situations, the results from the efficient CD approach and the Bayesian approach are drastically influenced by these outlying studies, whereas the results from the robust CD approaches are not. Thus, the robust CD approaches resist to model-misspecification, and outperform the efficient CD approach and the Bayesian approach when performing meta-analysis on data sets with potential outliers.

## 3.5 Discussion

In this paper, we consider multivariate normal CDs and construct individual CDs based on asymptotic normality. The asymptotic normality assumption might seem strong, but, in fact, all MLE type of combinings have assumed such an assumption (Fraser and McDunnough, 1984). Further, almost every CD used for combining is usually asymptotically normal, in the sense that all $H_i(\boldsymbol{\theta}, \boldsymbol{y}_i) \approx \boldsymbol{\Psi}\{\mathbf{S}_i^{-1/2}(\boldsymbol{\theta} - T_i(\boldsymbol{y}_i))\}$ for some statistic $T_i(\boldsymbol{y}_i)$ and scaling $\mathbf{S}_i$ (Hannig and Xie, 2012). At last, the multivariate normal CD is the only known multivariate CD under the $l$-CD definition – joint CD through Craeér-Wold device (Singh et al., 2007).

The proposed CD combining method is broadly applicable. It can be used to combine CDs from independent studies regardless the method used to construct such CDs. For example, one challenging meta-analysis problem is to synthesize studies had different types of outcomes (Dominici and Parmigiani, 2005). The conventional meta-analysis method can not combine studies reported binary outcomes (e.g., high or low), with studies reported continues outcomes (e.g., the number of blood platelets). The CD combining method instead can combine such studies by building CDs regarding to the same parameter(s), which can be measured from each individual studies.

The proposed CD combining method is also very flexible, in terms of the choice on function

$a_0(\cdot)$ and weights $\mathbf{W}_i$s. We focus on two choices of the function $a_0(\cdot)$ when applying our method to multivariate meta-analysis. The choice of $a_0(\cdot) = \Phi^{-1}(\cdot)$ in general provides efficient estimators, whereas $a_0(t) = t$ produces robust estimators. Other choices may be able to integrate such two features, for example, $a_0(t) = F^{-1}(t)$ where $F(t) = \dfrac{\Phi(t) - \Phi(-q)}{\Phi(q) - \Phi(-q)}$. Here $F(\cdot)$ in fact is the distribution function of a random variable following truncated normal distribution with support $[-q, q]$. Intuitively, this method leads to combined estimator mimic the properties of trimmed mean. As a result, such $a_0(\cdot)$ leads to a robust combined CD with higher relative efficiency when no contaminated studies involved in the meta-analysis. For example, the relative efficiency is $98.4238\% - 99.9974\%$ for $q = 1$ to $q = 4$ whereas the relative efficiency for using $a_0(t) = t$ is $97.7205\%$.

In meta-analysis, it is often the case that only the summary statistics of each endpoints but not their correlations, e.g., $\text{var}(y_{ij}), j = 1, \ldots, p$ instead of $\text{var}(\boldsymbol{y}_i)$, are reported in the literatures. As a result, the within study covariance matrices $\mathbf{S}_i$ would have correlations missing, or more generally, are misspecified. For the proposed method in Theorem 3.2, replace $\mathbf{S}_i$ by a surrogate "working" covariance matrix $\mathbf{S}_{\mathtt{i,W}}$, e.g., covariance matrix without correlations, can lead to an aCD function for $\boldsymbol{\theta}$, say $H_{\mathtt{w}}^{(c)}(\cdot)$. The center point of $H_{\mathtt{w}}^{(c)}(\cdot)$, $\widehat{\boldsymbol{\theta}}_{\mathtt{W}}$, is a consistent point estimator for $\boldsymbol{\theta}$, and it is asymptotically normally distributed with a "sandwich" covariance matrix, if: i) $\mathbf{S}_{\mathtt{i,W}}$ symmetric and positive semi-definite; and ii) $(n_i \mathbf{S}_{\mathtt{i,W}})^{-1} \to \mathbf{M}_i$ in probability as $n_i \to \infty$, where $\mathbf{M}_i$ is a fixed matrix (cf., Theorem 2.3 in Liu, 2012). Therefore, the proposed method is flexible to include studies with misspecified within study covariance matrices.

The robust meta-analysis approach is important because it provides protection on model misspecification. In the context of network meta-analysis, where direct and indirect evidence from disparate sources are combined, the "evidence inconsistency" can be interpreted in terms of "variance discrepancy". Intuitively, indirect evidence, which was obtained by

integrating different sources of information, often has larger underlying variance than direct evidence, due to potential compromise of study control. For example, a set of clinical trials is gathered for making inference on the relative performance of treatments A vs C. Indirect evidence, which comes from combining studies compared A vs B and B vs C, would in general has larger variance than direct evidence, which comes from studies directly compared A vs C. The reason is that the randomization is controlled only within individual studies, and such randomization might be compromised when combining two studies for obtaining indirect evidence. As a result, the contaminated model (3.9) might be more appropriate than the conventional random-effects model (3.7).

At last, we want to demonstrate that the proposed method can be applied to various multivariate meta-analysis problems. It is computationally efficient and easy to be built in a statistical packages, as *mvmeta* (White, 2009).

## 3.6 Appendix

**Proof of Theorem 3.1:**

(i) Point estimator:

For any $0 < \epsilon < \frac{1}{2}$ and $\boldsymbol{\lambda} \in \Re^p$, define $H_{(n),\boldsymbol{\lambda}}(t) = \Phi((\boldsymbol{\lambda}^{\mathbf{T}}\mathbf{S}_{(n)}\boldsymbol{\lambda})^{-1/2}(t - \boldsymbol{\lambda}^{\mathbf{T}}\bar{\boldsymbol{y}}_{(n)}))$ and consider $L_{(n),\boldsymbol{\lambda}} = H^{-1}_{(n),\boldsymbol{\lambda}}(1-\epsilon) - H^{-1}_{(n),\boldsymbol{\lambda}}(\epsilon)$. If $\sigma^2_{\mathtt{min}} \propto 1/n \to 0$ then $\boldsymbol{\lambda}^{\mathbf{T}}\mathbf{S}_{(n)}\boldsymbol{\lambda} \to 0$, and thus $L_{(n),\boldsymbol{\lambda}} \to 0$, in probability, as the sample size $n \to \infty$. By Singh et al. (2007), $\boldsymbol{\lambda}^{\mathbf{T}}\widehat{\boldsymbol{\theta}}_{(n)}$ is a consistent estimator for $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$.

Take $\boldsymbol{\lambda} = \mathbf{e}_j$ where $\mathbf{e}_j$ is a $p \times 1$ vector has 1 on the $j$-th position and zero on all others for

$j = 1, \ldots, p$, then it follows that, for any $\delta > 0$,

$$\begin{aligned}
\Pr\left\{||\widehat{\boldsymbol{\theta}}_{(n)} - \boldsymbol{\theta}_0||_2 \geq \delta\right\} &= \Pr\left\{\sum_{j=1}^{p}(\widehat{\theta}_{(n),j} - \theta_{0,j})^2 \geq \delta\right\} \\
&\leq \sum_{j=1}^{p}\Pr\left\{(\widehat{\theta}_{(n),j} - \theta_{0,j})^2 \geq \delta\right\} \\
&= \sum_{j=1}^{p}\Pr\left\{|\widehat{\theta}_{(n),j} - \theta_{0,j}| \geq \sqrt{\delta}\right\} \\
&= \sum_{j=1}^{p} 2\left(1 - \Phi\left(\sqrt{\delta}/(\mathbf{e}_j^{\mathbf{T}}\mathbf{S}_{(n)}\mathbf{e}_j)^{1/2}\right)\right) \\
&\leq \sum_{j=1}^{p} 2\left(1 - \Phi(\sqrt{\delta}/\sigma_{\texttt{min}})\right) \\
&\to 0 \quad \text{as } n \to \infty
\end{aligned} \tag{3.31}$$

where $\boldsymbol{\theta}_0$ is the true parameter value. Thus, $\widehat{\boldsymbol{\theta}}_{(n)}$ is a consistent estimator for $\boldsymbol{\theta}$.

(ii) Hypothesis testing:

We first show that the conclusion is valid for $B_{\texttt{rect}} = \prod_{j=1}^{p}\mathbf{J}_j$ where $\mathbf{J}_j$ represents an interval $(b_{j1}, b_{j2})$.

Let $\boldsymbol{\lambda}_j^{\mathbf{T}} = \mathbf{e}_j^{\mathbf{T}}\mathbf{S}_{(n)}^{-1/2}$ then

$$H_{(n)}(\boldsymbol{\theta}) = \prod_{j=1}^{p} H_{(n),\boldsymbol{\lambda}_j}(\boldsymbol{\lambda}_j^{\mathbf{T}}\boldsymbol{\theta}). \tag{3.32}$$

Let $\widetilde{\mathbf{J}}_j = \big(\min(\boldsymbol{\lambda}_j^{\mathbf{T}}\mathbf{b}_1, \boldsymbol{\lambda}_j^{\mathbf{T}}\mathbf{b}_2), \max(\boldsymbol{\lambda}_j^{\mathbf{T}}\mathbf{b}_1, \boldsymbol{\lambda}_j^{\mathbf{T}}\mathbf{b}_2)\big)$, where $\mathbf{b}_1 = (b_{11}, b_{21}, \ldots, b_{p1})^{\mathbf{T}}$ and $\mathbf{b}_2 = (b_{12}, b_{22}, \ldots, b_{p2})^{\mathbf{T}}$. By Singh et al. (2007),

$$\Pr\{\boldsymbol{\lambda}_j^{\mathbf{T}}\boldsymbol{\theta}_0 \in \widetilde{\mathbf{J}}_j\} = \int_{\widetilde{\mathbf{J}}_j} dH_{(n),\boldsymbol{\lambda}_j}(t), \tag{3.33}$$

Note that the probability in (3.32) are independent for $j = 1, \ldots, p$, and thus

$$\begin{aligned} \Pr\{\mathbf{S}_{(n)}^{-1/2}\boldsymbol{\theta}_0 \in \prod_{j=1}^{p} \widetilde{\mathbf{J}}_j\} &= \prod_{j=1}^{p} \Pr\{\boldsymbol{\lambda}_j^{\mathbf{T}}\boldsymbol{\theta}_0 \in \widetilde{\mathbf{J}}_j\} \\ &= \prod_{j=1}^{p} \int_{\boldsymbol{\lambda}_j^{\mathbf{T}}\boldsymbol{\theta} \in \widetilde{\mathbf{J}}_j} dH_{(n),\boldsymbol{\lambda}_j}(\boldsymbol{\lambda}_j^{\mathbf{T}}\boldsymbol{\theta}) \\ &= \int_{\mathbf{S}_{(n)}^{-1/2}\boldsymbol{\theta} \in \prod_{j=1}^{p} \widetilde{\mathbf{J}}_j} d \prod_{j=1}^{p} H_{(n),\boldsymbol{\lambda}_j}(\boldsymbol{\lambda}_j^{\mathbf{T}}\boldsymbol{\theta}) \\ &= \int_{\mathbf{S}_{(n)}^{-1/2}\boldsymbol{\theta} \in \prod_{j=1}^{p} \widetilde{\mathbf{J}}_j} dH_{(n)}(\boldsymbol{\theta}) \end{aligned} \tag{3.34}$$

which implies

$$\Pr\{\boldsymbol{\theta}_0 \in B_{\mathtt{rect}}\} = \Pr\left\{\boldsymbol{\theta}_0 \in \prod_{j=1}^{p} \mathbf{J}_j\right\} = \int_{\prod_{j=1}^{p} \mathbf{J}_j} dH_{(n)}(\boldsymbol{\theta}) = \int_{B_{\mathtt{rect}}} dH_{(n)}(\boldsymbol{\theta}) \tag{3.35}$$

and therefore

$$\Pr\{\boldsymbol{\theta}_0 \in B_{\mathtt{rect}}^c\} = 1 - \Pr\{\boldsymbol{\theta}_0 \in B_{\mathtt{rect}}\} = \int_{B_{\mathtt{rect}}^c} dH_{(n)}(\boldsymbol{\theta}) = H_{(n)}(B_{\mathtt{rect}}^c) \tag{3.36}$$

provides the $p$-value for $K_0 : \boldsymbol{\theta}_0 \in B_{\mathtt{rect}}$ versus $K_1 : \boldsymbol{\theta}_0 \in B_{\mathtt{rect}}^c$.

Based on Dudley (2002), any set $B \in \mathcal{B}$ is unions and/or intersections of some $B_{\mathtt{rect}}$ in $\Re^p$, and this completes the proof.

(iii) Confidence region:

Let $B = \{\boldsymbol{\theta} : ||\mathbf{S}_{(n)}^{-1/2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{(n)})||_2 \leq q_{1-\alpha}(\chi_p^2)\}$, then $\Pr\{\boldsymbol{\theta}_0 \in B\} = \int_B dH_{(n)}(\boldsymbol{\theta}) = 1 - \alpha$ and the conclusion follows by the dual relationship between hypothesis testing and confidence region.

**Proof of Theorem 3.2:**

We first show that the function $H(\cdot)$ defined in in Theorem 3.2 is an asymptotically cumulative distribution function on the parameter space given $\boldsymbol{Y}_i$.

Suppose $\mathbb{E}[\mathbf{a}_0[\mathbf{U}_i]] = m\mathbf{1}$ and $\text{var}(\mathbf{a}_0[\mathbf{U}_i]) = v\mathbf{I}$, where $\mathbf{1}$ is $p \times 1$ vector with all entries equal to 1 and $\mathbf{I}$ is $p \times p$ identity matrix. Without loss of generality, suppose $a_0(\cdot)$ monotonic non-decreasing and has first derivative, then asymptotically,

$$G^{(c)}(\mathbf{t}) = \boldsymbol{\Psi}\left(\left(v \cdot \sum_{i=1}^{k} \mathbf{W}_i\right)^{-1/2}\left(t - \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \cdot m\mathbf{1}\right).\right) \tag{3.37}$$

Thus,

$$\begin{aligned} H^{(c)}(\boldsymbol{\theta}) &= G^{(c)}\left\{\textstyle\sum_{i=1}^{k} \mathbf{W}_i^{1/2}\mathbf{a}_0[\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta})]\right\} \\ &= \boldsymbol{\Psi}\left(\left(v \cdot \sum_{i=1}^{k} \mathbf{W}_i\right)^{-1/2}\left(\sum_{i=1}^{k} \mathbf{W}_i^{1/2}(\mathbf{a}_0[\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta})] - m\mathbf{1})\right)\right) \end{aligned} \tag{3.38}$$

The inference based on (3.38) is asymptotically equivalent to an M-estimation approach solves:

$$\sum_{i=1}^{k} \mathbf{W}_i^{1/2}(\mathbf{a}_0[\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta})] - m\mathbf{1}) = \mathbf{0} \tag{3.39}$$

Suppose $\boldsymbol{\theta}_0$ is the true parameter value and consider the Taylor expansion of $a_0(H_{ij}(\boldsymbol{\lambda}_{ij}^{\mathbf{T}}\boldsymbol{\theta}))$ at $Z_{ij} = \boldsymbol{\lambda}_{ij}^{\mathbf{T}}(\boldsymbol{\theta}_0 - \boldsymbol{Y}_i)$:

$$\begin{aligned} a_0(H_{ij}(\boldsymbol{\lambda}_{ij}^{\mathbf{T}}\boldsymbol{\theta})) &= a_0(\Phi(\boldsymbol{\lambda}_{ij}^{\mathbf{T}}(\boldsymbol{\theta} - \boldsymbol{Y}_i))) \\ &= a_0(\Phi(Z_{ij})) + a_0'(\Phi(Z_{ij}))\phi(Z_{ij}) \cdot \boldsymbol{\lambda}_{ij}^{\mathbf{T}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(n^{-1/2}) \end{aligned} \tag{3.40}$$

Thus,

$$\begin{aligned} & a_0[\mathbf{H}_i^{\mathbf{\Lambda}}(\boldsymbol{\theta})] \\ = & \begin{bmatrix} a_0(\Phi(Z_{i1})) \\ \vdots \\ a_0(\Phi(Z_{ip})) \end{bmatrix} + \begin{bmatrix} a_0'(\Phi(Z_{i1}))\phi(Z_{i1}) & & \\ & \ddots & \\ & & a_0'(\Phi(Z_{ip}))\phi(Z_{ip}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{i1}^{\mathbf{T}} \\ \vdots \\ \boldsymbol{\lambda}_{ip}^{\mathbf{T}} \end{bmatrix} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(n^{-1/2}\mathbf{1}) \\ = & a_0[\mathbf{U}_i] - c\mathbf{\Lambda}_i(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(n^{-1/2}\mathbf{1}) \end{aligned} \tag{3.41}$$

where $c = \mathbb{E}[a_0'(\Phi(Z_{ij}))\phi(Z_{ij})]$ and $Z_{ij}$ are independent standard normal random variable, $\mathbf{U}_i = (U_{i1}, \ldots, U_{ip})^{\mathbf{T}}$, and $U_{ij}$ are independent $\mathbf{U}[0,1]$ random variables.

Therefore, the Taylor expansion of M-estimating equation (3.39) around $\boldsymbol{\theta}_0$ yields:

$$\widehat{\boldsymbol{\theta}}^{(c)} - \boldsymbol{\theta}_0 = \left( c \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \mathbf{\Lambda}_i \right)^{-1} \left( \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \left( a_0[\mathbf{U}_i] - m\mathbf{1} \right) \right) + o_p(n^{-1/2}\mathbf{1}) \tag{3.42}$$

which implies $\text{var}(\widehat{\boldsymbol{\theta}}^{(c)}) \to \left( c \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \mathbf{\Lambda}_i \right)^{-1} \left( v \sum_{i=1}^{k} \mathbf{W}_i \right) \left( c \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \mathbf{\Lambda}_i \right)^{-\mathbf{T}}$ as $k \to \infty$, and

$$H^{(c)}(\boldsymbol{\theta}) \to \mathbf{\Psi}\left( \mathbf{S}^{(c)^{-1/2}} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(c)}) \right) \quad \text{as} \quad k \to \infty \tag{3.43}$$

where $\mathbf{S}^{(c)} = \left( c \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \mathbf{\Lambda}_i \right)^{-1} \left( v \sum_{i=1}^{k} \mathbf{W}_i \right) \left( c \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \mathbf{\Lambda}_i \right)^{-\mathbf{T}}$.

Thus, $H^{(c)}(\boldsymbol{\theta})$ is an asymptotically cumulative distribution function for $\boldsymbol{\theta}$.

Because $H^{(c)}(\boldsymbol{\theta})$ is a distribution function on $\Re^p$ for given $\boldsymbol{Y}_i, i = 1, \ldots, k$, by construction theorem (Chung and AitSahlia, 2003), we can construct a random vector on $\Re^p$ such that $\boldsymbol{\xi}^{(c)} | \boldsymbol{Y}_i, i = 1, \ldots, k \sim H^{(c)}(\cdot)$ for each given $\boldsymbol{Y}_i, i = 1, \ldots, k$. To show $H^{(c)}(\boldsymbol{\theta})$ is an asymptotic multivariate normal CD for $\boldsymbol{\theta}$, define $H_{\boldsymbol{\lambda}}(t) = \Pr\{\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\xi}^{(c)} \leq t | \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k\}$ for any given vector $\boldsymbol{\lambda} \in \Re^p$, based on Definition 3.1, it suffices to show that $H_{\boldsymbol{\lambda}}(t)$ is univariate normal CD function for $\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\theta}$.

First note that $H_{\boldsymbol{\lambda}}(t)$ goes from 0 to 1 monotonically, as $t$ goes from $-\infty$ to $\infty$. Thus, $H_{\boldsymbol{\lambda}}(t)$ is a cumulative distribution function. Second, at the true parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\mathbf{a}_0[\mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta})] = \mathbf{a}_0[\boldsymbol{\Phi}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta} - \boldsymbol{Y}_i)] = \mathbf{a}_0[\boldsymbol{\Phi}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}_0 - \boldsymbol{Y}_i)] = \mathbf{a}_0[\mathbf{U}_i]. \tag{3.44}$$

Let $\boldsymbol{\eta}_i = \mathbf{a}_0[\mathbf{U}_i]$ then,

$$\begin{aligned} H^{(c)}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k; \boldsymbol{\theta}_0) &= \boldsymbol{\Psi}\left(\left(v \cdot \sum_{i=1}^{k} \mathbf{W}_i\right)^{-1/2} \left(\sum_{i=1}^{k} \mathbf{W}_i^{1/2} \boldsymbol{\eta}_i - \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \cdot m\mathbf{1}\right)\right) \\ &= \boldsymbol{\Psi}\left(\left(v \cdot \sum_{i=1}^{k} \mathbf{W}_i\right)^{-1/2} \left(\sum_{i=1}^{k} \mathbf{W}_i^{1/2} (\boldsymbol{\eta}_i - m\mathbf{1})\right)\right) \\ &\to \boldsymbol{\Psi}\left(v^{-1/2} \left(\sum_{i=1}^{k} \mathbf{W}_i\right)^{-1/2} \left(\left(\sum_{i=1}^{k} \mathbf{W}_i\right)^{1/2} (\boldsymbol{\eta} - m\mathbf{1})\right)\right) \\ &\qquad\qquad \text{as } k \to \infty \\ &= \boldsymbol{\Psi}\left(v^{-1/2}(\boldsymbol{\eta} - m\mathbf{1})\right) \end{aligned} \tag{3.45}$$

where $\boldsymbol{\eta}$ is $p \times 1$ random vector with distribution $N(m\mathbf{1}, v\mathbf{I})$. And it is straightforward to show that, at the true parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\Pr\{H_{\boldsymbol{\lambda}}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k) \le s\} = \Pr\left\{\Phi\left((||\boldsymbol{\lambda}||_2 \cdot v)^{-1/2} \cdot \boldsymbol{\lambda}^{\mathbf{T}}(\boldsymbol{\eta} - m\mathbf{1})\right) \le s\right\} = s \tag{3.46}$$

Thus, we have established that, at the true $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and as a function of the sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k$, $H_{\boldsymbol{\lambda}}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k)$ follows the uniform distribution $U[0, 1]$. Q.E.D. □

**Proof of Theorem 3.3:**

Let $\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}$ denote the center point of $\widetilde{H}_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$ in (3.11). Then, it is also the solution of the M-estimating equation (3.12), and conclusion (1) follows from the standard argument of an M-estimating equation (c.f., Huber, 1964).

The conclusion (2) follows from the standard M-estimation as well, except that in our case we have to incorporate the given weights $\mathbf{e}_j^{\mathbf{T}}\mathbf{W}_i^{1/2}\mathbf{e}_j$ for some vector $\mathbf{e}_j \in \Re^p$. Note that $v_{i(p)} \leq \mathbf{e}_j^{\mathbf{T}}\mathbf{W}_i^{1/2}\mathbf{e}_j \leq v_{i(1)}$ for any vector $\mathbf{e}_j \in \Re^p$. On the other hand, $H_{ij}(t)$ is bounded between 0 and 1, so $H_{ij}(t) - 1/2$ is bounded between $-1/2$ and $1/2$. Thus, the minimum and maximum contribution of the $i$-th study to the equation is $\pm v_{i(p)}/2$ and $\pm v_{i(1)}/2$, respectively.

For the estimation of the $j$-th component of $\boldsymbol{\theta}$, it is sufficient and necessary to require that the summation of $\mathbf{e}_j^{\mathbf{T}}\mathbf{W}_i^{1/2}\mathbf{e}_j$ over the outlying studies dominates the summation of $\mathbf{e}_j^{\mathbf{T}}\mathbf{W}_i^{1/2}\mathbf{e}_j$ over the non-outlying studies in order to break down the estimating equation so that the solution of the estimating equation approaches infinity. In consideration of the worst possible case, the breakdown point is obtained as the one stated in conclusion (2). Q.E.D. □

**Proof of Theorem 3.4:**

The Taylor expansion of $H_{ij}(\boldsymbol{\lambda}_{ij}^{\mathbf{T}}\boldsymbol{\theta})$ at $Z_{ij} = \boldsymbol{\lambda}_{ij}^{\mathbf{T}}(\boldsymbol{\theta}_0 - \boldsymbol{Y}_i)$ gives:

$$H_{ij}(\boldsymbol{\lambda}_{ij}^{\mathbf{T}}\boldsymbol{\theta}) = \Phi(\boldsymbol{\lambda}_{ij}^{\mathbf{T}}(\boldsymbol{\theta} - \boldsymbol{Y}_i)) = \Phi(Z_{ij}) + \boldsymbol{\lambda}_{ij}^{\mathbf{T}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \cdot \phi(Z_{ij}) + o_p(n^{-1/2}) \tag{3.47}$$

Therefore,

$$\begin{aligned} \mathbf{H}_i^{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) &= \begin{bmatrix} \Phi(Z_{i1}) \\ \vdots \\ \Phi(Z_{ip}) \end{bmatrix} + \begin{bmatrix} \phi(Z_{i1}) & & \\ & \ddots & \\ & & \phi(Z_{ip}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{i1}^{\mathbf{T}} \\ \vdots \\ \boldsymbol{\lambda}_{ip}^{\mathbf{T}} \end{bmatrix} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(n^{-1/2}\mathbf{1}) \\ &= \mathbf{U}_i - \tfrac{1}{2\sqrt{\pi}}\mathbf{W}_i^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(n^{-1/2}\mathbf{1}) \end{aligned} \tag{3.48}$$

where $Z_{ij}$ are independent standard normal random variable, $\mathbf{U}_i = \begin{bmatrix} U_{i1} \\ \vdots \\ U_{ip} \end{bmatrix}$, and $U_{ij}$ are independent $\mathbf{U}[0,1]$ random variables.

Given Model (3.7) is true, $\widetilde{H}_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$ and $H_{\mathtt{R}}^{(c)}(\boldsymbol{\theta})$ are essentially equivalent. Their center point $\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}$ is the solution of the M-estimating equation:

$$\sum_{i=1}^{k} \mathbf{W}_i^{1/2} \left( \mathbf{H}_i^{\mathbf{\Lambda}}(\boldsymbol{\theta}) - \frac{1}{2} \cdot \mathbf{1} \right) = \mathbf{0} \tag{3.49}$$

The Taylor expansion of above M-estimating equation around $Z_{ij}$ yields,

$$\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)} - \boldsymbol{\theta}_0 = \left( \frac{1}{2\sqrt{\pi}} \sum_{i=1}^{k} \mathbf{W}_i \right)^{-1} \left( \sum_{i=1}^{k} \mathbf{W}_i^{1/2} \left( \mathbf{U}_i - \frac{1}{2}\mathbf{1} \right) \right) + o_p(n^{-1/2}\mathbf{1}) \tag{3.50}$$

which implies $\mathrm{var}(\widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}) \to \frac{\pi}{3}(\sum_{i=1}^{k} \mathbf{W}_i)^{-1}$ as $k \to \infty$, and

$$H_{\mathtt{R}}^{(c)}(\boldsymbol{\theta}) \to \boldsymbol{\Psi} \left( {\mathbf{S}_{\mathtt{R}}^{(c)}}^{-1/2} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\mathtt{R}}^{(c)}) \right) \text{ as } k \to \infty \tag{3.51}$$

where $\mathbf{S}_{\mathtt{R}}^{(c)} = \frac{\pi}{3}(\sum_{i=1}^{k} \mathbf{W}_i)^{-1}$.

From the efficient combination approach in (3.8):

$$H_{\mathtt{E}}^{(c)}(\boldsymbol{\theta}) = \boldsymbol{\Psi} \left( {\mathbf{S}_{\mathtt{E}}^{(c)}}^{-1/2} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{\mathtt{E}}^{(c)}) \right) \tag{3.52}$$

where $\mathbf{S}_{\mathtt{E}}^{(c)} = \left( \sum_{i=1}^{k} \mathbf{W}_i \right)^{-1}$ and $\widehat{\boldsymbol{\theta}}_{\mathtt{E}}^{(c)} = \left( \sum_{i=1}^{k} \mathbf{W}_i)^{-1} \sum_{i=1}^{k} \mathbf{W}_i \boldsymbol{y}_i \right)$.

Therefore, the asymptotic relative efficiency of $\widetilde{H}_{\mathtt{R}}^{(c)}(\cdot)$ compared to $H_{\mathtt{E}}^{(c)}(\cdot)$ is $\sqrt{3/\pi} \approx 0.9772$ as $n \to \infty$ and $k \to \infty$. Q.E.D. $\square$

# Chapter 4

# gmeta: An R Package Unified Meta-Analysis Methods Through Combining Confidence Distributions

In this additional chapter, we put forward a computing software that realizes the unified univariate meta-analysis framework proposed in Xie et al. (2011). The R package **gmeta** can provide an all-in-one solution for univariate meta-analysis problems.

## 4.1 Introduction

In this article, we introduce an R package **gmeta** that can help users carrying out all standard meta-analysis through a single function `gmeta()`. Simply speaking, meta-analysis is a statistical procedure that synthesizes findings from independent sources for decision making (Glass, 1976). It is widely used in scientific fields such as biology, chemistry, psychology and clinical trials analysis to combine current study results with historical results based on systematically reviewing published literatures. A throughout review of standard meta-analysis methods can be found in Normand (1999).

The **gmeta** package has a unified structure for performing meta-analysis. All methods are invoked from a single function `gmeta()`. Specifically, methods incorporated in the `gmeta()` function include, but are not limited to, $p$-value combination, fixed-effect and random-effects model-based meta-analysis, Mantel-Haenszel and Peto's method for synthesizing

2x2 tables. The `gmeta()` function imitates the structure of the well known `glm()` function, where generalized linear models are unified under a single function through the interrelated options on arguments `family` and `link`. Likewise, the `gmeta()` function unifies all standard meta-analysis methods through the options on arguments `method` and `linkfunc`. The option on `method` reflects the assumptions made for meta-analysis model, e.g., fixed-effect or random-effects model. The option on `linkfunc` defines the way information from individual study is handled for integration, e.g., for Fisher efficiency or Bahadur efficiency.

The `gmeta()` function also implements several robust meta-analysis methods. The motivation is to provide protection from model-misspecification, and limit the impact of unknown outlying studies. These robust methods are developed along the conception of using sample median, instead of sample mean, for the estimation of the population mean. Though the development requires a thorough understanding of confidence distribution (see, Section 4 of Xie et al., 2011), its practical use does not. The users only need to distinguish that one type of methods, `'-robust1'`, are designed for performing meta-analysis on a number of large studies, e.g. in the case where studies sample sizes $n_i \to \infty$, whereas the other type of methods, `'-robust2'`, are designed for performing meta-analysis on a large number of studies, e.g., in the case where the number of studies, $k \to \infty$. Moreover, the performance of these two types of robust methods are similar under common practical situations, so even unawareness of the difference is in general not an issue.

In medical researches, clinical trials are often presented in 2x2 tables, where the number of events are often assumed following Binomial distribution. The `gmeta()` function can take 2x2 tables as inputs and combine them through Mantel-Haenszel or Peto's method. However, both methods are based on large sample theory such as asymptotic normality, which leads to invalid inference when the sample size is small. The meta-analysis of clinical trials are also challenged by studies with rare events. For examples, studies with zero total

event, where neither the treatment nor the control group observe an event. Such studies are often excluded from the analysis or applied artificial continuity corrections to zero event (Nissen and Wolski, 2007; Efron, 1996). However, exclusion or continuity correction, either way, leads to suspiciously biased results (J Sweeting et al., 2004; Tian et al., 2009; Liu et al., 2013). Thus, it is desirable to have methods that can appropriately account studies with small sample size and/or rare events. Therefore, the `gmeta()` function incorporates methods proposed by Tian et al. (2009) and Liu et al. (2013). These two methods are called "exact" methods because of using Binomial distribution, instead of asymptotic normality, during the development. As a result, these methods can use all available studies without any artificial continuity corrections on the zero event, and provide correct inference, in terms of Type I error rate, on risk difference (RD) and log odd ratio (LOR), respectively.

As a final remark, we want to point out that the `gmeta()` function not only unifies different meta-analysis methods operationally, but also implements these methods under the same structure of combining confidence distributions (CDs). Intuitively, the CD combining method extends the traditional model-based meta-analysis, in a similar way as the copula extended the linear correlation for describing the dependence between random variables. The copula allows one to separate the estimation of the distribution of random vectors to the estimation of marginal and copula. It then use parameters to describe the strength and structure of the dependence. Likewise, the `gmeta()` function separates the process of summarizing information from individual studies and synthesizing those information. It summarizes the evidence from each study into a CD, and then combines these CDs using a unified general formula. This combination is flexible in the sense that various `linkfunc` can be taken and each provides different features, e.g., efficiency or robustness. We hold the introduction of the CD combining framework until illustrating the usage of **gmeta** with some simple examples, see Singh et al. (2005, 2007) and Xie et al. (2011) for the development of

CD concept and the CD combining framework.

This article proceeds as follows. Section 4.2 provides an overview of the **gmeta** package with simple examples. Section 4.3 reviews the concept of confidence distribution and the general CD combining method. It is the theoretical support for the `gmeta()` function unifying all meta-analysis methods. Section 4.4 illustrates the use of `gmeta()` function for $p$-value combination, model-based meta-analysis, and 2x2 tables synthesis through real data examples. Section 4.5 concludes the article with a discussion of planned further developments on the package.

## 4.2 Overview of the gmeta Package

The **gmeta** package implements the unified meta-analysis method described in Xie et al. (2011) using S3 methods. The main function `gmeta()` has the following arguments:

```
gmeta <- function(gmi, gmi.type = c('pivot', 'cd', 'pvalue', '2x2'),
    method = c('fixed-mle',
        'fixed-robust1', 'fixed-robust2', 'fixed-robust2(sqrt12)',
        'random-mm', 'random-reml', 'random-tau2',
        'random-robust1', 'random-robust2', 'random-robust2(sqrt12)',
        'fisher', 'normal', 'stouffer', 'min', 'tippett', 'max', 'sum',
        'MH', 'Mantel-Haenszel', 'Peto', 'exact1', 'exact2'),
    linkfunc = c('inverse-normal-cdf', 'inverse-laplace-cdf'),
    weight = NULL, study.names = NULL, gmo.xgrid = NULL, ci.level = 0.95,
    tau2 = NULL, mc.iteration = 10000, eta = 'Inf', verbose = FALSE,
    report.error = FALSE)
```

The key arguments are `method` and `linkfunc`, where `method` reflects the assumptions made for meta-analysis model, and `linkfunc` defines the way information from individual studies handled for integration. The following examples show the use of `gmeta()` without assuming any knowledge on the concept of CD.

### 4.2.1 A Small Example - Conventional Model-based Meta-analysis

Let us consider a hypothetical toy example, suppose four studies regarding the same parameter of interest are gathered for meta-analysis. The summary statistics reported are `y=c(-2,0,1,2)` and `s=c(1,1,1,1)` for study-specific means and associated standard deviations. As in line with Normand (1999), conventional fix-effect and random-effects model-based meta-analysis can be done through the following script.

```
> y   <- c(-2,0,1,2)
> s   <- c( 1,1,1,1)
> # original data set
> gdf <- data.frame(mns=y, sds=s)
>
> # conventional fixed-effect meta-analysis
> gmt.fix <- gmeta(gmi=gdf, method='fixed-mle',
+ gmo.xgrid=seq(-10,10,by=0.001))
> # summary of the combining results
> summary(gmt.fix)
                Model-Based Meta-Analysis through CD-Framework

Call:
```

```
gmeta.default(gmi = gdf, method = "fixed-mle", gmo.xgrid = seq(-10,
    10, by = 0.001))

Summary of Combined CD:
            mean median    stddev  ci.lower ci.upper
Combined CD 0.25   0.25 0.5000002 -0.729982 1.229982

Confidence level = 0.95

Summary of Individual CDs:
        mean median stddev    ci.lower    ci.upper
study-1   -2     -2      1 -3.95996402 -0.04003598
study-2    0      0      1 -1.95996402  1.95996402
study-3    1      1      1 -0.95996402  2.95996402
study-4    2      2      1  0.04003598  3.95996402

Confidence level = 0.95
>
> # conventional random-effects meta-analysis
> gmt.rdm <- gmeta(gmi=gdf, method='random-mm',
+ gmo.xgrid=seq(-10,10,by=0.001))
> # summary of the combining results
> summary(gmt.rdm)
                Model-Based Meta-Analysis through CD-Framework

Call:
```

```
gmeta.default(gmi = gdf, method = "random-mm", gmo.xgrid = seq(-10,
    10, by = 0.001))

Summary of Combined CD:
            mean median    stddev  ci.lower ci.upper
Combined CD 0.25   0.25 0.8539126 -1.423638 1.923638

Confidence level = 0.95

Summary of Individual CDs:
        mean median stddev  ci.lower ci.upper
study-1   -2     -2      1 -5.347276 1.347276
study-2    0      0      1 -3.347276 3.347276
study-3    1      1      1 -2.347276 4.347276
study-4    2      2      1 -1.347276 5.347276

Confidence level = 0.95
```

For model-based meta-analysis, the input `gmi` are summary statistics, means and standard deviations. The summary statistics are organized into a two columns data.frame or matrix, where each row takes the reported mean and standard deviation from an individual study. The name of each study can be specified through the row names of the data.frame or matrix, or through the argument `study.names`. If `NULL`, the default values are `study-[row.index]`. The argument `method` specifies the method used for meta-analysis, for example, `method='fixed-mle'` for fixed-effect model and `method='random-mm'` for random-effects model with moment estimator for heterogeneity. The argument `gmo.xgrid`

specifies the range and gridding points for evaluating the individual and combined CDs. It should cover the range of $y_i \pm 3s_i$ with fine gridding. We will discuss these arguments in detail later.

In this small example, the fixed-effect method obtains a combined point estimate (`mean`) 0.25 with standard deviation (`stddev`) 0.50. The random-effects method obtains the same point estimate (`mean`) 0.25 with a slightly larger standard deviation (`stddev`) 0.85. This is because the random-effects model assumes an additional layer of randomness on the study-specific means, whereas fixed-effect model assumes all studies have the same underlying true value of the mean parameter (Normand, 1999).

### 4.2.2 A Small Example - Robust Model-based Meta-analysis Methods

The conventional model-based meta-analysis assumes that all studies involved in the meta-analysis have the same underlying true parameter or hyper-parameter value. Such assumption is vulnerable to the unaware inclusion of outlying studies. For example, suppose a transcription error causes the reported summary mean statistics `y=c(-2,0,1,2)` to be recorded as `y.cntm=(-2,0,1,20)`, e.g., a 10 times larger mean value for the 4th study. In this case, the conventional model-based meta-analysis results are significantly impacted by this outlying study. For example, the fixed-effect model obtains a combined mean value 4.75, compared to 0.25 based on original data set. The random-effects model obtains similar results.

```
> y.cntm <- c(-2,0,1,20)
> # contaminated data set
> gdf.cntm <- data.frame(mns=y.cntm, sds=s)
> rownames(gdf.cntm) <- c('study-1','study-2','study-3','study-4(outlying)')
```

```
>
> # conventional fixed-effect meta-analysis
> gmt.cntm.fix <- gmeta(gmi=gdf.cntm, method='fixed-mle',
+ gmo.xgrid=seq(-10,30,by=0.001))
> summary(gmt.cntm.fix)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = gdf.cntm, method = "fixed-mle", gmo.xgrid = seq(-10,
    30, by = 0.001))

Summary of Combined CD:
                mean median    stddev ci.lower ci.upper
Combined CD 4.749863   4.75 0.4999998 3.770018 5.729784

Confidence level = 0.95

Summary of Individual CDs:
                  mean median stddev  ci.lower    ci.upper
study-1             -2     -2      1 -3.959964 -0.04003598
study-2              0      0      1 -1.959964  1.95996402
study-3              1      1      1 -0.959964  2.95996402
study-4(outlying)   20     20      1 18.040036 21.95996402

Confidence level = 0.95
>
```

```
> # conventional random-effects meta-analysis
> gmt.cntm.rdm <- gmeta(gmi=gdf.cntm, method='random-mm',
+ gmo.xgrid=seq(-10,30,by=0.001), tau2=gmt.rdm$tau2)
> summary(gmt.cntm.rdm)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = gdf.cntm, method = "random-mm", gmo.xgrid = seq(-10,
    30, by = 0.001), tau2 = gmt.rdm$tau2)

Summary of Combined CD:
            mean median    stddev ci.lower ci.upper
Combined CD 4.75   4.75 0.8539126 3.076362 6.423638

Confidence level = 0.95

Summary of Individual CDs:
                  mean median stddev  ci.lower  ci.upper
study-1             -2     -2      1 -5.347276  1.347276
study-2              0      0      1 -3.347276  3.347276
study-3              1      1      1 -2.347276  4.347276
study-4(outlying)   20     20      1 16.652724 23.347276

Confidence level = 0.95
```

Here, we suppose the heterogeneity `tau2` for the random-effects is known as $\tau^2 = 1.92$, the value estimated through original data set.

```
> # tau2 - the heterogeneity parameter in random-effects model
> gmt.rdm$tau2
[1] 1.916667
```

The robust meta-analysis methods, on the other hand, can limit the impact of outlying study, and provide protection on failing to consider the outlying studies in the model assumptions. As shown in the following script, the method `fixed-robust1` provides pooled mean estimate 0.64 with standard deviation 0.70 using the original data set, and pooled mean estimate 0.32 with standard deviation 0.81 using the contaminated data set. The pooled estimate based on the contaminated data set is even smaller, because the 4th study in the contaminated data set is further away from the majorities, and as a result, it is further down-weighted during the combining step (cf., Section 4 of Xie et al., 2011).

```
> # robust meta-analysis method
> # on original data set
> gmt.rbst <- gmeta(gmi=gdf, method='fixed-robust1',
+ gmo.xgrid=seq(-10,30,by=0.001), tau2=gmt.rdm$tau2)
> summary(gmt.rbst)
                Model-Based Meta-Analysis through CD-Framework


Call:
gmeta.default(gmi = gdf, method = "fixed-robust1", gmo.xgrid = seq(-10,
    30, by = 0.001), tau2 = gmt.rdm$tau2)
```

```
Summary of Combined CD:
                 mean    median    stddev   ci.lower ci.upper
Combined CD 0.6409065 0.6409065 0.4117606 -0.1661293 1.447942

Confidence level = 0.95

Summary of Individual CDs:
        mean median stddev    ci.lower    ci.upper
study-1   -2     -2      1 -3.95996402 -0.04003598
study-2    0      0      1 -1.95996402  1.95996402
study-3    1      1      1 -0.95996402  2.95996402
study-4    2      2      1  0.04003598  3.95996402

Confidence level = 0.95
>
> # on contaminated data set
> gmt.cntm.rbst <- gmeta(gmi=gdf.cntm, method='fixed-robust1',
+ gmo.xgrid=seq(-10,30,by=0.001), tau2=gmt.rdm$tau2)
> summary(gmt.cntm.rbst)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = gdf.cntm, method = "fixed-robust1", gmo.xgrid = seq(-10,
    30, by = 0.001), tau2 = gmt.rdm$tau2)

Summary of Combined CD:
```

```
                mean    median    stddev   ci.lower ci.upper
Combined CD 0.3215122 0.3215122 0.4760393 -0.6115077 1.254532

Confidence level = 0.95

Summary of Individual CDs:
                  mean median stddev  ci.lower    ci.upper
study-1             -2     -2      1 -3.959964 -0.04003598
study-2              0      0      1 -1.959964  1.95996402
study-3              1      1      1 -0.959964  2.95996402
study-4(outlying)   20     20      1 18.040036 21.95996402

Confidence level = 0.95
```

To understand the robust meta-analysis methods and other advanced features, it is preferred to have a working knowledge on the concept of confidence distributions (CDs). We present a brief review of using CD for meta-analysis in Section 4.3. For the practical use of the **gmeta** package, it can be skipped and referred back when needed while going through the examples in Section 4.4.

## 4.3 Confidence Distribution

### 4.3.1 Confidence Distribution

A CD uses a sample-dependent distribution function to estimate the unknown parameter. Loosely speaking, it is a distribution function on the parameter space that can represent

all level confidence intervals for the parameter of interest. It contains more information comparing to the point or interval estimators, and is an ideal candidate as information carrier for combination. The following formal definition is proposed by Schweder and Hjort (2002) and Singh et al. (2005, 2007):

**Definition 4.1** *Suppose $\Theta$ is the parameter space of the unknown parameter of interest $\theta$, and $\mathcal{Y}$ is the sample space corresponding to data $\mathbf{Y} = \{y_1, \ldots, y_n\}$. Then a function $H(\cdot) = H(\mathbf{Y}, \cdot)$ on $\mathcal{Y} \times \Theta \to [0, 1]$ is a confidence distribution (CD) if:*
*(i) For each given $\mathbf{Y} \in \mathcal{Y}$, $H(\cdot)$ is a continuous cumulative distribution function on $\Theta$; and*
*(ii) $Pr\{\theta_0 \leq H^{-1}(\alpha)\} = \alpha, \forall 0 \leq \alpha \leq 1$, where $\theta_0$ is the true parameter value.*
*The function $H(\cdot)$ is an asymptotic CD (aCD) if $\lim_{n\to\infty} Pr\{\theta_0 \leq H^{-1}(\alpha)\} \to \alpha, \forall 0 \leq \alpha \leq 1$, and the continuity requirement on $H(\cdot)$ is dropped.*

The second condition facilitates desirable frequentist properties, such as unbiasedness, consistency and/or efficiency, of the estimates derived from $H(\cdot)$. We show through the following examples that the CD concept covers from regular parametric cases to $p$-value functions, see Singh et al. (2005) and Xie and Singh (2013) for more examples on normalized likelihood functions, bootstrap distributions and Bayesian posteriors, among others.

**Example 4.1** *(Normal mean and variance) Suppose a normal sample $X_i \sim \mathcal{N}(\mu, \sigma^2), i = 1, \ldots, n$ is observed and $\sigma$ is known, then the function $\psi(\mathbf{X}, \mu) = \dfrac{\mu - \overline{X}}{\sigma/\sqrt{n}}$ is monotonically increasing in $\mu$ and has distribution $\mathcal{N}(0, 1)$. Thus, the function*

$$H_\Phi(\mu) = \Phi\left(\frac{\mu - \overline{X}}{\sigma/\sqrt{n}}\right) \tag{4.1}$$

*satisfies the two conditions in Definition 4.1 and thus is a CD for $\mu$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. If the parameter $\sigma$ is*

*not known, then $H_\Phi(\mu)$ is no longer a sample dependent function on the space of $\mu$ since it also involves $\sigma$, and thus it is not a CD for $\mu$. In this case, consider the t-pivotal quantity $T(\mathbf{X}, \mu) = \dfrac{\mu - \overline{X}}{s/\sqrt{n}}$, it has a Student t-distribution with df $n-1$, and is asymptotic normal as $n \to \infty$. Thus, the functions*

$$H_T(\mu) = F_{T_{n-1}}\left(\frac{\mu - \overline{X}}{s/\sqrt{n}}\right) \quad \textit{and} \quad H_A(\mu) = \Phi\left(\frac{\mu - \overline{X}}{s/\sqrt{n}}\right) \tag{4.2}$$

*are CD and aCD for the parameter $\mu$, respectively. Here, $s$ is the sample standard deviation and $F_{T_{n-1}}(\cdot)$ is the cumulative distribution function of the t-distribution with df $n-1$.*

*For the parameter $\sigma^2$, the sample-dependent cumulative distribution function*

$$H_{\chi^2}(\theta) = 1 - F_{\chi^2_{n-1}}((n-1)s^2/\theta) \tag{4.3}$$

*is a CD for $\sigma^2$, where $F_{\chi^2_{n-1}}(\cdot)$ is the cumulative distribution function of the $\chi^2_{n-1}$-distribution. Here, we take $H_{\chi^2}(\cdot) = 1 - F_{\chi^2_{n-1}}(\cdot)$ because the quantity $\psi(\mathbf{X}, \sigma^2) = \dfrac{(n-1)s^2}{\sigma^2}$ is monotonic decreasing in $\sigma^2$.*

Example 4.1 illustrates using pivot statistics to construct confidence distribution. Given an observed sample, a pivot statistic is a function of observations and unobservable parameters, whose probability distribution does not depend on any unknown parameter. For example, the $z$-statistic and $t$-statistic used in (4.1) and (4.2). In general, suppose $\mathbf{x}$ is an observed sample, $\theta$ is the parameter of interest, $\psi(\mathbf{x}, \theta)$ is a (an asymptotic) pivot statistic that is (asymptotically) distributed as $F(\cdot)$. Without loss of generality, we suppose $F(\cdot)$ is monotonic non-decreasing in $\theta$. Then, a CD (an aCD) function for $\theta$ can be constructed by $H(\theta) = F(\psi(\mathbf{x}, \theta))$.

The CD is loaded with information in the sense that various inferences for the parameter of interest can be drawn from it, including but not limit to point estimates, confidence intervals, and hypothesis testing. For example, Figure 4.1 part (a) plots the CD for the mean parameter $\mu$ based on formula (4.1) in Example 4.1. It shows that the mean/median of the distribution function $H(\mu)$ can be used as a point estimator for $\mu$, the interval $(H^{-1}(\alpha_1), H^{-1}(1-\alpha_2))$ forms a level $100(1-\alpha_1-\alpha_2)\%$ confidence interval (CI) for any given $0 \le \alpha_1 \le 1-\alpha_2 \le 1$, and the $p$-value for the one-sided hypothesis testing problem $K_0 : \mu \ge \mu_0$ versus $K_a : \mu < \mu_0$ can be obtained through the tail mass $1 - H(\mu_0)$. Figure 4.1 also plots the confidence density $h(\mu) = \partial H(\mu)/\partial\mu$, and the confidence curve $\mathrm{CCV}(\mu) = 1 - 2|H(\mu) - 0.5| = 2\min\{H(\mu), 1 - H(\mu)\}$, on part (b) and (c), respectively. The confidence density and confidence curve are useful in the theoretical development, for example, Blaker (2000) relies the confidence curve to derive improved exact confidence intervals for a general discrete distribution (cf. Section 7 of Xie and Singh, 2013).

An extension of the Example 4.1 is dropping the normality assumption on the sample $\mathbf{X}$ and only assuming $\mathbf{X}$ comes from a population with finite mean $\mu$ and variance $\sigma^2$.

**Example 4.2** *(aCD based on Central Limit Theorem (CLT)) Suppose $X_1, \ldots, X_n$ are i.i.d. samples from a distribution with finite first and second cental moments, $\mu$ and $\sigma^2$, respectively, where $\mu$ is the parameter of interest and $\sigma^2$ is known. Let $\overline{X} = \sum X_i/n$, then according to CLT, $\sqrt{n}(\mu - \overline{X}) \to \mathcal{N}(0, \sigma^2)$ as $n \to \infty$. Therefore, $\frac{\mu - \overline{X}}{\sigma/\sqrt{n}}$ is an asymptotic pivot statistic. As a result, $\Phi\left(\frac{\mu - \overline{X}}{\sigma/\sqrt{n}}\right)$ is an aCD function.*

As an application of Example 4.1 and 4.2, the following Example 4.3 illustrates how to build aCD for log odds ratio (LOR) of 2x2 table. In clinical trials, the outcomes are often dichotomous, e.g., observed an event or not. A 2x2 table, as shown in Table 4.1, is commonly used for summarizing the study outcomes.

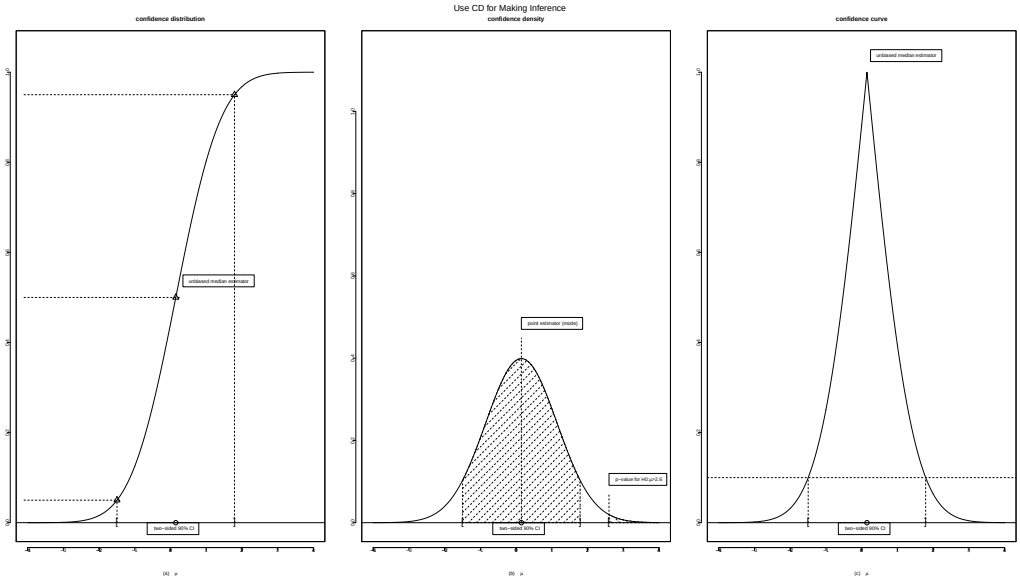


Figure 4.1: An example of using CD for making inference. The solid curves in figure (a), (b) and (c) are confidence distribution, confidence density, and confidence curve, respectively, for the parameter $\mu$ based on a sample generated from $x_i \sim$ i.i.d $\mathcal{N}(\mu(\text{true}) = 0, \sigma^2(\text{known}) = 1)$. Here, the confidence distribution is obtained by: $H_\Phi(\mu) = \Phi(\sqrt{n}(\mu - \bar{x})/\sigma)$. The confidence density is derived by taking the first derivative w.r.t $\mu$ in $H(\mu)$, and the confidence curve is defined as $\text{CCV}(\mu) = 1 - 2|H(\mu) - 0.5|$. The figure also illustrates the procedure to obtain point estimates, confidence intervals, and $p$-values for hypothesis testing problems through the confidence distribution, confidence density, or confidence curve.

Table 4.1: 2x2 Table From One Study

| | Events | Non-Events | Total |
|---|---|---|---|
| Treatment(Drugs) | $X_i$ | $n_i - X_i$ | $n_i$ |
| Control(Placebo) | $Y_i$ | $m_i - Y_i$ | $m_i$ |
| Total | $t_i$ | $N_i - t_i$ | $N_i$ |

**Example 4.3** *(aCD for Log Odds Ratio (LOR) of 2x2 Tables) Suppose that a clinical trial is designed to investigate the relative performance of a novel treatment (drugs) versus the current standard medical practice (control or placebo), and the dichotomous (e.g., stroke*

*or not) outcomes are organized as in Table 4.1. The study sample size is large, and thus no entries in Table 4.1 is zero. The estimated LOR and its standard deviation based on the Table 4.1 are*

$$\widehat{\theta}_i = \log\left(\frac{X_i/(n_i - X_i)}{Y_i/(m_i - Y_i)}\right) \quad \textit{and} \quad s_i = \sqrt{\frac{1}{X_i} + \frac{1}{n_i - X_i} + \frac{1}{Y_i} + \frac{1}{m_i - Y_i}} \tag{4.4}$$

*It is known that the sample distribution of $\widehat{\theta}_i$ converges to $\mathcal{N}(\theta_i, s_i^2)$ quickly as the sample size $N_i \to \infty$, where $\theta_i$ is the underlying true value of the LOR. Thus, asymptotically the statistic $(\theta_i - \widehat{\theta}_i)/s_i \sim \mathcal{N}(0,1)$ and therefore an aCD for $\theta_i$ can be constructed by*

$$H_{asymp}(\theta_i) = \Phi\left((\theta_i - \widehat{\theta}_i)/s_i\right) \tag{4.5}$$

An advanced method of constructing CD for LOR, without asymptotic normality, is through the instrumentality of significance functions or $p$-value functions (Fraser, 1991). Let us consider the hypothesis testing problem $K_0 : \theta \leq t_0$ versus $K_a : \theta > t_0$. The $p$-values, $p(\cdot)$, can be interpreted as a function of $t_0$, i.e., $p = p(t_0)$ while $t_0$ varies over the parameter space. This $p(t_0)$ is known as significance function or $p$-value function (Fraser, 1991), and this significance function $p(\cdot) = p(\cdot, \mathbf{x})$ almost always satisfies the definition of CD (Singh et al., 2005, 2007).

**Example 4.4** *(Exact CD for Log Odds Ratio (LOR) of 2x2 Tables based on p-value functions derived by Fisher's Exact Test) Follow the Example 4.3, but now suppose that the sample size is small. In this case, the large sample theory such as asymptotic normality is not suitable, and thus the aCD in (4.5) is not suitable for making inference on $\theta_i$. Further, if the event rate is small enough such that any entry in Table 4.1 is zero, then neither $\widehat{\theta}_i$ nor $s_i$ is meaningful unless adjusting the zeros by artificially continuity corrections. For*

*example, Table 4.2.*

Table 4.2: An Observed 2x2 Table From a Small Study

| | Events | Non-Events | Total |
|---|---|---|---|
| Treatment | $4(X_i)$ | $3(n_i - X_i)$ | $7(n_i)$ |
| Control | $4(Y_i)$ | $0(m_i - Y_i)$ | $4(m_i)$ |
| Total | $8(t_i)$ | $3(N_i - t_i)$ | $11(N_i)$ |

*Different from Example 4.3, let us consider the hypothesis testing problem:*

$$K_0 : \Psi_i = \Psi^* \quad vs. \quad K_a : \Psi_i > \Psi^* \tag{4.6}$$

*with respect to odds ratio $\Psi_i$, where $\Psi^*$ is an arbitrary but fixed value on the parameter space. The p-value function $p_i(\Psi)$ based on the mid-p adaptation of Fisher's Exact Test (Fisher, 1922) is*

$$p_i(\Psi) \equiv p_i(\Psi; x_i, y_i) = Pr_\Psi(X_i > x_i | T_i = t_i) + \frac{1}{2} Pr_\Psi(X_i = x_i | T_i = t_i) \tag{4.7}$$

*where $T_i$ is defined as $T_i = X_i + Y_i$, and the distribution of $X_i$ conditional on $T_i = t_i$ follows noncentral hypergeometric distribution:*

$$Pr_\Psi(X_i = x | T_i = t_i) = \frac{\binom{n_i}{x}\binom{m_i}{t_i - x}\Psi^x}{\sum_{s=L_i}^{U_i}\binom{n_i}{s}\binom{m_i}{t_i - s}\Psi^s}, \quad L_i \leqslant x_i \leqslant U_i \tag{4.8}$$

*where $L_i = \max(0, t_i - m_i), U_i = \min(n_i, t_i)$. Then, the p-value function $H_\Psi(\Psi) = p_i(\Psi)$ is an exact CD function for $\Psi_i$, and thus the CD function for the LOR, $\theta_i$, is simply $H_{exact}(\theta) = H_\Psi(\exp(\theta))$ (Liu et al., 2013).*

*In particular, the exact CD function $H_{exact}(\cdot)$ is able to make correct inference for studies*

*contain zero event, whereas the aCD $H_{asymp}(\cdot)$ has to be constructed with artificial continuity corrections and thus is biased (J Sweeting et al., 2004). Further, the exact CD is more efficient than the aCD when the study sample size is small (Liu et al., 2013).*

The concept of CD is broad. Besides pivotal quantities' distributions and $p$-value functions, it also covers bootstrap distributions, normalized likelihood functions, Bayesian posteriors, among others (Singh et al., 2007; Xie and Singh, 2013). In fact, any sample dependent distribution function "can" be used as a CD, just as any sample dependent scalar value "can" be used as a point estimator. However, CD constructed following the requirements in Definition 4.1 can ensure the estimates derived from such CD have desirable properties. For example, the unbiasedness, consistency, or efficiency of the point estimates based on the mean or median of CD.

### 4.3.2 Combining CDs and a Unified Meta-analysis Approach

Assume $k$ independent studies are available, all regarding the same parameter $\theta$. Based on the methods introduced in Example 4.1- 4.4, we can construct CDs $H_i(\theta) = H_i(X_i, \theta)$, $i = 1, \ldots, k$ for $\theta$ based on sample $X_i$ from each individual study. To combine individual CDs and make an overall inference on $\theta$, a general combining method is proposed by Singh et al. (2005). The key is choosing a coordinate-wise monotonic function $g^{(c)}(u_1, \ldots, u_k)$ that maps the $k$-dimensional cube $[0, 1]^k$ to real line. In particular, the combined CD can be built through:

$$H^{(c)}(\theta) = G^{(c)}\{g^{(c)}(H_1(\theta), \ldots, H_k(\theta))\} \tag{4.9}$$

where $G^{(c)}(t) = \Pr\{g^{(c)}(U_1, \ldots, U_k) \leq t\}$ is the cumulative distribution function of $g^{(c)}(U_1, \cdots, U_k)$. Here, $U_1, \ldots, U_k$ are independent $\mathbf{U}[0, 1]$ random variables. The combined function $H^{(c)}(\cdot)$ is a valid CD for $\theta$ as long as the inputs $H_i(\theta)$ share the same

underlying true parameter $\theta$ and independent to each other. The methods used to obtain individual CDs $H_i(\theta)$ are irrelevant.

To apply this general combining method to meta-analysis, a special family of $g^{(c)}(\cdot)$ is proposed by Xie et al. (2011):

$$g^{(c)}(u_1, \ldots, u_k) = w_1 a_0(u_1) + \cdots + w_k a_0(u_k) \tag{4.10}$$

where $a_0(\cdot)$ is a given continuous and monotonic (without loss of generality, say increasing) function and $w_i \geq 0$, with at least one $w_i \neq 0$, are generic weights for the combination. Though other choices may possible, the **gmeta** implements the general combining method with $g^{(c)}(\cdot)$ in formula (4.10). It indeed unifies almost all methods currently used in meta-analysis, including $p$-value combination methods, fixed-effect and random-effects model-based meta-analysis methods, the Mantel-Haenszel method, the Peto's method, the method proposed in Tian et al. (2009) by combining confidence intervals, and also the method proposed in Liu et al. (2013) by combining significance functions (see Xie et al., 2011, and subsequent research).

The following example shows that the CD combining method (4.9), (4.10) covers model-based meta-analysis as a special case by taking $a_0(\cdot) = \Phi^{-1}(\cdot)$ in (4.10).

**Example 4.5** *(aCD for Log odds ratio (LOR) of 2x2 Table - continued) Suppose k clinical trials are performed for comparing the relative performance of the same novel treatment versus control. In Example 4.3, we built a CD $H_i(\theta_i)$ for the study-specific LOR, $\theta_i$, based on the 2x2 table from the i-th study, and we want to combine them together for estimating the overall LOR, $\theta$. Follow Normand (1999), we assume a random-effects model*

*$\theta_i \sim \mathcal{N}(\theta, \tau^2)$ and $\tau^2$ is known. Thus, we have*

$$\widehat{\theta}_i|(\theta_i, s_i) \sim \mathcal{N}(\theta_i, s_i^2), \quad \theta_i|(\theta, \tau) \sim \mathcal{N}(\theta, \tau^2), \quad i = 1, \ldots, k \tag{4.11}$$

*Therefore, asymptotically $\widehat{\theta}_i \sim \mathcal{N}(\theta, s_i^2 + \tau^2)$, and an aCD for $\theta$ based on the $i$-th study can be found by*

$$H_i(\theta) = \Phi\left((\theta - \widehat{\theta}_i)/v_i\right) \tag{4.12}$$

*where $v_i^2 = s_i^2 + \tau^2$. Take $a_0(\cdot) = \Phi^{-1}(\cdot)$ and $w_i = v_i^{-1}$, and apply the general combining method (4.9) with $g^{(c)}(\cdot)$ given by (4.10), we find that*

$$g^{(c)}(H_1(\theta), \ldots, H_k(\theta)) = \sum_{i=1}^{k} \frac{1}{v_i} \Phi^{-1}\left(\Phi\left(\frac{\theta - \widehat{\theta}_i}{v_i}\right)\right) = \left(\sum_{i=1}^{k} \frac{1}{v_i^2}\right)\left(\theta - \frac{\sum_{i=1}^{k} \widehat{\theta}_i/v_i^2}{\sum_{i=1}^{k} 1/v_i^2}\right)$$

*Since $g^{(c)}(U_1, \ldots, U_k) = \sum_{i=1}^{k} w_i \Phi^{-1}(U_i) \sim \mathcal{N}(0, \sum_{i=1}^{k} w_i^2)$, which implies $G^{(c)}(t) = \Phi\left(t/\sqrt{\sum_{i=1}^{k} w_i^2}\right)$. Thus, the combined CD can be constructed by*

$$\begin{aligned} H^{(c)}(\theta) &= G^{(c)}\{g^{(c)}(H_1(\theta), \ldots, H_k(\theta))\} = \Phi\left(\frac{g^{(c)}(H_1(\theta), \ldots, H_k(\theta))}{\sqrt{\sum_{i=1}^{k} w_i^2}}\right) \\ &= \Phi\left(\left(\sum_{i=1}^{k} \frac{1}{v_i^2}\right)^{1/2}\left(\theta - \frac{\sum_{i=1}^{k} \widehat{\theta}_i/v_i^2}{\sum_{i=1}^{k} 1/v_i^2}\right)\right) \end{aligned} \tag{4.13}$$

*which is normally distributed with mean $\widehat{\theta}_c = \frac{\sum_{i=1}^{k} \widehat{\theta}_i/v_i^2}{\sum_{i=1}^{k} 1/v_i^2}$ and variance $v_c^2 = \frac{1}{\sum_{i=1}^{k} 1/v_i^2}$.*

*If the parameter $\tau^2$ is unknown, then we can replace it with the sample estimates, say the Dersimonian-Laird moment estimator $\widehat{\tau}^2_{DL}$ (DerSimonian and Laird, 1986), or the restricted-maximum-likelihood estimator $\widehat{\tau}^2_{REML}$. As long as it is a consistent estimator for*

*$\tau^2$, the combined function (4.13) is still an aCD for $\theta$.*

*On the other hand, the conventional random-effects meta-analysis method assumes the same model (4.11) (Normand, 1999). As a result, the DerSimonian and Laird estimator for $\theta$ is $\widehat{\theta}_{DL} = \frac{\sum_1^k \widehat{\theta}_i/(s_i^2+\widehat{\tau}^2_{DL})}{\sum_1^k 1/(s_i^2+\widehat{\tau}^2_{DL})}$ with variance $v^2_{DL} = \frac{1}{\sum_{i=1}^k 1/(s_i^2+\widehat{\tau}^2_{DL})}$. Likewise, the REML estimator for $\theta$ is $\widehat{\theta}_{REML} = \frac{\sum_1^k \widehat{\theta}_i/(s_i^2+\widehat{\tau}^2_{REML})}{\sum_1^k 1/(s_i^2+\widehat{\tau}^2_{REML})}$ with variance $v^2_{REML} = \frac{1}{\sum_{i=1}^k 1/(s_i^2+\widehat{\tau}^2_{REML})}$ (see Table IV of Normand, 1999). The point estimates derived by CD, $\widehat{\theta}_c$, by replacing $\tau^2$ with $\widehat{\tau}^2_{DL}$ and $\widehat{\tau}^2_{REML}$, matches $\widehat{\theta}_{DL}$ and $\widehat{\theta}_{REML}$, respectively. Therefore, the CD-based meta-analysis method includes the conventional random-effects model-based meta-analysis as a special case by taking $a_0(\cdot) = \Phi^{-1}(\cdot)$. Similar results hold for fixed-effect meta-analysis method, which is equivalent to assume $\tau^2 = 0$ (see Section 3 of Xie et al., 2011).*

The next section shows that different choices on `linkfunc`, $a_0(\cdot)$, lead to different combined CDs, with the derived point estimators having different properties, e.g., efficiency or robustness. Specifically, it uses real data examples to show the commands for performing conventional, robust and exact meta-analysis, along with the methods of constructing corresponding combined CDs.

## 4.4 Examples

We provide several examples of using functions in the **gmeta** package for meta-analysis. The **gmeta** package can be installed as any other R package through the `install.packages()` command. There are three key functions in the package for our purpose, the `gmeta()` function is responsible for performing meta-analysis, the `summary()` function summarizes the meta-analysis results, and the `plot()` function displays the results through extended forest plots. Hereafter, we assume that the **gmeta** package is installed and loaded in the

current R session.

```
# install and load package
> install.packages('gmeta_2.2-3.zip', repos = NULL)
> library("gmeta")        # load the gmeta package
> ls("package:gmeta", all = TRUE) # list functions included in gmeta package
> data(package = "gmeta")$results # list data sets included in gmeta package
```

The `ulcer` data is included in the **gmeta** package and used throughout the following examples. It gathers 41 randomized clinical trials conducted between 1980 and 1989, all compared the performance of a novel surgical treatment versus the old treatment (control), in terms of reducing the occurrence of an adverse event – recurrent bleeding (Sacks et al., 1990; Efron, 1996). The data is organized in an R `data.frame`, where each row contains the number of event and non-event from novel treatment group followed by the number of event and non-event from control group. In fact, each row represents a 2x2 table as in Table 4.1 in order of $(X_i, n_i - X_i, Y_i, m_i - Y_i)$. The parameter of interest is the overall log odds ratio (LOR), $\theta$, in favor of the new treatment. Thus, our objective is making inference on $\theta$ by summarizing all evidence from these 41 clinical trials.

```
> data(ulcer)      # load the ulcer dataset
> ulcer
   TrtEvent TrtNonevent CtrlEvent CtrlNonevent
1         7           8        11            2
[...]
41        0           9         0           16
```

### 4.4.1 Classical $p$-value Combination

For clinical trials compared new and old treatments, literatures often report the $p$-value of hypothesis testing problem $K_0 : \theta \geq 0$ vs $K_a : \theta < 0$, where $\theta$ is the log odds ratio in favor of the new treatment. Thus, we can obtain an overall $p$-value by combining individual $p$-values from each individual studies. The classical $p$-values combination methods, including Fisher method (Fisher, 1932), Stouffer (Normal) method (Stouffer et al., 1949), Tippett (Min) method, Max method, and Sum method, are summarized in Marden (1991).

For the test $K_0 : \theta \geq t_0$ vs $K_a : \theta < t_0$, the $p$-value, $p(\cdot)$, can be viewed as a function of $t_0$, i.e, $p(\cdot) = p(t_0)$ as $t_0$ varies over the parameter space. This $p$-value function is in general a CD or an aCD function, and thus the classical $p$-value combination is generally a special case of CD-based meta-analysis. For example, Fisher method suggest to use

$$p^{(c)} = \Pr\left\{\chi^2_{2k} \geq -2\sum_{i=1}^{k}\log(p_i)\right\} \tag{4.14}$$

as the combined $p$-value from all $k$ studies, and this combined $p$-value is Bahadur optimal (Littell and Folks, 1973). From the CD combining framework, suppose $p_i(s)$ are the $p$-value functions from $i$-th study, $i = 1, \ldots, k$, for the hypothesis testing $K_0 : \theta \geq s$ vs $K_a : \theta < s$. Take $a_0(\cdot) = \log(\cdot)$ and $w_i = 1$ in (4.10), then the combined CD following the general method (4.9) is

$$H^{(c)}(s) = \Pr\left\{\chi^2_{2k} \geq -2\sum_{i=1}^{k}\log(p_i(s))\right\} \tag{4.15}$$

It is obvious that $p^{(c)} = H^{(c)}(t_0)$, and thus Fisher method is a special case of CD-based meta-analysis method.

Table 4.3 lists the $p$-value combination methods unified under the CD combining framework, along with the choice of $g^{(c)}(\cdot)$ and the combined CD $H^{(c)}(\cdot)$. Here, the $\mathcal{C}_k$ is a random variable distributed as the sum of $k$ independent $\mathbf{U}[0,1]$ random variables, such distribution is given by Potuschak and Müller (2009).

Table 4.3: List of the $p$-value combination methods unified under the CD combining framework

| | |
|---|---|
| Fisher method | |
| classical $p$-value combination method | $p^{(c)} = \Pr\left\{\chi^2_{2k} \geq -2\sum_{i=1}^{k}\log(p_i)\right\}$ |
| choice of $g^{(c)}(\cdot)$ | $g^{(c)}(u_1,\ldots,u_k) = \log(u_1) + \ldots + \log(u_k)$ |
| $H^{(c)}(\cdot)$ | $H^{(c)}(s) = \Pr\left\{\chi^2_{2k} \geq -2\sum_{i=1}^{k}\log(p_i(s))\right\}$ |
| Stouffer (Noraml) method | |
| classical $p$-value combination method | $p^{(c)} = \Phi(1/\sqrt{k}[\Phi^{-1}(p_1) + \ldots + \Phi^{-1}(p_k)])$ |
| choice of $g^{(c)}(\cdot)$ | $g^{(c)}(u_1,\ldots,u_k) = \Phi^{-1}(u_1) + \ldots + \Phi^{-1}(u_k)$ |
| $H^{(c)}(\cdot)$ | $H^{(c)}(s) = \Phi(1/\sqrt{k}[\Phi^{-1}(p_1(s)) + \ldots + \Phi^{-1}(p_k(s))])$ |
| Tippett (Min) method | |
| classical $p$-value combination method | $p^{(c)} = 1 - (1 - \min(p_1,\ldots,p_k))^k$ |
| choice of $g^{(c)}(\cdot)$ | $g^{(c)}(u_1,\ldots,u_k) = \min(u_1,\ldots,u_k)$ |
| $H^{(c)}(\cdot)$ | $H^{(c)}(s) = 1 - (1 - \min(p_1(s),\ldots,p_k(s)))^k$ |
| Max method | |
| classical $p$-value combination method | $p^{(c)} = \max(p_1,\ldots,p_k)^k$ |
| choice of $g^{(c)}(\cdot)$ | $g^{(c)}(u_1,\ldots,u_k) = \max(u_1,\ldots,u_k)$ |
| $H^{(c)}(\cdot)$ | $H^{(c)}(s) = \max(p_1(s),\ldots,p_k(s))^k$ |
| Sum method | |
| classical $p$-value combination method | $p^{(c)} = \Pr\left\{\mathcal{C}_k \leq \sum_{i=1}^{k} p_i\right\}$ |
| choice of $g^{(c)}(\cdot)$ | $g^{(c)}(u_1,\ldots,u_k) = u_1 + \ldots + u_k$ |
| $H^{(c)}(\cdot)$ | $H^{(c)}(s) = \Pr\left\{\mathcal{C}_k \leq \sum_{i=1}^{k} p_i(s)\right\}$ |

The function `gmeta()` can perform all these $p$-value combinations by choosing corresponding methods. In particular, the input `gmi` is a vector of $p$-values from individual studies for the same hypothesis, and correspondingly `gmi.type='p-value'`. The combination `method` can be chosen from `'fisher'`,`'stouffer'`,`'normal'`,`'tippett'`,`'min'`,`'max'`, and `'sum'`. Example 4.6 shows the results of using `gmeta()` function to combine $p$-values

for the hypothesis $K_0 : \theta \geq 0$ vs $K_a : \theta < 0$ for the ulcer data set.

**Example 4.6** *(p-value combination) To obtain meaningful $p_i$ from the i-th study, we add 0.5 to the twelve zero entries in the data frame, as suggested by Efron (1996) and others. The sample statistic $\widehat{\theta}_i$ is approximately normally distributed with mean $\theta$ and variance $s_i^2$, where the summary statistics $\widehat{\theta}_i$ and $s_i$ are given in Example 4.3. Thus p-values for the hypothesis: $K_0 : \theta \geq 0$ vs $K_a : \theta < 0$ based on each individual study can be calculated by:*

```
> # keep original data set in ulcer.o
> ulcer.o <- as.matrix(ulcer)
> # impute 0.5
> ulcer <- ifelse(ulcer.o == 0, 0.5, ulcer.o)
> # summary statistics
> ulcer.theta <- log( (ulcer[,1]*ulcer[,4]) / (ulcer[,2]*ulcer[,3]) )
> ulcer.sigma <- sqrt(1/ulcer[,1] + 1/ulcer[,2] + 1/ulcer[,3] + 1/ulcer[,4])
> # p-values from individual studies for K0: LOR >=0 vs. Ka: LOR < 0
> ulcer.pvalues <- 1 - pnorm(0, mean=ulcer.theta, sd=ulcer.sigma)
> ulcer.pvalues
 [1] 2.364514e-02 3.204119e-01 7.170408e-01 7.631117e-01 9.045478e-01
 [6] 1.435593e-01 2.786246e-02 2.963888e-04 2.499260e-01 1.713687e-03
[11] 3.850228e-02 3.061835e-02 8.296315e-01 8.454856e-01 3.066366e-02
[16] 1.417192e-01 6.571353e-03 7.575993e-02 3.877901e-03 1.602023e-02
[21] 2.165047e-04 8.145715e-01 6.764660e-01 2.568914e-01 8.483522e-03
[26] 1.027274e-01 4.056633e-01 1.735560e-03 1.929439e-02 1.113040e-02
[31] 8.729020e-03 5.737073e-02 1.024082e-01 5.638801e-02 7.908861e-01
[36] 5.038550e-03 6.858163e-01 9.294398e-02 1.960676e-02 1.402044e-05
```

```
[41] 6.108883e-01
```

*Apply the Stouffer (Normal) method to combine these p-values, we have:*

```
> # apply Stouffer (Normal) method
> gmo.pvl.normal = gmeta(ulcer.pvalues, gmi.type='pvalue', method='normal')
> # display results
> #gmo.pvl.normal
> #print(gmo.pvl.normal)
> summary(gmo.pvl.normal)
                P-value combination through CD-Framework

Call:
gmeta.default(gmi = ulcer.pvalues, gmi.type = "pvalue", method = "normal")

Combine Method:    normal

Combined p-value:  1.10779e-16

Individual p-values:
 [1] 2.364514e-02 3.204119e-01 7.170408e-01 7.631117e-01 9.045478e-01
 [6] 1.435593e-01 2.786246e-02 2.963888e-04 2.499260e-01 1.713687e-03
[11] 3.850228e-02 3.061835e-02 8.296315e-01 8.454856e-01 3.066366e-02
[16] 1.417192e-01 6.571353e-03 7.575993e-02 3.877901e-03 1.602023e-02
[21] 2.165047e-04 8.145715e-01 6.764660e-01 2.568914e-01 8.483522e-03
[26] 1.027274e-01 4.056633e-01 1.735560e-03 1.929439e-02 1.113040e-02
[31] 8.729020e-03 5.737073e-02 1.024082e-01 5.638801e-02 7.908861e-01
```

```
[36] 5.038550e-03 6.858163e-01 9.294398e-02 1.960676e-02 1.402044e-05
[41] 6.108883e-01
```

*The following code shows the results from all methods mentioned in Table 4.3.*

```
## table of the result
> pvalue.combine.methods = c('fisher', 'stouffer', 'tippet', 'max', 'sum')
> combined.pvalue.vector = rep(NA,length(pvalue.combine.methods))
> for ( i in 1:length(pvalue.combine.methods) ) {
+ combined.pvalue.vector[i] <- gmeta(ulcer.pvalues, gmi.type='pvalue',
+                          method=pvalue.combine.methods[i])$cmbd.pvalue
+ }
> mthds = 'method'
> pvlus = 'p-value'
> for ( i in 1:length(pvalue.combine.methods) ) {
+ mthds = paste(mthds, pvalue.combine.methods[i], sep='\t\t& ')
+ pvlus = paste(pvlus, combined.pvalue.vector[i], sep='\t& ')
> }
> cat('\np-value combination result\n', mthds, '\n', pvlus, '\n')

p-value combination result:
 method   & fisher     & stouffer   & tippet     & max        & sum
 p-value  & 2.1684e-19 & 1.1078e-16 & 5.7468e-04 & 1.6357e-02 & 4.8591e-09
```

The results show that the new treatment achieves significant reduction on the recurrence bleeding events, since the combined $p$-value is significant at 5% level, no matter what method used for the combination.

### 4.4.2 Conventional Fixed-effect and Random-effects Meta-analysis

Let us consider the estimation of the log odds ratio $\theta$ under the following meta-analysis models. In particular, the fixed-effect model:

$$\widehat{\theta}_i|(\theta, s_i) \sim \mathcal{N}(\theta, s_i^2), \quad i = 1, \ldots, k \tag{4.16}$$

And, the random-effects model:

$$\widehat{\theta}_i|(\theta_i, s_i) \sim \mathcal{N}(\theta_i, s_i^2), \quad \theta_i|(\theta, \tau) \sim \mathcal{N}(\theta, \tau^2), \quad i = 1, \ldots, k \tag{4.17}$$

where the fixed-effect model can be viewed as a special case of random-effects model by taking $\tau^2 = 0$. Example 4.5 shows that these model-based meta-analysis can be unified under the CD combining framework (see Section 3 of Xie et al., 2011, for more details).

The following code exemplifies of using `gmeta()` function to perform conventional model-based meta-analysis. The argument `method` specifies the model assumptions, e.g. `method='fixed-mle'` for using fixed-effect model, and `method='random-mm'` and `method='random-reml'` for using random-effects model with DL and REML estimator for estimating heterogeneity, respectively. The argument `linkfunc` specifies the function $a_0(\cdot)$. The default value is `linkfunc='inverse-normal-cdf'` for $a_0(\cdot) = \Phi^{-1}(\cdot)$. The other choice is `linkfunc='inverse-laplace-cdf'` for $a_0(\cdot) = \mathrm{DE}^{-1}(\cdot)$, where $\mathrm{DE}(t) = \frac{1}{2}\exp(t)\mathbf{I}\{t \leq 0\} + (1 - \frac{1}{2}\exp(-t))\mathbf{I}\{t > 0\}$. The argument `weight` specifies the study-specific weights. The default value is `NULL`, where the weights will be assigned depends on `linkfunc`. For example, if `linkfunc='inverse-normal-cdf'` then inverse standard deviation weights will be used for achieving Fisher efficiency, on the other hand, if `linkfunc='inverse-laplace-cdf'` then weights of all ones will be used for obtaining

Bahadur efficiency (Xie et al., 2011). Or else, the user can specify `weight` for each study using a numeric vector.

The `summary()` function associated with the `gmeta()` output object presents a summary of the results, including mean, median, standard deviation, lower and upper 95% confidence interval (CI) boundary points derived from each individual CDs and the combined CD function.

**Example 4.7** *(conventional fixed-effect meta-analysis) The simplest way to do the conventional meta-analysis is using the summary statistics, means and standard deviations. To save space, we delete the results of most individual CDs and focus on displaying the results of the combined CDs.*

```
> # use summary statistics
> ulcer.pdata <- cbind(ulcer.theta, ulcer.sigma)
> # default value linkfunc='inverse-normal-cdf' - Fisher efficiency
> gm1fp.mle <- gmeta(ulcer.pdata, method='fixed-mle',
+ gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1fp.mle)
               Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pdata, method = "fixed-mle", gmo.xgrid = seq(-20,
    20, by = 0.001))

Summary of Combined CD:
                 mean     median    stddev  ci.lower   ci.upper
```

```
Combined CD -0.8875844 -0.8875844 0.1255535 -1.133665 -0.6415031

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev    ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -3.6545713 -0.02198789
[...]
study-41  0.5753641  0.5753641 2.0429418  -3.4287283  4.57945657

Confidence level = 0.95
>
> # using DE link for combining with default weight - Bahadur efficiency
> gm1fp.mle.DE <- gmeta(ulcer.pdata, method='fixed-mle',
+ linkfunc='inverse-laplace-cdf', gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1fp.mle.DE)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pdata, method = "fixed-mle",
    linkfunc = "inverse-laplace-cdf", gmo.xgrid = seq(-20, 20, by = 0.001))

Summary of Combined CD:
                mean    median    stddev  ci.lower   ci.upper
Combined CD -1.063764 -1.064124 0.1061674 -1.271884 -0.8534409
```

```
Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev    ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -3.6545713 -0.02198789
[...]
study-41  0.5753641  0.5753641 2.0429418  -3.4287283  4.57945657

Confidence level = 0.95
```

*These results show that the new treatment is significantly better than the old treatment in reducing recurrent bleeding rates, and match those presented in Xie et al. (2011).*

**Example 4.8** *(conventional fixed-effect meta-analysis - cont.) In the previous examples, we add 0.5 to the twelve entries of zero, in order to obtain meaningful sample estimates* $\widehat{\theta}_i$*. Instead, we could exclude studies with zero event, if artificial continuity correction is not preferred. Specifically, we can extract the summary statistics of studies with none zero events and construct a data.frame* `ulcer.nzo`*.*

```
> # exclude studies with zero event
> nzoidx <- c(1:4,7:10,12:24,26:27,30:33,35:39)
> ulcer.nzo <- ulcer[nzoidx,]
```

*Specify the studies names through the* `rownames` *of the data frame, so that the indices in* `study-index` *are corresponding to the indices in the original data set.*

```
> # construct data.frame with summary statistics
```

```
> ulcer.nzo.pdata <-
+ data.frame(mn=ulcer.theta[nzoidx],sd=ulcer.sigma[nzoidx])
> # specify row name of the data.frame so that it matches original data set
> rownames(ulcer.nzo.pdata) <- paste('study-',
+ formatC(nzoidx,width=2,format='d',flag='0'), sep='')
```

*Use* ***gmeta()*** *function to perform the fixed-effect model-based meta-analysis.*

```
> gm1fp.nzo.mle <- gmeta(ulcer.nzo.pdata, method='fixed-mle',
+ gmo.xgrid=seq(-20,20,by=0.001))
```

*The results are similar as the previous results, where studies with zero event were included by continuity corrections. Here, the absolute value of estimated LOR is smaller,* $-0.7970$ *versus* $-0.8876$*, partially due to the exclusion of* ***study-40****, which is strongly in favor of the new treatment.*

```
> summary(gm1fp.nzo.mle)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.nzo.pdata, method = "fixed-mle",
    gmo.xgrid = seq(-20, 20, by = 0.001))

Summary of Combined CD:
                  mean     median    stddev ci.lower  ci.upper
Combined CD -0.7969914 -0.7969914 0.1287665 -1.04937 -0.544613
```

```
Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964 -3.6545713 -0.02198789
study-02 -0.3184537 -0.3184537 0.6825753 -1.6562771  1.01936968
study-03  0.4111958  0.4111958 0.7162780 -0.9926836  1.81507498
study-04  0.4881568  0.4881568 0.6814521 -0.8474651  1.82377861
study-07 -1.3457091 -1.3457091 0.7033884 -2.7243253  0.03290697
study-08 -4.1743873 -4.1743873 1.2152872 -6.5563067 -1.79246785
study-09 -0.5371429 -0.5371429 0.7960944 -2.0974595  1.02317352
[...]
study-39 -1.1050848 -1.1050848 0.5359444 -2.1555170 -0.05465263

Confidence level = 0.95
```

*Further, in this case, we can construct exact individual CDs using significance functions, as shown in Example 4.4. The* `gmeta()` *function can take (exact) CDs as input, instead of summary statistics, which implicitly assumes normality. Here, we must explicitly specify* `gmi.type='cd'` *since the default value is* `gmi.type='pivot'`*, i.e., the summary statistics, means and standard deviations. Also, we specify the study names by the argument* `study.names`*, so that they match with the original data set.*

```
> # use a list of (exact) CDs as input
> # construct CD for each individual study
> ulcer.nzo.cdata <- list()
```

```
> idx = 1
> for (i in nzoidx) {
+ # CD function based on Fisher Exact Test
+ cdi.exact.func <- function(theta) {
+ 1 - pFNCHypergeo(ulcer[i,1], ulcer[i,1]+ulcer[i,2],
+ ulcer[i,3]+ulcer[i,4], ulcer[i,1]+ulcer[i,3], theta) +
+ + dFNCHypergeo(ulcer[i,1], ulcer[i,1]+ulcer[i,2],
+ ulcer[i,3]+ulcer[i,4], ulcer[i,1]+ulcer[i,3], theta) / 2
+ }
+ # CD evaluated at gridding points
+ li <- ulcer.theta[i] - 4*ulcer.sigma[i]
+ ui <- ulcer.theta[i] + 4*ulcer.sigma[i]
+ xi = seq(from=li, to=ui, by=0.001)
+ cdi.exact <- sapply(exp(xi), cdi.exact.func)
+ # a list of CD and correspoding gridding points
+ ulcer.nzo.cdata[[idx]] = cbind(xi, cdi.exact)
+ idx = idx + 1
+ }
> # combine individual CDs to obtain a combined CD function
> gm1fc.nzo.mle <- gmeta(ulcer.nzo.cdata, gmi.type='cd', method='fixed-mle',
+ gmo.xgrid=seq(-10,10,by=0.001), study.names=paste('study-',
+ formatC(nzoidx,width=2,format='d',flag='0'), sep=''))
> # summary of the result (mean, median, stddev, lower95CI, upper95CI)
> summary(gm1fc.nzo.mle)
                Model-Based Meta-Analysis through CD-Framework
```

```
Call:
gmeta.default(gmi = ulcer.nzo.cdata, gmi.type = "cd", method = "fixed-mle",
    study.names = paste("study-", formatC(nzoidx, width = 2,
        format = "d", flag = "0"), sep = ""), gmo.xgrid = seq(-10,
        10, by = 0.001))

Summary of Combined CD:
                 mean     median    stddev  ci.lower   ci.upper
Combined CD -0.7897418 -0.7892671 0.1266928 -1.039529 -0.5426535

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower     ci.upper
study-01 -1.7860442 -1.7301555 0.9660527 -3.9254255 -0.009779846
study-02 -0.3097563 -0.3070509 0.6981395 -1.6941752  1.058931405
study-03  0.4070910  0.3989053 0.7393358 -1.0529698  1.918549161
study-04  0.4839813  0.4674239 0.7003865 -0.8668185  1.937976104
study-07 -1.3162494 -1.2943594 0.7204508 -2.8145085  0.048953619
study-08 -4.0217210 -3.8706791 1.2837491 -7.2912533 -1.839365752
study-09 -0.5292889 -0.5028864 0.8240183 -2.2653177  1.039309378
[...]
study-39 -1.0912408 -1.0808067 0.5439268 -2.1979966 -0.046737430

Confidence level = 0.95
```

*The* `plot` *function associated with the* `gmeta()` *output objects provides an extended forest plot, for graphically illustrating the individual and combined CDs. The argument* `studies` *specifies the indices of individual studies to show on the plot, e.g.,* `studies=[vector of indices of individual studies for plotting]`*. The default value is* `NULL`*, which shows only the combined CD. The argument* `plot.option` *specifies the form of CD for plotting, the default value is* `'confidence-density'`*, other available choices are* `'confidence-curve'` *or* `'cv'`*, and* `'confidence-distribution'` *or* `'cdf'`*. For each CD, the median is marked by open circle, and the lower and upper* 95%*CI bounds are marked by* `[` *and* `]`*, respectively, on the line below the CDs. The individual and combines CDs from summary statistics and exact CDs are shown in Figure 4.2.*

```
> # plot the individual CDs and the combined CD (extended forest plot)
> # compare the exact CDs and asymptotic normal CDs used for combination
> postscript(file='ulcer-plot1-extended-forest-plot.eps',
+ paper="special", width=14, height=11, horizontal=FALSE)
> par(mfrow=c(1,2), mar=c(2,2,2,2)+0.1, mgp=c(2,1,0), oma=c(3,0,4,0))
> # left sub-figure, CDs based on asymptotic normality.
> plot(gm1fp.nzo.mle, studies=c(5,6,8,16,17), xlim=c(-10,5))
> title("Extended Forest Plot - Confidence Densities
+ From Asymptotic Normality")
> # right sub-figure, exact CDs based on Fisher Exact Test.
> plot(gm1fc.nzo.mle, studies=c(5,6,8,16,17), xlim=c(-10,5),
+ plot.option='confidence-density') # default is 'confidence-density'.
> title("Extended Forest Plot - Confidence Densities
+ From Significance Functions")
> # overall title
```

```
> mtext("Compare Individual CDs Based on Asymptotic Normality
+ and Significance Functions for Fixed-effect Meta-analysis",
+ NORTH<-3, line=0, adj=0.5, cex=1.2, col="black", outer=TRUE)
> dev.off()
null device
          1
```

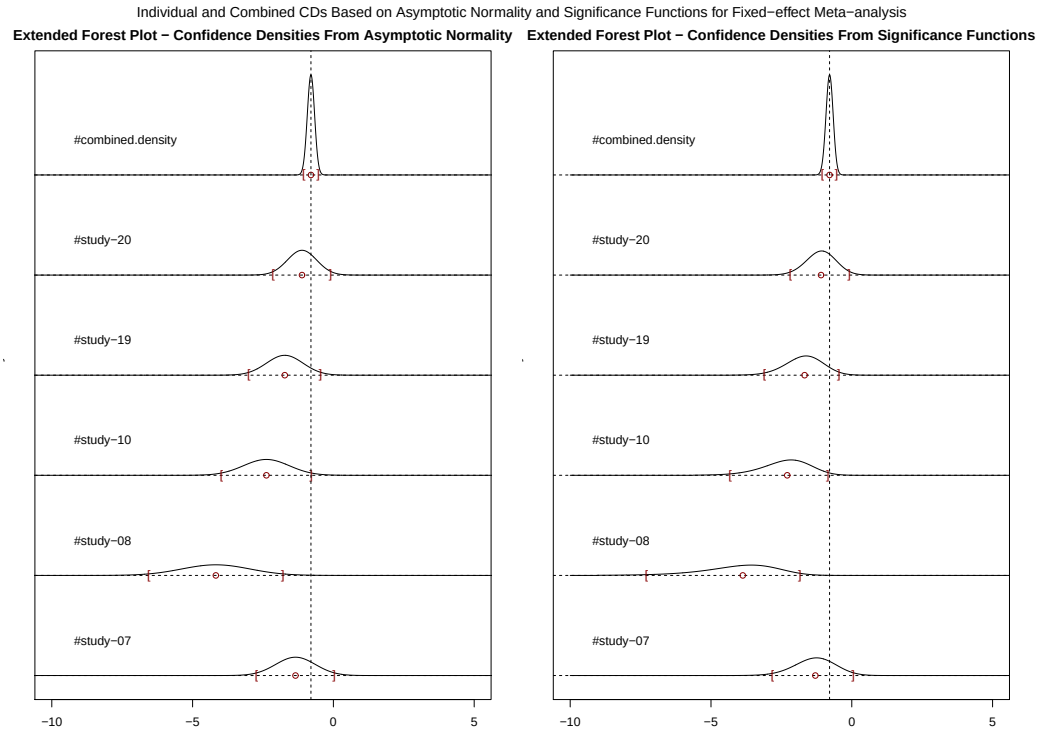


Figure 4.2: Individual and combined CDs based on asymptotic normality and significance functions for fixed-effect meta-analysis.

*The results of using exact CDs as input are very close to the previous ones of using summary statistics, the estimated LOR is* $-0.7897$ *versus* $-0.7970$. *However, Figure 4.2 shows that*

*the exact individual CDs are skewed whereas the normal individual CDs based on summary statistics are symmetric. For example,* `study-08` *and* `study-10`*. Thus, only the exact CDs are able to pass the skewness from individual studies to the combined CD.*

The CD-based meta-analysis framework also incorporates Bayesian methods by viewing the prior distribution as a CD function. In Example 4.9, we suppose that prior information does not show the new treatment is superior than the old treatment, e.g., the prior is given by $\mathcal{N}(0, 0.01)$. This prior can be viewed as a CD and included in meta-analysis.

**Example 4.9** *(Bayesian meta-analysis) To incorporate the prior information in meta-analysis, simply put it in the data.frame as a separate row. We use the ulcer data set with 0.5 continuity correction for studies with zero event. The results is closer to zero as a result of incorporating the evidence in prior distribution.*

```
> # prior information in the first row of the data frame
> ulcer.pdata.bayesian <- rbind(c(0,sqrt(0.01)), ulcer.pdata)
> study.names.bayesian <- c('prior', paste('study-',
+ formatC(c(1:41),width=2,format='d',flag='0'), sep=''))
> # bayesian meta-analysis, prior N(0,0.01).
> gm1fp.bayesian <- gmeta(ulcer.pdata.bayesian, method='fixed-mle',
+ gmo.xgrid=seq(-20,20,by=0.001), study.names=study.names.bayesian)
> summary(gm1fp.bayesian)
                Model-Based Meta-Analysis through CD-Framework


Call:
gmeta.default(gmi = ulcer.pdata.bayesian, method = "fixed-mle",
    study.names = study.names.bayesian, gmo.xgrid = seq(-20,
```

```
        20, by = 0.001))

Summary of Combined CD:
                  mean     median     stddev   ci.lower   ci.upper
Combined CD -0.3445117 -0.3445117 0.07822237 -0.4978241 -0.1911991

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev    ci.lower    ci.upper
prior     0.0000000  0.0000000 0.1000000  -0.1959964  0.19599643
study-01 -1.8382795 -1.8382795 0.9266964  -3.6545713 -0.02198789
[...]
study-40 -8.4390154 -8.4390154 2.0146522 -12.3876613 -4.49036952
study-41  0.5753641  0.5753641 2.0429418  -3.4287283  4.57945657

Confidence level = 0.95
```

*Figure 4.3 plots the combined CD based on fixed-effect model and the Bayesian method. For Bayesian method, the combined CD is pulled closer to zero by the evidence in prior distribution (prior CD).*

```
plot to show prior information pull the combined CD to zero
> postscript(file='ulcer-plot2-fix-mle-bayesian-plot.eps',
+ paper="special", width=14, height=11, horizontal=FALSE)
> par(mfrow=c(1,1), mar=c(2,2,2,2)+0.1, mgp=c(2,1,0), oma=c(1,0,1,0))
> plot(0,0, type='n', xlim=c(-3,2), ylim=c(0,5.5),
```

```
+ xlab=expression(theta), ylab='confidence density')
> lines(gm1fp.mle$x.grids, gm1fp.mle$combined.density, lty=1, lwd=2)
> lines(gm1fp.mle.DE$x.grids, gm1fp.mle.DE$combined.density, lty=2, lwd=2)
> lines(gm1fp.bayesian$x.grids, gm1fp.bayesian$combined.density, lty=3, lwd=2)
> title('Bayesian Meta-analysis through CD-based Meta-analysis Framework')
> legend(x=0.5, y=4, legend=c("fixed-mle with Normal", "fixed-mle with DE",
+ "Bayesian method"), lty=c(1,2,3), lwd=c(2,2,2), seg.len=4, bty='n')
> dev.off()
null device
          1
```

The fixed-effect model assumes $\theta_i = \theta$, which is too restrictive in practice. Instead, the random-effects model assumes $\theta_i \sim \mathcal{N}(\theta, \tau^2)$, and allows heterogeneity among study-specific means. The following example uses `gmeta()` function to perform random-effects model-based meta-analysis.

**Example 4.10** *(conventional random-effect model) The random-effects meta-analysis can be performed by specifying argument* `method='random-mm'` *where* $\tau^2$ *is estimated by moment estimator (DL method), or* `method='random-reml'` *where* $\tau^2$ *is estimated by REML estimator. The results are close to each other, are similar as those obtained by fixed-effect model, and match the results presented by Xie et al. (2011).*

```
> # random-effects meta-analysis, DL estimator for tau2
> gm1rp.mm <- gmeta(ulcer.pdata, method='random-mm',
+ gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1rp.mm)
```

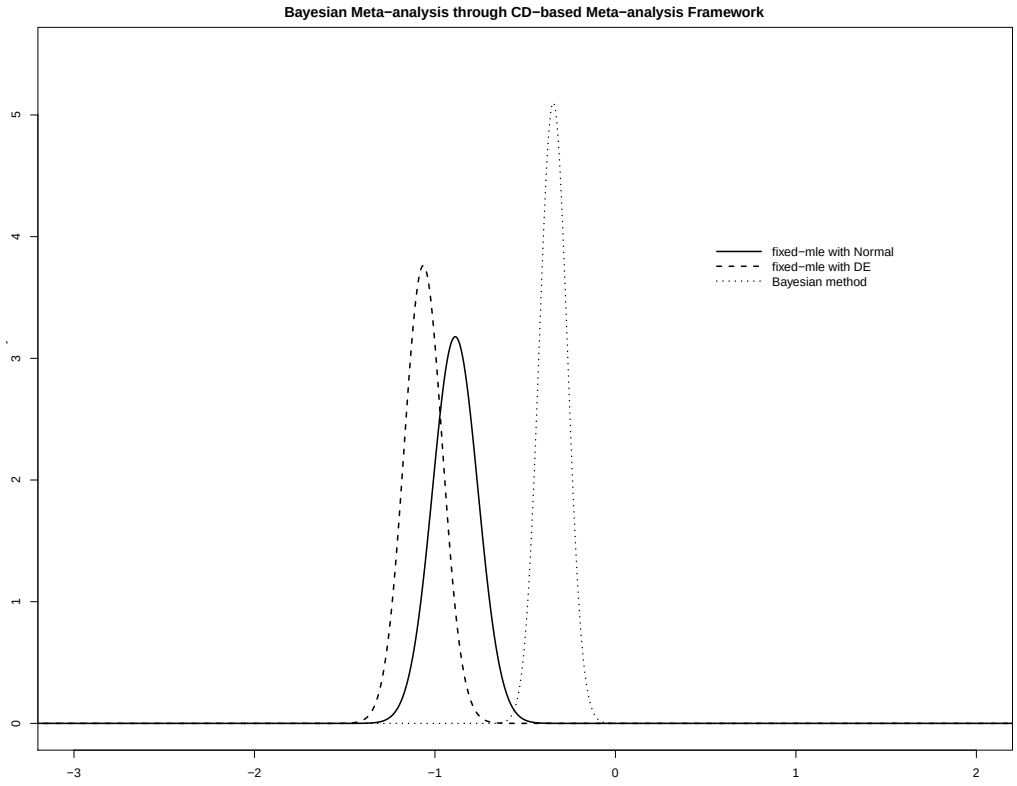


Figure 4.3: The meta-analysis results from fixed-effect model and Bayesian method. The solid, dashed and dotted curves are the combined CD densities based upon fixed-effect model with $a_0(\cdot) = \Phi^{-1}(\cdot)$ and $a_0(\cdot) = \mathrm{DE}^{-1}(\cdot)$, and Bayesian method, respectively. For Bayesian method, the combined CD is pulled closer to zero by the evidence in prior distribution (prior CD).

```
        Model-Based Meta-Analysis through CD-Framework


Call:

gmeta.default(gmi = ulcer.pdata, method = "random-mm", gmo.xgrid = seq(-20,

    20, by = 0.001))


Summary of Combined CD:
```

```
                mean    median    stddev  ci.lower   ci.upper
Combined CD -1.097574 -1.097574 0.2103224 -1.509798 -0.6853496

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -4.474761  0.79820226
[...]
study-41  0.5753641  0.5753641 2.0429418  -3.861400  5.01212827

Confidence level = 0.95
> # random-effects meta-analysis, REML estimator for tau2
> gm1rp.reml <- gmeta(ulcer.pdata, method='random-reml',
+ gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1rp.reml)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pdata, method = "random-reml", gmo.xgrid = seq(-20,
    20, by = 0.001))

Summary of Combined CD:
                mean    median    stddev  ci.lower   ci.upper
Combined CD -1.091384 -1.091384 0.2069317 -1.496963 -0.6858053
```

```
Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -4.436960  0.76040118
[...]
study-41  0.5753641  0.5753641 2.0429418  -3.839042  4.98977023

Confidence level = 0.95
```

*The* **gmeta** *package handles the random-effects model with more versatility. For example, we can use* `method='random-tau2'` *and assign a user-specified* $\tau^2$ *by* `tau2=[value of tau2]`*. We can also use* `method='random-tau2'` *and choose an implemented methods for estimating* $\tau^2$ *using* `tau2=[method to estimate tau2]`*. The available methods include* `'DL', 'EB', 'HE', 'HS', 'ML', 'SJ', 'REML'`*, represents for DerSimonian-Laird method (DerSimonian and Laird, 1986), Empirical-Bayesian method (Morris, 1983), Hedges method (Hedges, 1983), Hunter-Schmidt method (Hunter and Schmidt, 2004), Maximum-Likelihood estimates, Sidik-Jonkman method (Sidik and Jonkman, 2005a,b), and Restricted Maximum-Likelihood estimates, respectively.*

```
> # use user-specified tau2
> gm1rp.tau2 <- gmeta(ulcer.pdata, method='random-tau2',
+ gmo.xgrid=seq(-20,20,by=0.001), tau2=1.00)

> # use implemented method for estimating tau2
> gm1rp.tau2HS <- gmeta(ulcer.pdata, method='random-tau2',
+ gmo.xgrid=seq(-20,20,by=0.001), tau2='HE')
```

```
# compare the estimation on tau2
> round(gm1rp.reml$tau2,4)
[1] 0.8992
> round(gm1rp.tau2HS$tau2,4)
[1] 2.4173
```

*Figure 4.4 plots the combined CD densities based on fixed-effect model, and random-effects model with DL and REML estimates on $\tau^2$. For random-effects models, the combined CDs are more dispersed.*

### 4.4.3 Robust Meta-analysis Methods

The CD combining framework also develops robust meta-analysis methods that provide a means of protection to model misspecification, and limit the impact of outlying studies. These methods are developed uniquely under the CD-based meta-analysis framework, by taking the advantage of using different `linkfunc`, $a_0(\cdot)$ when specifying the $g^{(c)}(\cdot)$ function in (4.10).

In particular, the method `robust1` is developed under fixed-effect model for a set of large studies. Assume a set of $k$ studies with sample sizes $n_1, \ldots, n_k$, where $n_i$ goes to infinity at the same rate $n$. Let us consider the following model as an extension of conventional fixed-effect model in (4.16) (Xie et al., 2011) :

$$\begin{aligned} \widehat{\theta}_i|(\theta_i, s_i) \sim \mathcal{N}(\theta_i, s_i^2), \quad \text{sample size } n_i \propto 1/s_i^2 \\ \theta \equiv \text{median}(\theta_i, i = 1, \ldots, k), \quad \mathcal{I} = \{i : \theta_i = \theta\}, \quad |\mathcal{I}| > [k/2] \end{aligned} \tag{4.18}$$

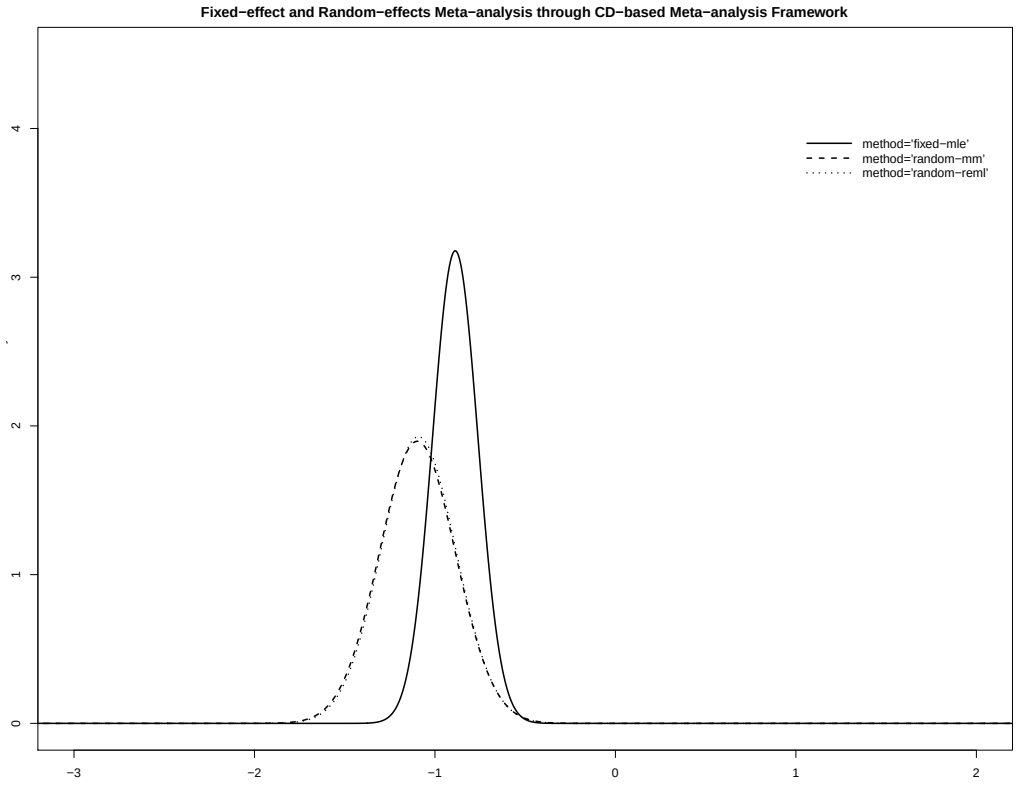


Figure 4.4: The meta-analysis results from fixed-effect and random-effects models. The solid, dashed and dotted curves are the combined CD densities based upon fixed-effect model, random-effects model with DL estimate on heterogeneity, and random-effects model with REML estimate on heterogeneity, respectively. For random-effects models, the combined CDs are more dispersed.

where $[\cdot]$ is the function for rounding to integers. Here, the parameter $\theta$ is the true parameter value of the majority of the studies, i.e., a small fraction of the studies are outlying studies with different true parameter value.

The key idea of `robust1` method is down-weighting the studies far apart the majorities by using adaptive weights, and use the median combined CDs $H^{(c)}_{\texttt{R1,median}}(\theta)$, or simply $H^{(c)}_{\texttt{R1}}(\theta)$, for making inference on $\theta$. Specifically, it can be shown that $\widehat{\theta}^{(c)}_{\texttt{R1}} = {H^{(c)}_{\texttt{R1}}}^{-1}(1/2)$ is a consistent and an asymptotic efficient estimator (as if $\mathcal{I}$ is known) for $\theta$, as $n \to \infty$ (Xie

et al., 2011).

The estimates based on `robust1` method is not consistent under the random-effects model, even as $k \to \infty$. Therefore, `robust2` method is developed under random-effects model for producing consistent estimates of $\theta$ given a large number of $k$ studies, where the study-specific parameter values come from a contaminated distribution. Let us consider the following model as an extension of conventional random-effects model in (4.17) (Huber, 1964; Xie et al., 2011):

$$\widehat{\theta}_i|(\theta_i, s_i) \sim \mathcal{N}(\theta_i, s_i^2) \quad \text{and} \quad \theta_i|(\theta, \tau) \sim (1-\epsilon)\mathcal{N}(\theta, \tau^2) + \epsilon D_\epsilon(\theta), \quad i = 1, \ldots, k, \quad (4.19)$$

where $D_\epsilon(\cdot)$ is some unknown contaminating population. If $\epsilon \equiv 0$, model (4.19) degenerates to the conventional random-effects model (4.17). If not, it implies that the parameter value of some studies comes from contaminated distribution $D_\epsilon(\cdot)$, which is not appropriately accounted in the conventional random-effects model.

Let $\theta_0$ and $\theta_*$ be the population mean of the target distribution $\mathcal{N}(\theta, \tau^2)$ and the contaminated distribution $(1-\epsilon)\mathcal{N}(\theta, \tau^2) + \epsilon D_\epsilon(\theta)$, respectively. Then, $\theta_* = \theta_0$ if $\epsilon = 0$ or $D_\epsilon(\cdot)$ is symmetric around $\theta_0$. The `robust2` method is in essential equivalent to solve an M-estimating equation $\sum_{i=1}^k w_i\{H_i(\theta) - 1/2\} = 0$. It takes $a_0(t) = t$ and thus $g^{(c)}(u_1, \ldots, u_k) = \sum_{i=1}^k w_i u_i$, which leads to the combined aCD function, $H_{\texttt{R2}}(\theta)$. Specifically, it can be shown that $\widehat{\theta}_{\texttt{R2}}^{(c)} = H_{\texttt{R2}}^{(c)^{-1}}(1/2)$ is a consistent estimator of $\theta_*$ as $k \to \infty$. Further, in the case of no contamination, the asymptotic relative efficiency of $H_{\texttt{R2}}^{(c)}(\cdot)$ compared to the Fisher efficient CD $H^{(c)}(\cdot)$ in (4.13) is $\sqrt{3/\pi} = 0.9972$ as $k \to \infty$ and sample sizes $n \to \infty$, assuming all $n_i$ goes to infinity at the same rate $n$ (Xie et al., 2011).

In practice, the performance of `robust1` and `robust2` are relative close to each other. The `gmeta()` function can perform these robust meta-analysis by choosing

method='fixed-robust1', method='fixed-robust2', method='random-robust1', or method='random-robust2', as illustrated in the following example.

**Example 4.11** *(robust meta-analysis methods) Apply the robust meta-analysis on the ulcer data set, we obtain similar results as the conventional fixed-effect and random-effects meta-analysis obtained in the previous examples. This implies that the data in the original data set are consistent to each other, and no outlying study exists.*

```
# robust meta-analysis methods
> gm1fp.robust1 <- gmeta(ulcer.pdata, method='fixed-robust1',
+ gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1fp.robust1)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pdata, method = "fixed-robust1",
    gmo.xgrid = seq(-20, 20, by = 0.001))

Summary of Combined CD:
                 mean    median    stddev  ci.lower   ci.upper
Combined CD -1.179134 -1.179134 0.1467666 -1.466792 -0.8914757

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev    ci.lower    ci.upper
```

```
study-01 -1.8382795 -1.8382795 0.9266964  -3.6545713 -0.02198789
[...]
study-41  0.5753641  0.5753641 2.0429418  -3.4287283  4.57945657

Confidence level = 0.95
>
> gm1rp.robust2 <- gmeta(ulcer.pdata, method='random-robust2',
+ gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1rp.robust2)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pdata, method = "random-robust2",
    gmo.xgrid = seq(-20, 20, by = 0.001))

Summary of Combined CD:
                 mean    median    stddev  ci.lower  ci.upper
Combined CD -1.073812 -1.072808 0.2170723 -1.505937 -0.647466

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -4.436960  0.76040118
[...]
study-41  0.5753641  0.5753641 2.0429418  -3.839042  4.98977023
```

```
Confidence level = 0.95
```

*Figure 4.5 plots the combined CDs densities based on conventional and robust methods. It is obvious that the combined CDs from robust methods are slightly wider and thus less efficient than the corresponding ones from conventional methods, which is the trade off for robustness.*

To demonstrate the robustness property, we construct a contaminated data set by artificially creating some outlying studies. In particular, six studies (study-05,13,14,22,35,41) have $\widehat{\theta}_i > 0.5$, which are in favor of the old treatment and contradict to the overall conclusion. We make the situation even worse by multiplying these $\widehat{\theta}_i$s by a factor of 10. As a result, these six studies are "outlying" studies far apart from the majorities.

```
> # construct a contaminated data set
> ulcer.pdata.cntm <- cbind(
+ ifelse(ulcer.theta>0.5, 10*ulcer.theta, ulcer.theta), ulcer.sigma)
```

**Example 4.12** *(robust meta-analysis method - continued) Apply the conventional and robust methods to the original and contaminated data set.*

```
> # fixed-effect meta-analysis on original data set
> gm1fp.mle <- gmeta(ulcer.pdata, method='fixed-mle',
+ gmo.xgrid=seq(-20,20,by=0.001))
> gm1fp.robust1 <- gmeta(ulcer.pdata, method='fixed-robust1',
+ gmo.xgrid=seq(-20,20,by=0.001))
> gm1fp.robust2 <- gmeta(ulcer.pdata, method='fixed-robust2',
```

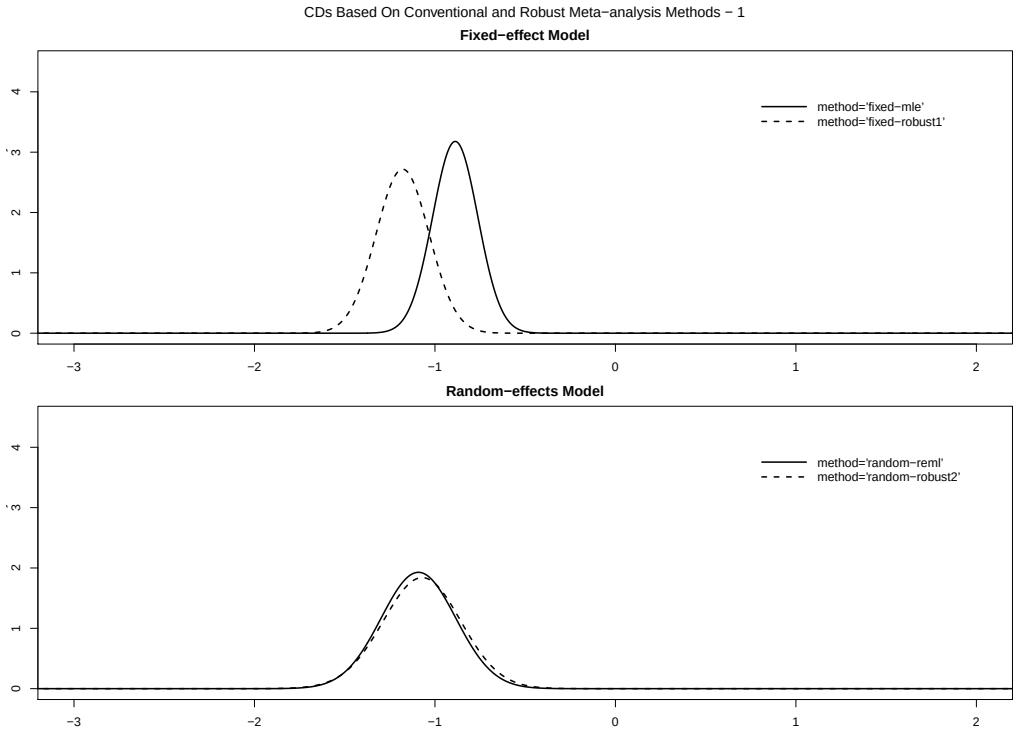


Figure 4.5: The meta-analysis results from conventional and robust fixed-effect and random-effects models. The top figure plots the combined CDs densities based on fixed-effect model, where the solid and dashed curves draw the results from conventional and robust methods, respectively. The bottom figure plots the combined CDs densities based on random-effects model, where the solid and dashed curves draw the results from conventional and robust methods, respectively. The combined CDs from robust methods are slightly wider and thus less efficient than the corresponding ones from conventional methods, which is the trade off for robustness.

```
+ gmo.xgrid=seq(-20,20,by=0.001))

>

> # fixed-effect meta-analysis on contaminated data set

> gm2fp.mle <- gmeta(ulcer.pdata.cntm, method='fixed-mle',

+ gmo.xgrid=seq(-20,20,by=0.001))
```

```
> gm2fp.robust1 <- gmeta(ulcer.pdata.cntm, method='fixed-robust1',
+ gmo.xgrid=seq(-20,20,by=0.001))
> gm2fp.robust2 <- gmeta(ulcer.pdata.cntm, method='fixed-robust2',
+ gmo.xgrid=seq(-20,20,by=0.001))
>
> # random-effect meta-analysis on original data set
> gm1rp.reml <- gmeta(ulcer.pdata, method='random-tau2',
+ gmo.xgrid=seq(-20,20,by=0.001), tau2=tau2REML)
> # user-specified tau2
> tau2REML <- gm1rp.reml$tau2
> gm1rp.robust1 <- gmeta(ulcer.pdata, method='random-robust1',
+ gmo.xgrid=seq(-20,20,by=0.001), tau2=tau2REML)
> gm1rp.robust2 <- gmeta(ulcer.pdata, method='random-robust2',
+ gmo.xgrid=seq(-20,20,by=0.001), tau2=tau2REML)
>
> # random-effect meta-analysis on contaminated data set
> gm2rp.reml <- gmeta(ulcer.pdata.cntm, method='random-tau2',
+ gmo.xgrid=seq(-20,20,by=0.001), tau2=tau2REML)
> gm2rp.robust1 <- gmeta(ulcer.pdata.cntm, method='random-robust1',
+ gmo.xgrid=seq(-20,20,by=0.001), tau2=tau2REML)
> gm2rp.robust2 <- gmeta(ulcer.pdata.cntm, method='random-robust2',
+ gmo.xgrid=seq(-20,20,by=0.001), tau2=tau2REML)
```

*The results of the conventional fixed-effect or random-effects meta-analysis methods are significantly influenced by the outlying studies in the contaminated data set. For example, the estimate of* $\theta$ *and its associated* $95\%CI$*, based on random-effects model, moves from*

$-1.0914(-1.4970, -0.6858)$ *to* $-0.0985(-0.5041, 0.3071)$. *Thus, the superiority conclusion for the new treatment is no longer valid. The fixed-effect model provides similar results.*

```
> summary(gm1rp.reml)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pdata, method = "random-tau2",
    gmo.xgrid = seq(-20, 20, by = 0.001), tau2 = tau2REML)

Summary of Combined CD:
                 mean    median    stddev  ci.lower   ci.upper
Combined CD -1.091384 -1.091384 0.2069317 -1.496963 -0.6858053

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -4.436960  0.76040118
[...]
study-41  0.5753641  0.5753641 2.0429418  -3.839042  4.98977023

Confidence level = 0.95
>
> summary(gm2rp.reml)
                Model-Based Meta-Analysis through CD-Framework
```

```
Call:
gmeta.default(gmi = ulcer.pdata.cntm, method = "random-tau2",
    gmo.xgrid = seq(-20, 20, by = 0.001), tau2 = tau2REML)

Summary of Combined CD:
                  mean      median    stddev   ci.lower  ci.upper
Combined CD -0.09849573 -0.09849573 0.2069316 -0.5040743 0.3070829

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -4.436960  0.76040118
[...]
study-41  5.7536414  5.7536414 2.0429418   1.339235 10.16804747

Confidence level = 0.95
```

*On the other hand, the robust methods resist to the impact of outlying studies. For example, the* ***random-robust2*** *method obtains* $-1.0738(-1.5059, -0.6475)$ *when applying to the original data set, and* $-1.0096(-1.4987, -0.5086)$ *when applying to the contaminated data set. The other methods,* ***fixed-robust1****,* ***fixed-robust2****, and* ***random-robust1*** *all give similar results as the* ***random-robust2*** *method.*

```
> summary(gm1rp.robust2)
```

```
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pdata, method = "random-robust2",
    gmo.xgrid = seq(-20, 20, by = 0.001), tau2 = tau2REML)

Summary of Combined CD:
                 mean    median    stddev  ci.lower  ci.upper
Combined CD -1.073812 -1.072808 0.2170723 -1.505937 -0.647466

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -4.436960  0.76040118
[...]
study-41  0.5753641  0.5753641 2.0429418  -3.839042  4.98977023

Confidence level = 0.95
>
> summary(gm2rp.robust2)
                Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pdata.cntm, method = "random-robust2",
    gmo.xgrid = seq(-20, 20, by = 0.001), tau2 = tau2REML)
```

```
Summary of Combined CD:
                 mean    median    stddev  ci.lower   ci.upper
Combined CD -1.009601 -1.011652 0.2489081 -1.498723 -0.5086346

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.8382795 -1.8382795 0.9266964  -4.436960  0.76040118
[...]
study-41  5.7536414  5.7536414 2.0429418   1.339235 10.16804747

Confidence level = 0.95
```

*Figure 4.6 plots results from all methods on both original data set and contaminated data set. The top three figures plot the results of fixed-effect methods. In particular, part (a) plots the results of* `fixed-mle` *method, where the solid and dashed curves are combined CD densities using the original and contaminated data set, respectively. It is obvious that the results are impacted by the outlying studies, and the combined CD is pulled closer to zero significantly when applying to the contaminated data set. Further, part (b) and (c) plot the results of* `fixed-robust1` *and* `fixed-robust2` *methods. The combined CDs using the original and contaminated data set are drawn by solid and dashed curves, and are close to each other. Thus, the robust methods limit the impact of outlying studies. Likewise, part (d), (e), (f) plot the results of* `random-reml`*,* `random-robust1`*, and* `random-robust2` *methods. The solid and dashed curves are combined CD densities using the original and*

*contaminated data set, respectively. Once again, the conventional meta-analysis method results are undermined by the outlying studies, whereas the robust meta-analysis methods results are not.*

### 4.4.4 Meta-analysis of 2x2 Tables

The CD combining framework includes the Mantel-Haenszel and Peto's method for 2x2 tables as a special case (Yang et al., 2012b). As a consequence, the `gmeta()` function can also perform meta-analysis on 2x2 tables using Mantel-Haenszel or Peto's method. Let us consider a set of $k$ studies, where the results of each study are reported as 2x2 table in Table 4.1. Assume

$$X_i \sim \text{Binomial}(n_i, p_1), \quad Y_i \sim \text{Binomial}(m_i, p_2), \quad i = 1, \ldots, k. \tag{4.20}$$

Then, the Mantel-Haenszel method provides an overall estimate on odd ratio (OR), $\Psi = \dfrac{p_1/(1-p_1)}{p_2/(1-p_2)}$, and the Peto's method provides an overall estimate on LOR, $\theta = \log(\Psi)$.

**Example 4.13** *(meta-analysis of 2x2 tables - Mantel-Haenszel and Peto's method) To use `gmeta()` function for meta-analysis of 2x2 tables, the input must be organized as a $4 \times k$ data.frame or matrix. In particular, each row of the data.frame or matrix represents outcomes from a single study, in the order of $(X_i, n_i, Y_i, m_i)$, following the notation in Table 4.1.*

```
> ulcer.2x2dt <- cbind(ulcer.o[,1], ulcer.o[,1]+ulcer.o[,2],
+ ulcer.o[,3], ulcer.o[,3]+ulcer.o[,4])
```

*The type of input must be specified as 2x2 table using `gmi.type='2x2'`. The Mantel-Haenszel method can be called using `method='Mantel-Haenszel'` or `method='MH'`, and*

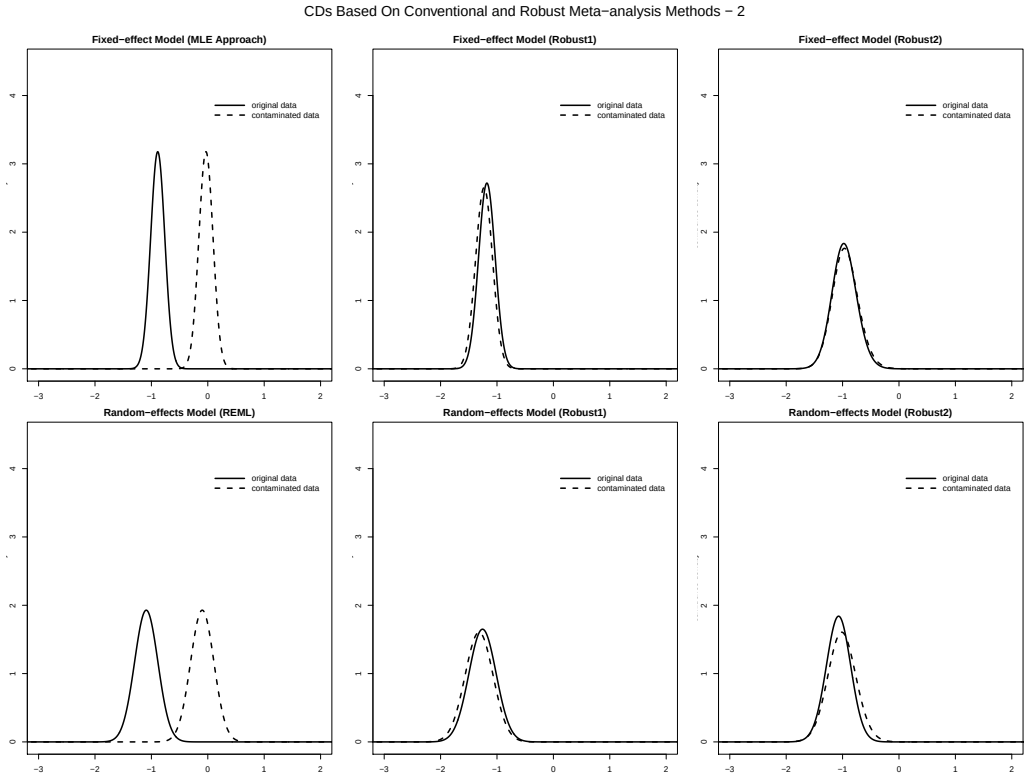


Figure 4.6: The meta-analysis results from conventional and robust meta-analysis methods. The top three figures plot the results of fixed-effect methods. In particular, part (a) plots the results of `fixed-mle` method, where the solid and dashed curves are combined CD densities using the original and contaminated data set. It is obvious that the results are impact by the outlying studies, and the combined CD is pulled closer to zero significantly when applying on the contaminated data set. Further, part (b) and (c) plot the results of `fixed-robust1` and `fixed-robust2` methods. The combined CD using the original and contaminated data set are drawn by solid and dashed curves, and are close to each other. Thus, the robust methods limit the impact of outlying studies. Likewise, part (d), (e), (f) plot the results of `random-reml`, `random-robust1`, and `random-robust2` methods. The solid and dashed curves are combined CD densities using the original and contaminated data set, respectively. Once again, the conventional meta-analysis method results are undermined by the outlying studies, whereas the robust meta-analysis methods results are not.

*the Peto's method can be invoked using* `method='Peto'`.

```
> # Mantel-Haenszel's odds ratio meta-analysis on 2x2 table
> gm1dt.MH <- gmeta(ulcer.2x2dt, gmi.type='2x2', method='MH',
+ gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1dt.MH)
                Exact Meta-Analysis Approach through CD-Framework

Call:
gmeta.default(gmi = ulcer.2x2dt, gmi.type = "2x2", method = "MH",
    gmo.xgrid = seq(-20, 20, by = 0.001))

Summary of Combined CD:
                 mean    median     stddev ci.lower  ci.upper
Combined CD 0.3370037 0.3370037 0.03746689  0.26357 0.4104375

Confidence level = 0.95

Summary of Individual CDs:
               mean     median     stddev     ci.lower   ci.upper
study-01 0.15909091 0.15909091 0.14742897 -0.129864570 0.44804639
study-02 0.72727273 0.72727273 0.49641843 -0.245689511 1.70023497
[...]
study-05        Inf        Inf        Inf          NaN        Inf
study-06 0.00000000 0.00000000        NaN          NaN        NaN
[...]
```

```
study-11 0.00000000 0.00000000        NaN          NaN        NaN
[...]
study-25 0.00000000 0.00000000        NaN          NaN        NaN
[...]
study-28 0.00000000 0.00000000        NaN          NaN        NaN
study-29 0.00000000 0.00000000        NaN          NaN        NaN
[...]
study-34 0.00000000 0.00000000        NaN          NaN        NaN
[...]
study-40 0.00000000 0.00000000        NaN          NaN        NaN
study-41        NaN        NaN        NaN          NaN        NaN

Confidence level = 0.95
> # on log odds ratio scale
> log(c(gm1dt.MH$combined.mean,gm1dt.MH$combined.ci))
[1] -1.0876612 -1.3334363 -0.8905316
>
> # Peto's log odds ratio meta-analysis on 2x2 table
> gm1dt.Pt <- gmeta(ulcer.2x2dt, gmi.type='2x2', method='Peto',
+ gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1dt.Pt)
                Exact Meta-Analysis Approach through CD-Framework

Call:
gmeta.default(gmi = ulcer.2x2dt, gmi.type = "2x2", method = "Peto",
    gmo.xgrid = seq(-20, 20, by = 0.001))
```

```
Summary of Combined CD:
                 mean    median    stddev  ci.lower   ci.upper
Combined CD -1.132064 -1.132064 0.1093015 -1.346291 -0.9178366

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower    ci.upper
study-01 -1.5938462 -1.5938462 0.7765803 -3.1159155 -0.07177678
study-02 -0.3090374 -0.3090374 0.6713267 -1.6248135  1.00673875
[...]
study-40 -3.9411765 -3.9411765 0.4814913 -4.8848822 -2.99747079
study-41        NaN        NaN       Inf        NaN         NaN

Confidence level = 0.95
```

*The results are similar as the results of fixed-effect and random-effects model-based meta-analysis. The* ***Inf*** *and* ***NaN*** *reflect the uncertainty introduced by the zeros – it is impossible to evaluate how extreme the OR or LOR will be on one or two sides when studies have zero event.*

The Mantel-Haenszel and Peto's method can perform meta-analysis on 2x2 tables and handle studies with zero event without continuity corrections. However, both methods are based on large sample theory, such as the asymptotic normality of the sample distribution of the summary statistics OR or LOR. Thus, both methods fail when the studies sample

sizes are small. Further, studies with zero-zero event, e.g. study-41 in ulcer data set, do not contribute to the estimation in Mantel-Haenszel method at all. Thus, the method in essential excludes those studies. Such approach often lead to bias and loss of efficiency (Nissen and Wolski, 2007; Tian et al., 2009; Liu et al., 2013).

To use all information without assuming asymptotic normality and without artificial continuity corrections, an exact meta-analysis method, developed uniquely under the CD combining framework, was proposed for estimating the overall LOR of 2x2 tables (Liu et al., 2013). The key idea is to use significance functions based on mid-p adaptation of Fisher's Exact Test (4.7) for constructing individual CDs, with some further adjustments for improving efficiency. In particular, it can be shown that the the proposed method, `'exact1'`, ensures correct Type I error rate and improves efficiency, when the Mental-Haenszel and Peto's methods failed (Liu et al., 2013).

**Example 4.14** *(meta-analysis of 2x2 tables - exact method for LOR) The exact meta-analysis method of 2x2 tables for LOR can be called using* `method='exact1'`.

```
> # exact 1 (LOR): Liu et al. 2012
> gm1dt.e1 <- gmeta(ulcer.2x2dt, gmi.type='2x2', method='exact1',
+ gmo.xgrid=seq(-20,20,by=0.001))
> summary(gm1dt.e1)
                Exact Meta-Analysis Approach through CD-Framework


Call:
gmeta.default(gmi = ulcer.2x2dt, gmi.type = "2x2", method = "exact1",
    gmo.xgrid = seq(-20, 20, by = 0.001))
```

```
Summary of Combined CD:
                 mean    median    stddev  ci.lower   ci.upper
Combined CD -1.239827 -1.240056 0.1270008 -1.485981 -0.9882136

Confidence level = 0.95

Summary of Individual CDs:
               mean     median    stddev   ci.lower     ci.upper
study-01 -1.7911586 -1.7313392 0.9660527 -3.9201201 -0.009971508
study-02 -0.3098157 -0.3068812 0.6981395 -1.6933698  1.058949659
[...]
study-05        Inf        Inf       Inf -0.4680651          Inf
study-06       -Inf       -Inf       Inf       -Inf  1.040483624
[...]
study-11       -Inf       -Inf       Inf       -Inf -0.332826382
[...]
study-25       -Inf       -Inf       Inf       -Inf -1.631226137
[...]
study-28       -Inf       -Inf       Inf       -Inf -2.098344052
study-29       -Inf       -Inf       Inf       -Inf -0.940709896
[...]
study-34       -Inf       -Inf       Inf       -Inf -0.160331137
[...]
study-40       -Inf       -Inf       Inf       -Inf -5.239567140
study-41        NaN       -Inf       Inf       -Inf          Inf
```

```
Confidence level = 0.95
```

*Figure 4.7 makes an extended forest plot of individual and combined CDs (confidence curves) based on Peto's method and exact1 method for meta-analysis of 2x2 tables. The exact1 method appropriately accounts the impact of zero event, where the confidence curves go to infinity without decreasing at the sides having zero event, see the confidence curves for* `study-05` *and* `study-06`*.*

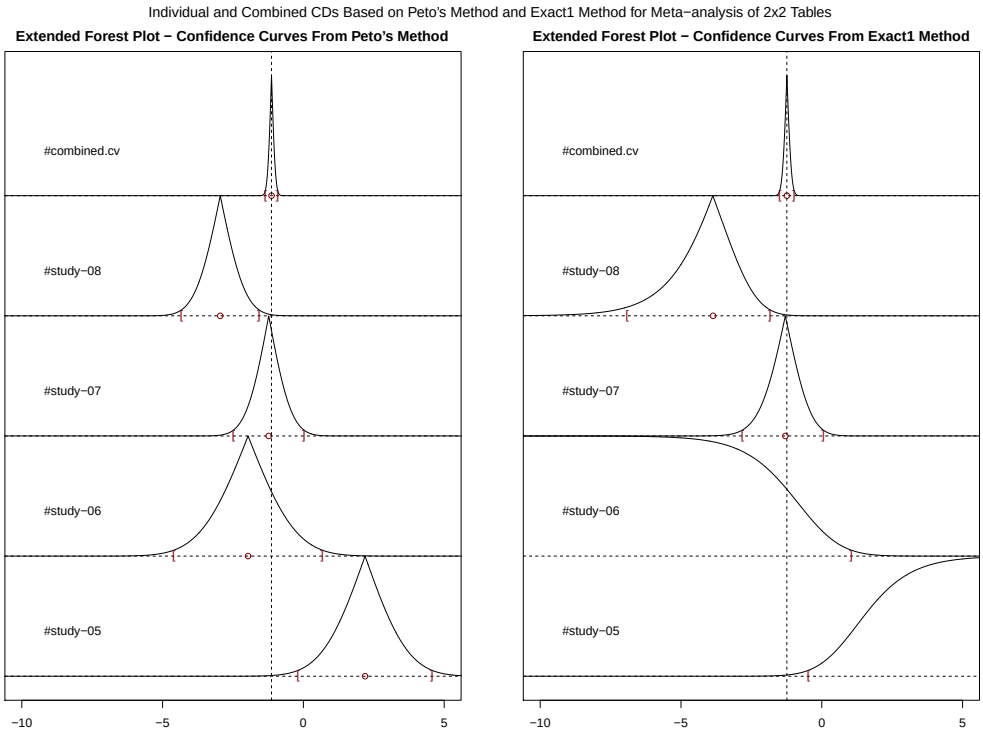


Figure 4.7: Individual and combined CDs (confidence curves) based on Peto's method and exact1 method for meta-analysis of 2x2 tables. The exact1 method appropriately accounts the impact of zero event, where the confidence curves go to infinity without decreasing at the sides having zero event, see the confidence curves for `study-05` and `study-06`.

The risk difference (RD), $\Delta = p_1 - p_2$, is another measurement often used in clinical trials. It might be preferred for including studies with zero event, since the sample estimates of LOR are not meaningful when either arm has zero event. An exact meta-analysis method of 2x2 tables for RD is proposed by Tian et al. (2009). Hereafter, the Tian's method. It provides correct Type I error rate, and includes studies with zero event without artificial continuity corrections.

The CD combining framework includes the Tian's method as a special case, by taking $a_0(u) = \sum_{j=1}^{J} \widetilde{w}_j(\mathbf{I}(u > 1 - \eta_j) - \eta_j)$ in (4.10), where $0 \leq \eta_1, \ldots, \eta_J \leq 1$ are the confidence levels and $\widetilde{w}_j$ are weights associated with those confidence levels (Yang et al., 2012a). For example, $\eta_j = 0.1 + 0.85 \times (j-1)/19$ and $\widetilde{w}_j = \{\eta_j(1-\eta_j)\}^{-1}, j = 1, \ldots, 20$, are suggested by Tian et al. (2009).

**Example 4.15** *(meta-analysis of 2x2 tables - exact method for RD) The exact meta-analysis method of 2x2 tables for RD (the Tian's method) can be called using* `method='exact2'`. *The vector* $\eta_j, j = 1, \ldots, J$ *can be specified by* `eta=seq(0.10, 0.95, length=20)`, *and weights associated with each confidence level is, in default,* $\widetilde{w}_j = \{\eta_j(1-\eta_j)\}^{-1}$. *These confidence levels associated weights,* $\widetilde{w}_j$, *are not allowed to change.*

```
> # exact 2 (RD): Tian et al. 2009
> gm1dt.e2o <- gmeta(ulcer.2x2dt, gmi.type='2x2', method='exact2',
+ gmo.xgrid=seq(-1,1,by=0.001), eta=seq(0.10, 0.95, length=20))
> summary(gm1dt.e2o)
                Exact Meta-Analysis Approach through CD-Framework

Call:
```

```
gmeta.default(gmi = ulcer.2x2dt, gmi.type = "2x2", method = "exact2",
    gmo.xgrid = seq(-1, 1, by = 0.001), eta = seq(0.1, 0.95,
        length = 20))

Summary of Combined CD:
                  mean     median     stddev   ci.lower   ci.upper
Combined CD -0.1816916 -0.1800597 0.02646488 -0.2396198 -0.1215608

Confidence level = 0.95

Summary of Individual CDs:
                mean       median     stddev    ci.lower      ci.upper
study-01 -0.37509270 -0.386095745 0.19755120 -0.69071037 -0.0209623136
study-02 -0.07953751 -0.103013552 0.20736116 -0.44624424  0.2850587896
[...]
study-25 -0.84767957           NA         NA          NA -0.5607693668
[...]
study-40 -0.97171083           NA         NA          NA -0.9143694415
study-41  0.01685562 -0.004672469 0.10196604 -0.17018078  0.2815502333

Confidence level = 0.95
```

*In the previous example, the argument* **`eta`** *takes a vector of values between 0 and 1 to indicate the levels of confidence intervals used in the analysis. The CD combining method,* **`'exact2'`***, can let the number of confidence intervals used in Tian's method goes to infinity, e.g.,* $J \to \infty$*, by taking* **`eta='Inf'`***. This is also the default value of argument* **`eta`***.*

```
> gm1dt.e2 <- gmeta(ulcer.2x2dt, gmi.type='2x2', method='exact2',
+ gmo.xgrid=seq(-1,1,by=0.001), eta='Inf')
> summary(gm1dt.e2)
                Exact Meta-Analysis Approach through CD-Framework

Call:
gmeta.default(gmi = ulcer.2x2dt, gmi.type = "2x2", method = "exact2",
    gmo.xgrid = seq(-1, 1, by = 0.001), eta = 'Inf')

Summary of Combined CD:
                  mean     median    stddev   ci.lower    ci.upper
Combined CD -0.2227558 -0.2214107 0.0182077 -0.2636318 -0.04452226

Confidence level = 0.95

Summary of Individual CDs:
                mean       median     stddev    ci.lower      ci.upper
study-01 -0.37509270 -0.386095745 0.19755120 -0.69071037 -0.0209623136
study-02 -0.07953751 -0.103013552 0.20736116 -0.44624424  0.2850587896
[...]
study-25 -0.84767957           NA         NA          NA -0.5607693668
[...]
study-40 -0.97171083           NA         NA          NA -0.9143694415
study-41  0.01685562 -0.004672469 0.10196604 -0.17018078  0.2815502333

Confidence level = 0.95
```

*Figure 4.8 shows an extended forest plot of the individual and combined CDs for RD from the Tian's method. The left part uses confidence curves, and the right part uses confidence densities. The confidence densities based on the Tian's method are often ragged since the distribution $G^{(c)}(\cdot)$ is acquired by simulation.*

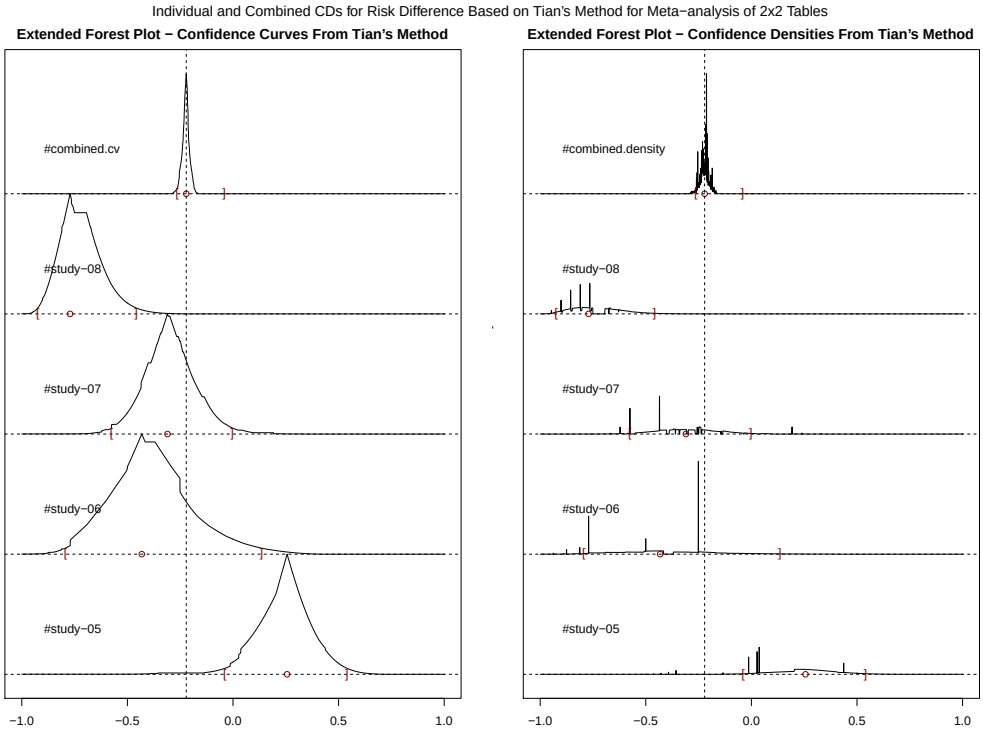


Figure 4.8: Extended forest plot of the individual and combined CDs for RD from the Tian's method. The left part uses confidence curves, and the right part uses confidence densities. The confidence densities based on the Tian's method are often ragged since the distribution $G^{(c)}(\cdot)$ are acquired by simulation.

### 4.4.5 Final Remarks

The `gmeta()` function is versatile, advanced users can build their own CDs for individual studies, and use any available methods, for example, `method='random-robust2'`, for the combination. The user can also add subjective opinions by assigning user-specified weights to individual studies, using `weight=[a vector of weights (real value), one for each study]`. In this way, the default weights will be overrode. In summary, the key idea of using CD combining framework for meta-analysis resides in selecting a suitable function $g^{(c)}(u_1, \ldots, u_k)$.

## 4.5 Further Work

The ultimate goal of the **gmeta** package is providing a simple unified meta-analysis approach. The unification is supported by the combining of independent CDs. This CD combining framework unifies all commonly used meta-analysis methods, including $p$-values combinations, fixed-effect and random-effects model-based meta-analysis, Mantel-Haenszel method, and Peto's method, under the same structure. It develops robust meta-analysis methods that provide a means of protection to model misspecification, and limit the impact of outlying studies. It also covers two recently developed exact meta-analysis methods for combining 2x2 tables with rare events (Tian et al., 2009; Liu et al., 2013).

For the further developments, adding user specified $g^{(c)}(\cdot)$ function is one important feature for providing more flexibility. The current version implements $g^{(c)}(\cdot)$ in form of (4.10), where the function $a_0(\cdot)$ is specified by argument `linkfunc` and the weights $w_i$ are specified by argument `weight`. The available choices for `linkfunc` are `'inverse-normal-cdf'`, and `'inverse-laplace-cdf'`. The option `'inverse-normal-cdf'` in general leads

to Fisher efficiency, whereas the option `'inverse-laplace-cdf'` would lead to Bahadur efficiency in most cases. Adding additional options for $a_0(\cdot)$ can bring other useful results in the same structure. For example, `linkfunc='identity'` can lead to robust results – same as the `'robust2'` methods currently implemented. Then, `method='fixed-robust2'` (`method='random-robust2'`) can be removed and instead is called by using `method='fixed-mle'` (`method='random-mm'`) with `linkfunc='identity'`. However, in this way, the `gmeta()` function is more powerful in the sense that the user can mix the `linkfunc` with other model assumptions and user-specified weights. Another example would be `linkfunc='logit'`, i.e., $a_0(u) = \log(\frac{u}{1-u})$, which leads to results similar as the Tian's method with $J \to \infty$.

The computation complexity is another issue. Advanced meta-analysis methods, like the exact methods, take 2-3 minutes for combining the 41 studies in the ulcer data set, using the author's personal computer (Intel®E7500 CPU, 2.93GHz). Translate the R code to C code can improve the computational efficiency. The translation is done for the $p$-value combination approaches. For other approaches, it is still undergoing.

At last, it is promising to add the ability to perform multivariate meta-analysis, and provide simultaneous inference on a vector of dependent parameters. The CD combining framework is generalized to multivariate cases, and correspondingly, efficient and robust multivariate meta-analysis methods are developed through the combining of independent multivariate CDs (Yang et al., 2013a,b). Implementation of such methods would be a good starting point.

# Chapter 5

# Concluding Remarks

In this dissertation, we develop new statistical methodologies for combining information from independent sources. These methods use confidence distributions as the combining tools, and support the development of efficient and new robust multivariate meta-analysis approaches. The confidence distribution combining method naturally includes the conventional point estimators combining methods as special cases. The new robust meta-analysis approaches can withstand to the impact of potential outlying studies and provide protections to model misspecification during the combining step of the meta-analysis.

In Chapter 2, we use the confidence distribution random vectors as the combining tool to develop a multivariate confidence distribution combining method. This method leads to the asymptotic efficient meta-analysis method under the standard fixed-effect and random-effects model. However, it is limited as not ready to develop other new methods, e.g., the robust meta-analysis methods. Thus, in Chapter 3, we directly combine multivariate normal confidence distribution functions, by using an equivalent set of univariate confidence distribution functions of each multivariate confidence distribution. This combining method is broad to include the the previous method as special cases and support the development of two new robust multivariate meta-analysis method. In the additional Chapter 4, we introduce a user-friendly R package **gmeta**, which realizes the unified univariate meta-analysis framework through combining confidence distributions proposed in Xie et al. (2011).

As a concluding remark, we want to point out that the confidence distribution combining method is very broad and flexible, which might be a potential approach to solve other open meta-analysis problems, e.g., meta-analysis of dependent studies.

# Bibliography

Arends, L. R. (2006). *Multivariate meta-analysis: modelling the heterogeneity mixing apples and oranges; dangerous or delicious?* Erasmus University Rotterdam.

Arends, L. R., Vokó, Z., and Stijnen, T. (2003). Combining multiple outcome measures in a meta-analysis: an application. *Statistics in medicine*, 22(8):1335–1353.

Bayes, T. (1763). An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418.

Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics*, 28(4):783–798.

Chung, K. L. and AitSahlia, F. (2003). *Elementary Probability Theory: with stochastic processes and an introduction to mathematical finance*. Springer.

Cox, D. (2013). Discussion of "Confidence distribution, the frequentist distribution estimator of a parameter — A Review" by Xie and Singh. *International Statistical Review*, 81(1):40–41.

Cramér, H. and Wold, H. (1936). Some theorems on distribution functions. *Journal of the London Mathematical Society*, 1(4):290–294.

De Blasi, P. and Schweder, T. (2012). Tail symmetry of confidence curves based on log-likelihood ratio. *Technical Report*.

Dempster, A. P. (2008). The Dempster-Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, 48:365–377.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188.

Dias, S., Welton, N., Caldwell, D., and Ades, A. (2010). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in medicine*, 29(7-8):932–944.

Dominici, F. and Parmigiani, G. (2005). Combining studies with continuous and dichotomous responses: a latent-variables approach. *Meta-analysis in medicine and health policy*, pages 105–125.

Dudley, R. M. (2002). *Real analysis and probability*, volume 74. Cambridge University Press.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika*, 80(1):3–26.

Efron, B. (1996). Empirical bayes methods for combining likelihoods. *Journal of the American Statistical Association*, 91(434):538–550.

Emerson, J. D., Hoaglin, D. C., and Mosteller, F. (1993). A modified random-effect procedure for combining risk difference in sets of 2×2 tables from clinical trials. *Journal of the Italian Statistical Society*, 2(3):269–290.

Fisher, R. A. (1922). On the interpretation of $\chi$ 2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, 85(1):87–94.

Fisher, R. A. (1930). Inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26:528–535.

Fisher, R. A. (1932). *Statistical methods for research workers (4th ed.)*. Edinburgh, Scotland: Oliver & Boyd.

Fraser, D. (1991). Statistical inference: Likelihood to significance. *Journal of the American Statistical Association*, pages 258–265.

Fraser, D. and McDunnough, P. (1984). Further remarks on asymptotic normality of likelihood and conditional analyses. *Canadian Journal of Statistics*, 12(3):183–190.

Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 5(10):3–8.

Hannig, J. and Xie, M. (2012). A note on Dempster-Shafer recombination of confidence distributions. *Electronic Journal of Statistics*, 6:1943–1966.

Hardy, R. and Thompson, S. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*, 15(6):619–629.

Hedges, L. (1983). A random effects model for effect sizes. *Psychological Bulletin*, 93(2):388.

Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, FL. MR0798597.

Higgins, J., Thompson, S., Deeks, J., and Altman, D. (2002). Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *Journal of Health Services Research & Policy*, 7(1):51.

Higgins, J., Thompson, S., Deeks, J., and Altman, D. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, 327(7414):557–560.

Hine, L. K., Laird, N., Hewitt, P., and Chalmers, T. C. (1989). Meta-analytic evidence

against prophylactic use of lidocaine in acute myocardial infarction. *Archives of internal medicine*, 149(12):2694.

Hoaglin, D. C., Hawkins, N., Jansen, J. P., Scott, D. A., Itzler, R., Cappelleri, J. C., Boersma, C., Thompson, D., Larholt, K. M., Diaz, M., and Barrett, A. (2011). Conducting indirect-treatment-comparison and network-meta-analysis studies: Report of the ispor task force on indirect treatment comparisons good research practices: Part 2. *Value in Health*, 14(4):429–437.

Huber, P. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.

Hunter, J. and Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications, Inc.

J Sweeting, M., J Sutton, A., and C Lambert, P. (2004). What to add to nothing? use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in medicine*, 23(9):1351–1375.

Jackson, D., Riley, R., and White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*, 30(20):2481–2498.

Jackson, D., White, I., and Thompson, S. (2010). Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Statistics in medicine*, 29(12):1282–1297.

Jansen, J., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., Lee, K., Boersma, C., Annemans, L., and Cappelleri, J. (2011). Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ispor task force

on indirect treatment comparisons good research practices: part 1. *Value in Health-Journal of the Inter Soc Pharmacoeconomics*, 14(4):417–428.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, pages 805–820.

Joseph, L., Du Berger, R., and Belisle, P. (1997). Bayesian and mixed bayesian/likelihood criteria for sample size determination. *Statistics in Medicine*, 16(7):769–781.

Li, T., Puhan, M. A., Vedula, S. S., Singh, S., Dickersin, K., et al. (2011). Network meta-analysis-highly attractive but more methodological research is needed. *BMC medicine*, 9(1):79.

Littell, R. C. and Folks, J. L. (1973). Asymptotic optimality of fisher's method of combining independent tests ii. *Journal of the American Statistical Association*, 68(341):193–194.

Liu, D. (2012). *Combining information for heterogeneous studies and rare events studies: a confidence distribution approach*. PhD thesis, Rutgers University.

Liu, D., Liu, R., and Xie, M. (2013). Exact meta-analysis approach for discrete data and its application to 2x2 tables with rare events. *Preprint*.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414.

Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh). *The Annals of Statistics*, 27(3):783–858.

Lu, G. and Ades, A. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*, 23(20):3105–3124.

Lu, G. and Ades, A. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, 101(474):447–459.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*, 21(16):2313–2324.

Marden, J. I. (1991). Sensitive and sturdy p-values. *The Annals of Statistics*, pages 918–934.

Martin, R. and Liu, C. (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *Journal of the American Statistical Association*, 108(501):301–313.

Morris, C. (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American Statistical Association*, pages 47–55.

Nissen, S. E. and Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine*, 356(24):2457–2471.

Normand, S. (1999). Tutorial in biostatistics meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18(3):321–359.

Pagliaro, L., D'Amico, G., Sörensen, T. I. A., Lebrec, D., Burroughs, A. K., Morabito, A., Tiné, F., Politi, F., and Traina, M. (1992). Prevention of first bleeding in cirrhosis. a meta-analysis of randomized trials of nonsurgical treatment. *Annals of Internal Medicine*, 117(1):59–70.

Potuschak, H. and Müller, W. G. (2009). More on the distribution of the sum of uniform random variables. *Statistical Papers*, 50(1):177–183.

Riley, R. D. (2009). Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(4):789–811.

Rousseeuw, P. J. and Ruts, I. (1996). Algorithm as 307: Bivariate location depth. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(4):516–526.

Rousseeuw, P. J. and Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, 8(3):193–203.

Sacks, H. S., Chalmers, T. C., Blum, A. L., Berrier, J., and Pagano, D. (1990). Endoscopic Hemostasis, an effective therapy for bleeding peptic ulcers. *JAMA: the journal of the American Medical Association*, 264(4):494–499.

Schweder, T. and Hjort, N. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics*, 29(2):309–332.

Sidik, K. and Jonkman, J. (2005a). A note on variance estimation in random effects meta-regression. *Journal of biopharmaceutical statistics*, 15(5):823–838.

Sidik, K. and Jonkman, J. (2005b). Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(2):367–384.

Sidik, K. and Jonkman, J. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine*, 26(9):1964–1981.

Simmonds, M. and Higgins, J. (2007). Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine*, 26(15):2982–2999.

Singh, K., Xie, M., and Strawderman, W. E. (2005). Combining information from independent sources through confidence distributions. *The Annals of Statistics*, 33(1):159–183.

Singh, K., Xie, M., and Strawderman, W. E. (2007). Confidence distribution (cd): distribution estimator of a parameter. *Lecture Notes-Monograph Series Vol. 54, Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, 54:132–150.

Smith, T., Spiegelhalter, D., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine*, 14(24):2685–2699.

Stettler, C., Wandel, S., Allemann, S., Kastrati, A., Morice, M., Schömig, A., Pfisterer, M., Stone, G., Leon, M., de Lezo, J., et al. (2007). Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *The Lancet*, 370(9591):937–948.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., and Williams Jr, R. M. (1949). The american soldier: Adjustment during army life, vol. i. *Studies in Social Psychology in World War II. Princeton: Princeton University Press.*

Sutton, A., Abrams, K., Jones, D., Jones, D., Sheldon, T., and Song, F. (2000). *Methods for meta-analysis in medical research.* J. Wiley.

Sutton, A. and Higgins, J. (2008). Recent developments in meta-analysis. *Statistics in medicine*, 27(5):625–650.

Thorlund, K., Wetterslev, J., Awad, T., Thabane, L., and Gluud, C. (2011). Comparison of statistical inferences from the dersimonian–laird and alternative random-effects model meta-analyses–an empirical assessment of 920 cochrane primary outcome meta-analyses. *Research Synthesis Methods*, 2(4):238–253.

Tian, L., Cai, T., Pfeffer, M. A., Piankov, N., Cremieux, P.-Y., and Wei, L. (2009). Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2× 2 tables with all available data but without artificial continuity correction. *Biostatistics*, 10(2):275–281.

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.

White, I. R. (2009). Multivariate random-effects meta-analysis. *Stata Journal*, 9(1):40.

Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review*, 81(1):2–39.

Xie, M., Singh, K., and Strawderman, W. (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association*, 106(493):320–333.

Yang, G., Liu, D., Liu, R., Xie, M., and Hoaglin, D. (2013a). A confidence distribution approach for an efficient network meta-analysis. *Manuscripts*.

Yang, G., Liu, D., and Xie, M. (2012a). Lu Tian's exact meta-analysis method (Tian et al. 2009) as a special example under the general framework of combining CDs. *Manuscripts, available at http://stat.rutgers.edu/home/gyang/researches/gmetaRpackage/*.

Yang, G., Liu, D., and Xie, M. (2013b). Combining multivariate normal confidence distributions and its application to multivariate meta-analysis. *Manuscripts*.

Yang, G., Shi, P., and Xie, M. (2012b). Mantel-Haenszel odds ratio and Peto's log odds ratio under the general framework of combining CDs. *Manuscripts, available at http://stat.rutgers.edu/home/gyang/researches/gmetaRpackage/*.

# Vita

## Guang Yang

**2009-2013** Ph.D. in Statistics and Biostatistics, Rutgers University

**2009-2011** M.Sc. in Computer Science, Rutgers University

**2007-2009** M.Sc. in Statistics and Biostatistics, Rutgers University

**2007-2009** M.Sc. in Microbiology and Molecular Genetics, Rutgers University

**2002-2006** B.Sc. in Biology, Fudan University, Shanghai, CHINA

**2009-2013** Teaching Assistant/Research Assistant, Department of Statistics and Biostatistics, Rutgers University

**2007-2009** Graduate Fellow/Research Assistant, Department of Microbiology and Molecular Genetics, Rutgers University