COMBINING INFORMATION FOR HETEROGENEOUS STUDIES AND RARE EVENTS STUDIES: A CONFIDENCE DISTRIBUTION APPROACH

BY DUNGANG LIU

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

Regina Liu and Minge Xie

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2012

ABSTRACT OF THE DISSERTATION

Combining Information for Heterogeneous Studies and Rare Events Studies: a Confidence Distribution Approach

by Dungang Liu

Dissertation Director: Regina Liu and Minge Xie

This dissertation develops efficient statistical methodologies for combining information from independent sources. The developments focus on two settings where the studies are heterogeneous or the studies involve rare events. In these settings, the conventional combining approaches often lead to inefficient or even invalid statistical inference. In this dissertation, we propose effective and efficient combining approaches using *confidence distributions*. The proposed approaches are justified both theoretically and numerically. They are also shown to be superior to the conventional approaches.

Combining information from multiple studies, often referred to as *meta-analysis* in the literature, has been used extensively in many fields, including health sciences, social sciences, and others. However, there remain many unresolved problems on how to effectively and efficiently combine information. For example,

- *Heterogeneous studies* – When the effect of interest is not estimable in heterogeneous studies (e.g., indirect evidence), how can we utilize these studies to

perform meta-analysis?

- *Rare events studies* – For clinical trials with rare events, how can we perform meta-analysis and incorporate studies with zero events in the analysis without using artificial continuity corrections or relying on large sample theory?

To address these challenging but recurrent problems, this dissertation develops new meta-analysis approaches based on combining confidence distributions. Roughly speaking, a confidence distribution refers to *a sample-dependent distribution function on the parameter space* with desirable inferential properties in terms of repeated sampling performance. It can be viewed as a frequentist counterpart to the posterior distribution in Bayesian inference. In this dissertation, we show the combination of confidence distributions has desirable properties which are lacking in the conventional approaches. Specifically, 1) in the presence of heterogeneous studies, the proposed approach integrates direct and indirect evidence and achieves asymptotic efficiency; 2) for rare event studies, the proposed approach yields exact inference and incorporates all the studies in the analysis without using artificial continuity corrections for zero events. These properties are demonstrated numerically in simulation studies and real data examples, including flight landing safety data collected by the Federal Aviation Administration and drug safety data collected in diabetes clinical trials.

Acknowledgements

I would like to express my gratitude to my advisors, Professor Regina Liu and Professor Minge Xie. They introduced me to many exciting research areas and provided me with invaluable guidance in my research. Without their continuous support and encouragement, the completion of this dissertation would not have been possible. In fact, their direction has helped me in many ways, far beyond my research projects. They even took time from their busy schedules to help me improve my communication and people skills. They also provided me with precious support in difficult times during these years.

I would like to thank the graduate director Professor John Kolassa, who has always been eager to help. He patiently and generously provided me with guidance year after year. I would also like to thank both Professor Donald R. Hoover and Professor Eun-Young Mun for their time and effort to serve on my thesis committee.

Finally, I am greatly indebted to the Department of Statistics and Biostatistics at Rutgers University for an excellent research environment, and continuous financial support throughout my graduate study with graduate fellowship, teaching assistantship and research assistantship. I will always remember the help and interaction with the faculty, staff and fellow students in this department.

Dedication

To my father

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

Meta-analysis provides a systematic methodology for combining information from independent scources. Most existing meta-analysis approaches make an overall inference by combining study-specific point estimators (see, e.g., Normand, 1999). However, these approaches are inefficient or even invalid in some recurrent settings, such as the presence of heterogeneous studies and rare events studies. There is a strong demand for efficient statistical methodologies for combining information. To this end, this dissertation proposes meta-analysis approaches using the concept of *confidence distributions*, and demonstrates their validity and usefulness with theories and real life applications.

In recent years, there has been increasing emphasis on evidence-based decision-making in sciences and practices (Lin and Zeng, 2010). Therefore, meta-analysis, as a methodology for evidence synthesis, is now used extensively in many fields, especially in medicine, epidemiology, sociology, and others. For example, for the research project "Analysis and Modeling Aircraft Landing Performance" sponsored by the Federal Aviation Administration (FAA), we need to develop a general guideline for identifying safe landing performance from the unsafe ones. This requires modeling the flight landing performance data by combining the observed flight landing data from aircrafts of different makes and models, e.g., Boeing 737, 757 or Airbus 320, 321, etc. We also consider another example in randomized clinical trial studies, for which it is a common practice to combine the data from different clinical centers in order to draw an overall conclusion on the treatment effect, such as drug efficacy and safety (see, e.g., Nissen and Wolski, 2007). In fact, there are numerous applications in many areas, from astronomy to zoology, where the benefits of meta-analysis have been recognized (Sutton

and Higgins, 2008).

Traditionally, meta-analysis is performed by weighting study-specific point estimators from relevant studies to form an overall inference for the parameter of interest. Such a combining approach, however, is inadequate for solving the following important problems that frequently arise in real life applications.

- **Problem 1 (Heterogeneous Studies).** In many occasions, the studies collected for meta-analysis are often found to be heterogeneous. The heterogeneity can arise from the differences among the studies in populations, designs, and outcomes (Sutton and Higgins, 2008). In the presence of heterogeneous studies, the parameter of interest may not be estimable in some of the studies, and thus the corresponding point estimates are not available. As a result, such studies are often excluded from the conventional meta-analysis. However, these studies contain indirect but valuable information for the inference. The question here is that: how can we incorporate in the analysis all the studies, including those heterogeneous studies?

- **Problem 2 (Rare Events Studies).** In meta-analysis of clinical trials with rare events, we are often confronted with a challenging situation, that is, the presence of *zero total event studies* where both treatment and control arms observe zero events. For such studies, the point estimates of some effect measures, such as the odds ratio and risk ratio, are undefined. As a result, the conventional meta-analysis approaches either exclude such studies from the analysis (e.g., Nissen and Wolski, 2007), or apply artificial continuity corrections to zero events. Both practices, however, are known to have undesirable consequences in inference (Tian et al., 2009; Sweeting et al., 2004). The question here is that: how can we utilize all available data without applying artificial continuity corrections for zero events?

To address the challenging problems above, this dissertation uses the concept of

confidence distribution to develop new approaches for combining information. Roughly speaking, the confidence distribution uses a sample-dependent distribution function on the parameter space to estimate an unknown parameter (see Appendix A for a brief review on the notions of confidence distribution). One attractive aspect of the confidence distribution is that it contains wealthy information for making frequentist inference. For example, let $H(\cdot) = H(\cdot; X)$ denote a confidence distribution (obtained based on the sample X) for the parameter θ, then the interval $(H^{-1}(\alpha), \infty)$ is a $100(1-\alpha)\%$ level one-sided confidence interval for θ. In fact, the way we draw inference from a confidence distribution is similar to the way we draw inference from a Bayesian posterior distribution. But confidence distribution is a pure frequentist concept, and it does not rely on any Bayesian reasoning. This concept subsumes a broad range of frequentist concepts. For example, we can show that likelihood function (after normalization), p-value function (or significance function; cf. Fraser, 1991) and bootstrap distribution can all be viewed as confidence distributions, under some mild conditions. Readers are referred to Xie and Singh (2012) for an in-depth review on recent developments of confidence distribution and its applications.

Since a confidence distribution carries much more information than a point estimator, it provides a useful device for combining information from multiple studies (Singh, Xie, and Strawderman, 2005; Xie, Singh, and Strawderman, 2011). In this dissertation, instead of combining point estimators, we propose to combine confidence distributions to make an overall inference for the parameter of interest. Specifically,

- to address Problem 1 involving heterogeneous studies, we obtain confidence distributions from maximum likelihood procedures associated with each of the studies. We propose to combine the obtained *confidence density functions* (density functions of confidence distributions). We show that the combining of confidence density functions enable us to integrate indirect evidence. Consequently, the proposed approach can incorporate in the analysis all the studies, including

heterogeneous studies. Under a general likelihood inference setting, we theoretically show that the proposed approach is asymptotically as efficient as the maximum likelihood approach using individual participant data (IPD) from all the studies. But unlike the IPD analysis, the proposed approach only uses summary statistics and does not require individual-level data. In addition, we show that our approach is robust in the sense that the resulting estimator remains consistent even if the covariance estimates from the studies are misspecified. These desirable properties are confirmed numerically in the analysis of the data simulated from the randomized clinical trials setting and the flight landing data collected by the FAA. These developments are presented in Chapter 2.

- to address Problem 2 involving rare events studies, we obtain *p-value functions* from the exact tests associated with each of the studies. These p-value functions can be viewed as confidence distributions in asymptotic sense. We propose to combine these p-value functions. This idea yields a new class of meta-analysis methods for making exact inference on the common parameter in a series of studies. In rare events settings, the proposed approach can utilize all available data without using any artificial corrections for zero events. For zero total event studies, this approach can capture the appreciable difference between the effects from large and small studies, distinguishing, for example, a zero total event study with 1000 cases and 1000 controls from that with 10 cases and 10 controls. An explicit formula is established for determining the type I error rate of the overall test from the tests associated with individual studies. This allows us to evaluate the performance of our approach and devise further adjustments to improve the power of the overall inference. In addition to the previous desirable small sample properties, our approach is also shown to be efficient in a large sample setting. Numerical examples using simulated and real data on the inference of odds ratio show that, in the setting of rare events, our approach is superior to the Mantel-Haenszel, Peto and classical conditional methods. These developments

are presented in Chapter 3.

The rest of this dissertation is organized as follows. In order to address Problem 1 that involves heterogeneous studies, we propose in Chapter 2 a new meta-analysis approach for combining summary statistics from independent studies. We show that the proposed approach has the following properties: i) it can integrate indirect evidence; ii) it does not have any asymptotic efficiency loss compared to the IPD analysis; iii) it is robust to misspecification of covariance estimates. In order to address Problem 2 that involves rare events studies, we propose in Chapter 3 a new meta-analysis approach for combining studies with discrete data. We show that the proposed approach has the following properties: i) it enables us to make exact inference for the common parameter across the studies; ii) in the rare events setting, it enables us to incorporate all available data in the analysis without using artificial continuity corrections for zero events; iii) it is efficient in a large sample setting, under some mild conditions. The dissertation is concluded in Chapter 4 with some additional remarks on the role confidence distributions have played in our developments. Additionally, we include in Appendix A a brief review on the notions of confidence distribution.

Chapter 2

An Efficient Meta-Analysis Approach Using Summary Statistics in the Presence of Heterogeneous Studies

2.1 Introduction

Traditionally, meta-analysis is performed by combining the summary statistics from relevant studies, such as weighting study-specific point estimates for the parameter of interest. On the other hand, collecting individual-level data from original studies is widely regarded as the "gold standard" approach to meta-analysis (Sutton and Higgins, 2008; Lin and Zeng, 2010). But retrieving individual-level data is formidable, unaffordable, and even impossible in many situations. A nature question is whether or not combining summary statistics has efficiency loss, compared to analyzing the original data on individual participants. This question is of both practical and theoretical importance, and it has been investigated in the literature in various settings. See, e.g., Olkin and Sampson (1998), Mathew and Nordstrom (1999), Simmonds and Higgins (2007), and Lin and Zeng (2010), among others. In particular, Lin and Zeng (2010) showed that, in general, there is no asymptotic efficiency loss by analyzing summary statistics under certain conditions. But in the presence of heterogeneous studies, the question on relative efficiency of analyzing summary statistics versus individual-level data becomes more challenging and remains unanswered.

In many practical occasions, the studies collected for meta-analysis are often found to be heterogeneous. As pointed out in Sutton and Higgins (2008), the heterogeneity may arise from the differences among the studies in i) populations, such as study-specific effects; ii) designs, such as missing covariate designs; or iii) outcomes, such

as mixed response types. All these types of heterogeneity lead to *parameter heterogeneity* among the studies in the sense that the estimable parameters are different from one study to another. In this situation, the parameter of meta-analytic interest may not be estimable in some of the studies. Since such studies do not provide direct inference information, such as point estimates, for the parameter of interest, they are often excluded from the conventional meta-analysis. Clearly, this practice can lead to non-negligible, or even substantial, loss of efficiency. To overcome this problem, we propose in this chapter a meta-analysis approach that can incorporate in the analysis all the studies, including heterogeneous studies. Under a general likelihood inference framework, which adapts to parameter heterogeneity, we theoretically show that the proposed approach is asymptotically as efficient as the maximum likelihood approach using individual-level data from all the studies. But unlike the IPD analysis, our approach only needs summary statistics from relevant studies and does not require the individual-level data.

To exemplify the problem discussed above, we consider the following fixed-effects linear model for K independent studies:

$$Y_{ij} = \alpha_i + \beta_1 X_{ij} + \beta_2 Z_{ij} + \beta_3 Z_{ij} X_{ij} + \varepsilon_{ij}, \quad i = 1, \ldots, K, \ j = 1, \ldots, n_i, \tag{2.1}$$

where Y_{ij} is the response for the j-th subject in the i-th study, X_{ij} denotes the treatment status ($1/0$ for treatment/control), Z_{ij} is the covariate of interest (e.g., drug dosage), and the noise variable $\varepsilon_{ij} \sim N(0, \sigma_i^2)$. Here, α_i's are the study-specific effects and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ is the common effect. This model is often used in meta-analysis of randomized clinical trials to examine the covariate effect, in addition to the treatment effect (Simmonds and Higgins, 2007). Using this model, Simmonds and Higgins (2007) investigated the power of different meta-analysis methods in detecting the treatment-covariate interaction effect β_3. They showed that the conventional meta-analysis method, simply weighting the point estimates of β_3 from each of the studies,

suffers loss of power, or equivalently, efficiency. Lin and Zeng (2010) showed that this loss of efficiency can be avoided if the point estimates of the vector parameter $\boldsymbol{\beta}$ are combined using the inverse of the corresponding covariance matrix as the weight. But both of these methods break down when heterogeneity is present among the studies, as illustrated in the following examples.

Example 2.1 (study-specific effects)**.** The heterogeneity in populations across the studies is often captured by study-specific parameters. These parameters are of interest when the study-specific inference is part of the research goal. For example, in medical research, it is often of interest to make inference for some subpopulations, in addition to the whole population. In this case, the study-specific parameters, such as α_i's in model (2.1), can represent different population characteristics, such as location, climate, race and others. However, it is important to note that the parameter α_1 is not estimable in Study 2, $\cdots$, Study K. Thus, these studies can not be utilized in the conventional meta-analysis for making inference on α_1.

Example 2.2 (missing covariate designs)**.** Missing covariate designs can be encountered frequently since the studies from different sources may not have exactly the same research goal. As a result, some studies may not design the covariate that is of current meta-analytic interest, as noted in Simmonds and Higgins (2007). For example, if Study 1 is not aimed to examine the effect of the covariate Z_{1j} (e.g., drug dosage), then the Z_{1j} values of all the subjects in this study are often controlled at a fixed level (e.g., same dosage), say $Z_{1j} \equiv z_1$. In this situation, model (2.1) becomes

$$Y_{1j} = (\alpha_1 + z_1\beta_2) + (\beta_1 + z_1\beta_3)X_{1j} + \varepsilon_{1j}, \quad j = 1, \ldots, n_1, \tag{2.2}$$

for Study 1. As Simmonds and Higgins (2007) pointed out, the interaction effect β_3 is not estimable here, and Study 1 can not be utilized in the conventional meta-analysis for making inference on β_3.

Example 2.3 (mixed response types)**.** A challenging problem in meta-analysis is how to combine studies with different response types. This problem arises when the studies have different data report policies. For example, even when the underlying outcome of interest is indeed continuous (e.g., blood loss in women labor), some studies may only report binary outcomes (e.g., "severe" or "not severe"), depending on the underlying continuous outcome exceeding a prefixed threshold or not (see, e.g., Whitehead, Bailey, and Elbourne, 1999). Suppose that Study 1 only reports binary response d_{1j} in such a way that $d_{1j} = 1$ if $y_{1j} \geq \tau_1$ and $d_{1j} = 0$ otherwise, where τ_1 is the prefixed threshold. It is straightforward to show that model (2.1) reduces to the following probit model:

$$\Pr(d_{1j} = 1) = \Phi\left(\frac{\alpha_1 - \tau_1}{\sigma_1} + \frac{\beta_1}{\sigma_1}X_{1j} + \frac{\beta_2}{\sigma_1}Z_{1j} + \frac{\beta_3}{\sigma_1}Z_{1j}X_{1j}\right), \quad j = 1,\ldots,n_1, \quad (2.3)$$

for Study 1. Again, the interaction effect β_3 is not estimable here, and Study 1 can not be utilized in the conventional meta-analysis for making inference on β_3.

Examples 2.1, 2.2 and 2.3 exemplify the situations when the parameter of interest is not estimable in some of the studies, due to the heterogeneity in populations, designs, and outcomes, respectively. Such studies do not provide summary information, such as point estimates, for the parameter of interest, and thus they are typically excluded from the conventional meta-analysis. But it is important to note that the problem here involves multiple parameters, and, through the correlation, the information for one parameter can potentially contribute to the inference for one another. In fact, heterogeneous studies contain *indirect* but valuable information, which can be absorbed in the analysis if individual-level data are available for all the relevant studies. In this situation, inference can be made based on multiplying individual-level likelihood functions from all the studies. Although this so-called IPD method is regarded as the "gold standard" for combining information in the literature, its implementation is difficult, costly, and often impractical for various reasons, including data confidentiality issues and reluctance of original researchers to release the full data.

Now, an important question has been raised: can we retain full efficiency using only summary statistics to perform meta-analysis in the presence of heterogeneous studies? The answer is affirmative. In this article, we propose a meta-analysis approach that can incorporate all the studies in the analysis. This approach only needs summary statistics, namely, the estimates of *estimable* parameters and the estimate of the corresponding covariance matrix. Under a general likelihood inference setting, which adapts to parameter heterogeneity, we theoretically show that the estimator derived from the proposed approach is asymptotically as efficient as the IPD estimator. Hence, there is no asymptotic efficiency loss by analyzing summary statistics. This theoretical conclusion is broader than those established in Olkin and Sampson (1998), Mathew and Nordstrom (1999), and Lin and Zeng (2010), in the sense that the settings considered in those papers are special cases of the broad setting considered in the present chapter.

The proposed approach combines *confidence density functions* from relevant studies. The concept of confidence density and its cumulative counterpart, confidence distribution, has been developed extensively in recent years. See, e.g., Efron (1993, 1998), Schweder and Hjort (2002), Singh, Xie, and Strawderman (2005, 2007), and Xie, Singh, and Strawderman (2011). A comprehensive review for the modern definition and interpretation of this concept is provided in Xie and Singh (2012). Simply speaking, the confidence distribution (density) uses a sample-dependent distribution (density) function to estimate the parameter of interest. It is a pure frequentist concept, and it can be regarded as a frequentist counterpart to the posterior distribution in Bayesian inference. In the context of this chapter, we utilize a key feature of the confidence density function, that is, it can preserve, carry and integrate the correlation information among multiple effects. This advantage can also be seen in Tian et al. (2010), where the confidence density is used to make joint inference about a set of constrained parameters in survival analysis.

This chapter also establishes an important robustness property of our proposed approach. Under mild conditions, we show that our approach is still valid if the covariance matrices estimates obtained from the studies are misspecified. More specifically, the resulting estimator remains consistent and asymptotically normal with a "sandwich" limiting covariance matrix. This asymptotic result in essence follows the theory for *generalized estimating equations*, which is first proposed in Liang and Zeger (1986) and further elaborated in Xie and Yang (2003) for different asymptotic settings. The established robustness property ensures that heterogenous studies can still be analyzed using our approach in the situation of misspecification of covariance matrices in summary statistics. Therefore, indirect evidence can remain in our analysis to improve the efficiency of inference. The robustness property considerably broadens the applicability of our approach. In particular, we show that it applies to the situation when only estimates of the variances, rather than the full covariance matrices, of the parameter estimates are available from the studies.

The rest of this chapter is organized as follows. In Section 2.2, we set up a general likelihood inference framework that can adapt to parameter heterogeneity, propose a meta-analysis approach that can incorporate heterogeneous studies, and theoretically show that the proposed approach is asymptotically as efficient as the IPD approach. In Section 2.3, we establish a robustness property of our approach and illustrate one of its important applications. The theoretical results established in Section 2.2 and 2.3 are numerically confirmed in the analysis of i) the simulated data from a randomized clinical trials setting, as presented in Section 2.4, and ii) the real data on flight landing retrieved from FAA, as presented in Section 2.5. Finally, we provide a discussion in Section 2.6 on methodological and practical implications of the development in this chapter.

2.2 Methodology

2.2.1 A general likelihood inference framework

Consider K independent studies, with n_i participants in the i-th study. Under fixed-effects models, we denote the effects of interest in the K studies by a $p \times 1$ vector parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$. For the i-th study, suppose that the likelihood function $L_i(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$ only through a $p_i \times 1$ ($p_i \leq p$) vector parameter $\boldsymbol{\gamma}_i = \boldsymbol{f}_i(\boldsymbol{\theta})$, namely $L_i(\boldsymbol{\theta}) \equiv L_i^*(\boldsymbol{f}_i(\boldsymbol{\theta})) \equiv L_i^*(\boldsymbol{\gamma}_i)$ for any value of $\boldsymbol{\theta}$. Here, $\boldsymbol{f}_i : R^p \to R^{p_i}$ is a smooth function. We also assume that the likelihood function $L_i^*(\boldsymbol{\gamma}_i)$ is identifiable with respect to the parameter $\boldsymbol{\gamma}_i$ in the sense that it can not be written in terms of a smaller set of $\boldsymbol{\gamma}_i$. Typically, the function $\boldsymbol{f}_i$ is determined by the nature of the i-th study. For example, in model (2.1), we can let $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_K, \boldsymbol{\beta})^T$, $\boldsymbol{\gamma}_i = (\alpha_i, \boldsymbol{\beta})^T$, and the function $\boldsymbol{f}_i$ is self-apparent.

Let $\hat{\boldsymbol{\gamma}}_i = \arg\max_{\boldsymbol{\gamma}_i} L_i^*(\boldsymbol{\gamma}_i)$ be the maximum likelihood estimate of $\boldsymbol{\gamma}_i$ obtained from the i-th study. Denote the observed information matrix by $\Gamma_i(\boldsymbol{\gamma}_i) = -\partial^2 \log L_i^*(\boldsymbol{\gamma}_i)/\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^T$. Let $\hat{\Sigma}_i = \Gamma_i^{-1}(\hat{\boldsymbol{\gamma}}_i)$ be the estimated covariance matrix for $\hat{\boldsymbol{\gamma}}_i$. We treat $\hat{\boldsymbol{\gamma}}_i$ and $\hat{\Sigma}_i$ as summary statistics from the i-th study throughout this chapter, unless stated explicitly otherwise. We consider the asymptotics when K is fixed and n_i goes to infinity. Write $n = \sum_{i=1}^K n_i$, and assume that $n_i/n \to c_i \in (0,1)$ as $n \to \infty$. Assume that the regularity conditions in Chapter 6.5 of Lehmann and Casella (1998) hold. Then, $\Gamma_i(\boldsymbol{\gamma}_i)/n_i \to I_i$ in probability and $n_i^{-1/2}\{\partial \log L_i^*(\boldsymbol{\gamma}_i)/\partial \boldsymbol{\gamma}_i\} \to \mathrm{MN}(\mathbf{0}, I_i)$ in distribution, as $n_i \to \infty$. Here, I_i is the $p_i \times p_i$ Fisher information matrix, and "MN" stands for multivariate normal distribution. When nuisance parameters are in presence, $L_i^*(\boldsymbol{\gamma}_i)$ can be viewed as profile likelihood function, and the regularity conditions can be replaced by those in Murphy and van der Vaart (2000).

This likelihood inference framework enables us to investigate the relative efficiency of analyzing summary statistics versus individual-level data to a broad extent. It subsumes virtually all commonly used parametric and semiparametric models. It also

allows different likelihood functions among the studies, and the likelihood functions are not necessarily from regression models. More importantly, the framework here can adapt to a large scope of heterogeneous studies, since $\boldsymbol{f}_i$ can be any complex function satisfying some mild smoothness conditions. The key here is that we establish explicit connections between the estimable parameters $\boldsymbol{\gamma}_i$'s in different studies by using the "link functions" $\boldsymbol{f}_i$'s to link $\boldsymbol{\gamma}_i$'s to the same vector parameter $\boldsymbol{\theta}$. In fact, the setting considered in Lin and Zeng (2010) can be viewed as a special case of this framework when $\boldsymbol{f}_i$'s are all identical transformations. Thus, our methodological and theoretical development in the next section reaches far beyond the scope of Lin and Zeng (2010).

2.2.2 Combining confidence density functions

Based on the results in Chapter 5 of Singh et al. (2007), the density function of $\text{MN}(\hat{\boldsymbol{\gamma}}_i, \hat{\Sigma}_i)$ is a confidence density for the parameter $\boldsymbol{\gamma}_i$ asymptotically. In other words, we use the density function of $\text{MN}(\hat{\boldsymbol{\gamma}}_i, \hat{\Sigma}_i)$ as an *estimator* for the parameter $\boldsymbol{\gamma}_i$. Denote this density function by $h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i)$, where $\boldsymbol{S}_i$ represents the sample in the i-th study. Specifically,

$$h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i) = \frac{1}{(2\pi)^{p_i/2}|\hat{\Sigma}_i|^{1/2}} \exp\left\{\frac{1}{2}(\boldsymbol{\gamma}_i - \hat{\boldsymbol{\gamma}}_i)^T \hat{\Sigma}_i^{-1}(\boldsymbol{\gamma}_i - \hat{\boldsymbol{\gamma}}_i)\right\}, \quad i = 1, \ldots, K. \tag{2.4}$$

Singh et al. (2007) showed that this sample-dependent density function contains wealthy information for frequentist inference. For example, the "central region" of this density function is a confidence region for $\boldsymbol{\gamma}_i$ asymptotically, where the "central region" can be defined using the notion of data depth (see, e.g., Liu, Parelius, and Singh, 1999). Moreover, the marginal density function derived from $h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i)$ can be used to obtain confidence intervals for the corresponding component of $\boldsymbol{\gamma}_i$. It is important to note that the confidence density function $h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i)$ is constructed solely based on the summary statistics $\hat{\boldsymbol{\gamma}}_i$ and $\hat{\Sigma}_i$.

We propose to combine the confidence density functions $h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i),\ i = 1, \ldots, K$, in

the same way as we combine likelihood functions for inference. Specifically, let

$$h(\boldsymbol{\theta};\boldsymbol{S}_1,\ldots,\boldsymbol{S}_K)=\prod_{i=1}^{K}h_i(\boldsymbol{\gamma}_i;\boldsymbol{S}_i)=\prod_{i=1}^{K}h_i(\boldsymbol{f}_i(\boldsymbol{\theta});\boldsymbol{S}_i). \tag{2.5}$$

For notational ease, we suppress the samples $\boldsymbol{S}_1,\ldots,\boldsymbol{S}_K$ in all confidence density functions hereafter. For example, we write $h(\boldsymbol{\theta})\equiv h(\boldsymbol{\theta};\boldsymbol{S}_1,\ldots,\boldsymbol{S}_K)$ and $h_i(\boldsymbol{f}_i(\boldsymbol{\theta}))\equiv h_i(\boldsymbol{f}_i(\boldsymbol{\theta});\boldsymbol{S}_i)$. Now, we obtain a point estimator by maximizing the multiplied confidence density function $h(\boldsymbol{\theta})$, namely

$$\hat{\boldsymbol{\theta}}_{CD}=\arg\max_{\boldsymbol{\theta}}h(\boldsymbol{\theta}). \tag{2.6}$$

The asymptotic properties of $\hat{\boldsymbol{\theta}}_{CD}$ are presented in the following theorem.

Theorem 2.1. *Under some regularity conditions, the estimator $\hat{\boldsymbol{\theta}}_{CD}$ obtained from (2.6) satisfies that:*

(a) The estimator $\hat{\boldsymbol{\theta}}_{CD}$ is consistent and asymptotically normally distributed:

$$n^{1/2}(\hat{\boldsymbol{\theta}}_{CD}-\boldsymbol{\theta})\xrightarrow{d}\mathrm{MN}\left(\mathbf{0},\left\{\sum_{i=1}^{K}c_iJ_i(\boldsymbol{\theta})^TI_iJ_i(\boldsymbol{\theta})\right\}^{-1}\right). \tag{2.7}$$

where $J_i(\boldsymbol{\theta})=\partial\boldsymbol{f}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T$ is the Jacobian of the function $\boldsymbol{f}_i$ with respect to $\boldsymbol{\theta}$.

(b) The covariance matrix of $n^{1/2}(\hat{\boldsymbol{\theta}}_{CD}-\boldsymbol{\theta})$ can be consistently estimated by $n\hat{\Sigma}_{CD}$, where

$$\hat{\Sigma}_{CD}=\left\{-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\log h(\hat{\boldsymbol{\theta}}_{CD})\right\}^{-1}. \tag{2.8}$$

On the other hand, if individual-level data are available, the IPD estimator can be obtained by maximizing the multiplied likelihood function $L(\boldsymbol{\theta})=\prod_{i=1}^{K}L_i(\boldsymbol{\theta})$, namely

$$\hat{\boldsymbol{\theta}}_{IPD}=\arg\max_{\boldsymbol{\theta}}L(\boldsymbol{\theta}). \tag{2.9}$$

Under some regularity conditions, the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$ is consistent and we show

in Appendix that it is asymptotically normally distributed:

$$n^{1/2}(\hat{\boldsymbol{\theta}}_{IPD}-\boldsymbol{\theta}) \xrightarrow{d} \text{MN}\left(\mathbf{0}, \left\{\sum_{i=1}^{K} c_i J_i(\boldsymbol{\theta})^T I_i J_i(\boldsymbol{\theta})\right\}^{-1}\right). \tag{2.10}$$

Moreover, the covariance matrix of $n^{1/2}(\hat{\boldsymbol{\theta}}_{IPD}-\boldsymbol{\theta})$ can be consistently estimated by $n\hat{\Sigma}_{IPD}$, where

$$\hat{\Sigma}_{IPD} = \left\{-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\log L(\hat{\boldsymbol{\theta}}_{IPD})\right\}^{-1}. \tag{2.11}$$

The results in (2.7)and (2.10) show that the estimators $\hat{\boldsymbol{\theta}}_{CD}$ and $\hat{\boldsymbol{\theta}}_{IPD}$ have the same limiting covariance matrix. Hence, we have established the following theorem.

Theorem 2.2. *The estimator $\hat{\boldsymbol{\theta}}_{CD}$ is asymptotically as efficient as the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$.*

Theorem 2.2 states that, under the general likelihood inference framework introduced in Section 2.2.1, analyzing summary statistics using our proposed approach has no asymptotic efficiency loss compared to analyzing individual-level data using the IPD approach.

In the literature, there is a great endeavor in examining the relative efficiency of using summary statistics versus IPD in meta-analysis. For some special settings, Olkin and Sampson (1998) and Mathew and Nordstrom (1999) theoretically showed that there is no efficiency loss by analyzing summary statistics. Lin and Zeng (2010) made the same conclusion under a general likelihood inference setting. But Lin and Zeng (2010) focused on a common parameter that is estimable in all the studies, and the meta-analysis approach under their investigation is essentially a linear weighting of point estimators. The development in this chapter is broader. It does not require that the parameter $\boldsymbol{\theta}$ be estimable in each of the studies. This allows us to include a broad class of heterogenous studies in the analysis. Moreover, the proposed approach of combining confidence density functions is fundamentally different from linearly weighting point estimators. It can process the summary statistics for $\boldsymbol{\gamma}_i = \boldsymbol{f}_i(\boldsymbol{\theta})$ efficiently for any

complex function $\boldsymbol{f}_i$ satisfying mild smoothness conditions. In fact, the main result in Section 2.1 of Lin and Zeng (2010) is a special case in this chapter by letting $\boldsymbol{f}_i$ be identity transformations of $\boldsymbol{\theta}$, for all $i = 1, \ldots, K$.

In a special case when $\boldsymbol{f}_i$'s are all linear functions of $\boldsymbol{\theta}$, the estimator $\hat{\boldsymbol{\theta}}_{CD}$ in (2.6) has an explicit solution:

$$\hat{\boldsymbol{\theta}}_{CD} = \left(\sum_{i=1}^{K} J_i^T \hat{\Sigma}_i^{-1} J_i \right)^{-1} \left(\sum_{i=1}^{K} J_i^T \hat{\Sigma}_i^{-1} \hat{\boldsymbol{\gamma}}_i \right). \tag{2.12}$$

Here $J_i = J_i(\boldsymbol{\theta}),\ i = 1, \ldots, K$, are deterministic matrices, not depending on $\boldsymbol{\theta}$. In this case, the estimator $\hat{\boldsymbol{\theta}}_{CD}$ in (2.12) is identical to the estimator derived using the multivariate generalized least squares approach proposed by Becker and Wu (2007). Hence, the multivariate generalized least squares approach can also be considered as a special case of our approach.

We have shown that $\hat{\boldsymbol{\theta}}_{CD}$ has the same asymptotic properties as $\hat{\boldsymbol{\theta}}_{IPD}$. Moreover, the formulas (2.8) and (2.11) for estimating corresponding covariance matrices are also similar. To close this section, we show that these similarities are not coincident. In fact, they stem from an inherent connection between the confidence density function $h_i(\boldsymbol{\gamma}_i) \equiv h_i(\boldsymbol{f}_i(\boldsymbol{\theta}))$ and the likelihood function $L_i(\boldsymbol{\theta})$, which is revealed in the next lemma.

Lemma 2.1. *The gradient of the log-confidence density function* $\log h_i(\boldsymbol{\gamma}_i) \equiv \log h_i(\boldsymbol{f}_i(\boldsymbol{\theta}))$, *with respect to* $\boldsymbol{\theta}$, *is asymptotically equivalent to the score function* $\boldsymbol{s}_i(\boldsymbol{\theta}) = \partial \log L_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$. *More precisely,*

$$\frac{\partial \log h_i(\boldsymbol{f}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = J_i(\boldsymbol{\theta})^T \hat{\Sigma}_i^{-1} (\hat{\boldsymbol{\gamma}}_i - \boldsymbol{f}_i(\boldsymbol{\theta})) = \boldsymbol{s}_i(\boldsymbol{\theta}) + o_p(1), \quad i = 1, \ldots, K. \tag{2.13}$$

Lemma 2.1 has an important implication; that is, solving the estimating equation

$$\sum_{i=1}^{K} \frac{\partial \log h_i(\boldsymbol{f}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \mathbf{0} \tag{2.14}$$

is asymptotically equivalent to solving the estimating equation

$$\sum_{i=1}^{K} \boldsymbol{s}_i(\boldsymbol{\theta}) = \mathbf{0}. \tag{2.15}$$

Since $\hat{\boldsymbol{\theta}}_{CD}$ is a solution to Equation (2.14) and $\hat{\boldsymbol{\theta}}_{IPD}$ is a solution to Equation (2.15), these two estimators have the same asymptotic properties with no wonder.

2.2.3 Efficiency gain of utilizing indirect evidence

The proposed approach of combining summary statistics can integrate direct as well as indirect evidence, whereas the conventional meta-analysis approach generally ignores indirect evidence. In this section, we theoretically examine the relative efficiency of our approach versus the conventional approach. We demonstrate that the utilization of indirect evidence can achieve asymptotic efficiency gain. Without loss of generality, we consider a setting where $\boldsymbol{\gamma}_i = (\alpha_i, \boldsymbol{\beta})$, for $i = 1, \ldots, K$. Here, α_i's are the study-specific parameters and $\boldsymbol{\beta}$ is the common vector parameter. This setting is used widely in parametric and semiparametric models. See, e.g., the examples in Simmonds and Higgins (2007) and Lin and Zeng (2010). It also subsumes Model (2.1) in the introduction section, and thus the results presented in this section apply to Example 2.1, 2.2 and 2.3.

First, suppose the study-specific parameters α_i's are of interest (see Example 2.1 for instance). The following corollary shows that our proposed estimator $\hat{\alpha}_{i,CD}$ is asymptotically more efficient than the study-specific estimator $\hat{\alpha}_i$. In the following, we include the IPD estimator as a benchmark in the comparison, and the notation "aVar" stands for asymptotic variance.

Corollary 2.1. *The asymptotic variances of the proposed estimator $\hat{\alpha}_{i,CD}$, the IPD estimator $\hat{\alpha}_{i,IPD}$ and the study-specific estimator $\hat{\alpha}_i$ satisfy that*

$$aVar(\hat{\alpha}_{i,CD}) = aVar(\hat{\alpha}_{i,IPD}) \leq aVar(\hat{\alpha}_i), \quad i = 1, \ldots, K. \tag{2.16}$$

Corollary 2.1 clearly shows the efficiency gain of our proposed estimator $\hat{\alpha}_{i,CD}$ over the study-specific estimator $\hat{\alpha}_i$. This implies that, for estimating the study-specific parameter α_i, the j-th study $(j \neq i)$ can also make contribution. This seems surprising considering that the j-th study does not involve the parameter α_i and it is independent of the i-th study. However, it is important to realize that the i-th and j-th studies share a common parameter $\boldsymbol{\beta}$, and that the information for α_i and $\boldsymbol{\beta}$ is often correlated within the i-th study. When the estimation of $\boldsymbol{\beta}$ is improved in the combination of multiple studies, the estimation of α_i can also be improved through the correlation. In other words, the common parameter $\boldsymbol{\beta}$ serves as a catalyst that enables borrowing information from other studies for estimating the study-specific parameter α_i in the i-th study. This phenomenon of borrowing strength from indirect evidence is not well appreciated in current meta-analysis practice, where the study-specific estimators $\hat{\alpha}_i$'s are often reported as the final estimators. Our simulation studies in Section 2.4 show that this ignorance of indirect evidence leads to substantial loss of efficiency. On the other hand, our proposed approach utilizes all the studies to estimate the study-specific parameter α_i in the i-th study, and Corollary 2.1 shows that the proposed estimator $\hat{\alpha}_{i,CD}$ is asymptotically as efficient as the IPD estimator $\hat{\alpha}_{i,IPD}$. This implies that our approach fully integrates the correlation between the summary statistics for α_i's and $\boldsymbol{\beta}$. The simulation studies in Section 2.4 and real data analysis in Section 2.5 show that $\hat{\alpha}_{i,CD}$ is numerically quite close to $\hat{\alpha}_{i,IPD}$.

Now, we consider the estimation of a common scalar parameter $\eta = g(\boldsymbol{\beta})$, where g is a scalar function of the common parameter vector $\boldsymbol{\beta}$ (see Example 2.2 and 2.3 for instance, where $\eta = \beta_3$). The conventional meta-analysis approach combines the study-specific estimators $\hat{\eta}_i = g(\hat{\boldsymbol{\beta}}_i)$ using $w_i = 1/\widehat{\text{aVar}}(\hat{\eta}_i)$ as the weights, provided the estimate $\hat{\eta}_i$ is available from the i-th study. More specifically, the conventional estimator $\hat{\eta}_{cvt} = \sum_{i=1}^{K} w_i \hat{\eta}_i / \sum_{i=1}^{K} w_i$. The following corollary shows that our proposed estimator is asymptotically more efficient than the conventional estimator $\hat{\eta}_{cvt}$.

Corollary 2.2. *The asymptotic variances of the proposed estimator $\hat{\eta}_{CD} = g(\hat{\boldsymbol{\beta}}_{CD})$, the*

IPD estimator $\hat{\eta}_{IPD} = g(\hat{\boldsymbol{\beta}}_{IPD})$ *and the conventional estimator* $\hat{\eta}_{cvt}$ *satisfy that*

$$aVar(\hat{\eta}_{CD}) = aVar(\hat{\eta}_{IPD}) \leq aVar(\hat{\eta}_{cvt}). \tag{2.17}$$

Corollary 2.2 shows the efficiency gain of our proposed estimator $\hat{\eta}_{CD}$ over the conventional estimator $\hat{\eta}_{cvt}$. The latter has efficiency loss because it does not fully utilize the correlation between the components of $\hat{\boldsymbol{\beta}}_i$ in the combination. This finding is also confirmed in Simmonds and Higgins (2007) and Lin and Zeng (2010). Corollary 2.2 also shows that our proposed estimator $\hat{\eta}_{CD}$ is asymptotically as efficient as the IPD estimator $\hat{\eta}_{IPD}$. This implies that our approach fully integrates the correlation information.

2.3 Robustness to misspecification of covariance matrices

In this section, we show that our proposed approach is robust in the sense that it remains valid even if the covariance matrices $\hat{\Sigma}_i,\ i = 1,\ldots,K$, are misspecified. The resulting estimator of $\boldsymbol{\theta}$ remains consistent with appropriately adjusted limiting covariance matrix. Moreover, heterogeneous studies can still be analyzed using our approach, and indirect information can be utilized to increase the efficiency of inference. The robustness property greatly enhances the flexibility of our approach in applications. In particular, we show that the robustness property has an important application to the situation when only the estimates of the variances, rather than the full covariance matrices, of $\hat{\boldsymbol{\gamma}}_i$ are reported.

Let $\hat{\Sigma}_{i,W}$ denote a "working" covariance matrix of $\hat{\boldsymbol{\gamma}}_i$ in the i-th study. Assume that $\hat{\Sigma}_{i,W}$ satisfies that i) it fulfills the requirement of being a covariance matrix; and ii) $(n_i\hat{\Sigma}_{i,W})^{-1} \to A_i$ in probability as $n_i \to \infty$, where A_i is a fixed matrix. In this section, we use $(\hat{\boldsymbol{\gamma}}_i, \hat{\Sigma}_{i,W})$ in place of $(\hat{\boldsymbol{\gamma}}_i, \hat{\Sigma}_i)$ as summary statistics and justify that our proposed approach remains valid. Specifically, denote by $\hat{\boldsymbol{\theta}}_W$ the new estimator obtained from (2.6) by replacing $\hat{\Sigma}_i$ with $\hat{\Sigma}_{i,W}$ in (2.5). The next theorem shows that $\hat{\boldsymbol{\theta}}_W$ is consistent

and asymptotically normally distributed with a "sandwich" covariance matrix.

Theorem 2.3. *Under some regularity conditions, the estimator $\hat{\boldsymbol{\theta}}_W$ is consistent and asymptotically normally distributed:*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{MN}(\mathbf{0}, \Delta), \tag{2.18}$$

where

$$\Delta = \left\{\sum_{i=1}^{K} c_i J_i(\boldsymbol{\theta})^T A_i J_i(\boldsymbol{\theta})\right\}^{-1} \left\{\sum_{i=1}^{K} c_i J_i(\boldsymbol{\theta})^T A_i I_i^{-1} A_i J_i(\boldsymbol{\theta})\right\} \left\{\sum_{i=1}^{K} c_i J_i(\boldsymbol{\theta})^T A_i J_i(\boldsymbol{\theta})\right\}^{-1}. \tag{2.19}$$

The covariance of $\hat{\boldsymbol{\theta}}_W$ given in Theorem 2.3 can be estimated by

$$\left\{\sum_{i=1}^{K} J_i(\hat{\boldsymbol{\theta}}_W)^T \hat{\Sigma}_{i,w}^{-1} J_i(\hat{\boldsymbol{\theta}}_W)\right\}^{-1} \left\{\sum_{i=1}^{K} J_i(\hat{\boldsymbol{\theta}}_W)^T \hat{\Sigma}_{i,w}^{-1} \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\gamma}}_i) \hat{\Sigma}_{i,w}^{-1} J_i(\hat{\boldsymbol{\theta}}_W)\right\} \left\{\sum_{i=1}^{K} J_i(\hat{\boldsymbol{\theta}}_W)^T \hat{\Sigma}_{i,w}^{-1} J_i(\hat{\boldsymbol{\theta}}_W)\right\}^{-1},$$

where $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\gamma}}_i) = \{\hat{\boldsymbol{\gamma}}_i - \boldsymbol{f}_i(\hat{\boldsymbol{\theta}}_W)\}\{\hat{\boldsymbol{\gamma}}_i - \boldsymbol{f}_i(\hat{\boldsymbol{\theta}}_W)\}^T$.

The proof of Theorem 2.3 is straightforward if we notice that the estimating equation (2.14) changes to

$$\sum_{i=1}^{K} \frac{\partial \log h_i(\boldsymbol{f}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{K} J_i(\boldsymbol{\theta})^T \hat{\Sigma}_{i,W}^{-1} (\hat{\boldsymbol{\gamma}}_i - \boldsymbol{f}_i(\boldsymbol{\theta})) = \mathbf{0}. \tag{2.20}$$

Since $\hat{\boldsymbol{\gamma}}_i$ is an asymptotically unbiased estimator of $\boldsymbol{f}_i(\boldsymbol{\theta})$, the solution to the above estimating equation holds consistency and asymptotic normality even if $\hat{\Sigma}_{i,W} \neq \hat{\Sigma}_i$. This robustness essentially follows the asymptotic theory for generalized estimating equations (GEE), that is, the solution to a GEE remains consistent and asymptotically normal when the second moment (covariance matrix) of the response is misspecified, provide only that the first moment (mean) of the response is correctly specified (see, e.g., Liang and Zeger, 1986; Fahrmeir and Tutz, 2001, pp 119-129; Xie and Yang, 2003). Note that, different from Liang and Zeger (1986), the result in Theorem 2.3 is establish when $n_i \to \infty$ but K is fixed. This asymptotic setting is considered in Xie

and Yang (2003). Similar to the GEE approach, although the specification of $\hat{\Sigma}_{i,W}$ does not alter the consistency, it determines the efficiency of the estimator $\hat{\boldsymbol{\theta}}_W$. Generally speaking, $\hat{\boldsymbol{\theta}}_W$ has higher efficiency when $\hat{\Sigma}_{i,W}$ is closer to $\hat{\Sigma}_i$. If $\hat{\Sigma}_{i,W} = \hat{\Sigma}_i$, $\hat{\boldsymbol{\theta}}_W = \hat{\boldsymbol{\theta}}_{CD}$ and it is asymptotically as efficient as the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$.

The robustness of our approach has important applications. In many publications, the authors report the estimates of the variances, rather than the full covariance matrix, for the corresponding point estimates of relevant parameters. In the following, we show that our approach can be applied given only the variance estimates, and furthermore, it can achieve greater efficiency with suitably chosen working correlation matrices. To see this more clearly, we let the working covariance matrix be

$$\hat{\Sigma}_{i,W} = \hat{V}_i^{1/2} R_{i,W}(\phi) \hat{V}_i^{1/2}, \tag{2.21}$$

where $\hat{V}_i = \text{diag}\{\hat{\Sigma}_i\}$ is a $p_i \times p_i$ diagonal matrix whose diagonal entries are the estimates of the variances of each component of $\hat{\boldsymbol{\gamma}}_i$, and $R_{i,W}(\phi)$ is a working correlation matrix fully characterized by a (possibly vector-valued) parameter ϕ. This specification of working covariance matrices is similar to that in Liang and Zeger (1986) proposed for longitudinal data analysis. Note that $\hat{V}_i$ is available from most publications. Theorem 2.3 guarantees that the resulting estimator $\hat{\boldsymbol{\theta}}_W$ is consistent with *any* choice of $R_{i,W}(\hat{\phi})$, provided that $\hat{\phi}$ is a consistent estimator of ϕ given $\boldsymbol{\theta}$. Here, $\hat{\phi}$ only needs be consistent for *some* ϕ, not the *true* ϕ_0, which does not exist for a working correlation matrix.

The working correlation matrix $R_{i,W}(\phi)$ can be chosen as a compromise between simplicity and loss of efficiency due to incorrect specification. Liang and Zeger (1986) provided several choices of $R_{i,W}(\phi)$ for longitudinal data analysis. In the context of meta-analysis, we provide in the following some choices of $R_{i,W}(\phi)$. Note that more elegant choice is possible based on the particular setting of the problem.

Example 2.4. Let $R_{i,W}(\phi) = R_{i,0}$, any given correlation matrix. For any $R_{i,0}$, the resulting estimator $\hat{\boldsymbol{\theta}}_W$ is consistent. Obviously, choosing $R_{i,0}$ closer to the true correlation matrix of $\hat{\boldsymbol{\gamma}}_i$ gives increased asymptotic efficiency. The simplest choice of $R_{i,0}$ is the identity matrix (i.e., using a working independence assumption), and in this case $\hat{\Sigma}_{i,W} = \hat{V}_i$. In our simulation studies, we show that the study designs, typically reported in the publications, can provide useful information for choosing $R_{i,0}$ to achieve greater efficiency.

Example 2.5. Let $R_{i,W}(\phi)$ be totally unspecified, that is, ϕ contains $p_i(p_i-1)/2$ parameters. Then $R_{i,W}$ can be estimated by

$$\hat{R}_{i,W} = \frac{1}{|\kappa(i)|} \sum_{j \in \kappa(i)} \hat{V}_j^{-1/2} \{\hat{\boldsymbol{\gamma}}_j - \boldsymbol{f}_j(\hat{\boldsymbol{\theta}}_W)\} \{\hat{\boldsymbol{\gamma}}_j - \boldsymbol{f}_j(\hat{\boldsymbol{\theta}}_W)\}^T \hat{V}_j^{-1/2}, \tag{2.22}$$

where $\kappa(i)$ is the index set that contains all the indexes j such that $\hat{\boldsymbol{\gamma}}_j$ has the same correlation structure as $\hat{\boldsymbol{\gamma}}_i$, and $|\kappa(i)|$ is the cardinality of the set $\kappa(i)$. Obviously, $\hat{R}_{i,W}$ is not consistent unless $|\kappa(i)|$ goes to infinity. This is also similar to the GEE approach where the number of clusters must go to infinity to yield a consistent estimator of the correlation (see, e.g., Liang and Zeger, 1986).

We stress that the robustness property ensures the utilization of heterogeneous studies in our analysis when correlation information is unavailable. Again, this is different from the conventional meta-analysis method which exclude those studies. Apparently, the utilization of all available information, including indirect evidence, can improve the overall inference. Our simulation studies in the next section show that the simplest independence working correlation matrices (i.e., identity matrices) yield unbiased estimators for all parameters of interest, including those non-estimable in individual studies. Furthermore, our simulation studies also show that well chosen working correlation matrices, such as those obtained from design-driven or data-driven methods, can lead to highly efficient inference.

To end this section, we point out that, when $\hat{\Sigma}_{i,W}$ is not equivalent to $\hat{\Sigma}_i$ asymptotically, the density function $h_i(\boldsymbol{\gamma}_i)$ in (2.4) with $\hat{\Sigma}_{i,W}$ replacing $\hat{\Sigma}_i$ is not a confidence density for the parameter $\boldsymbol{\gamma}_i$ asymptotically. When $\hat{\Sigma}_{i,W} = \hat{V}_i$, $h_i(\boldsymbol{\gamma}_i)$ can be regarded as a *pseudo* confidence density for the parameter $\boldsymbol{\gamma}_i$ in the same spirit of the pseudo likelihood function (Cox and Reid, 2004).

2.4 Simulation studies

We conduct simulation studies to numerically examine the theoretical results established in Section 2.2 and 2.3. We mimic meta-analysis of randomized clinical trials, and simulate $K = 3$ independent studies using model (2.1) as described in the section of introduction. For the i-th study, the treatment indicator X_{ij} is 1 with 50% probability, the covariate Z_{ij} (e.g., drug dosage) has three levels of 1, 2 and 5, and each level is assigned to $n_i/3$ subjects. In the following, we modify the design of Study 1 and make it different from the other two studies. Then, we analyze the three studies using the conventional meta-analysis method, the IPD method and the proposed method. The results are based on 1000 replicates when $n_1 = n_2 = n_3 = 150$ and the parameters $\alpha_1 = -1, \alpha_2 = 0, \alpha_3 = 1, \beta_1 = 1, \beta_2 = 2, \beta_3 = -1$, $\sigma_1 = 3, \sigma_2 = 4$ and $\sigma_3 = 3$.

For the first part of the simulation study, we let Study 1 have a missing covariate design, as described in Example 1. Specifically, we set the covariate Z_{1j} at a fixed level, say, $Z_{1j} \equiv 1$ for all $j = 1, \ldots, n_1$. This mimics the situation where a clinical trial is designed to solely examine the treatment effect and thus control the variable Z_{1j} to eliminate the covariate effect. In this case, as shown in Example 1, the model for Study 1 reduces to model (2.2). The estimable parameter $\boldsymbol{\gamma}_1 = (\alpha_1 + \beta_1, \beta_1 + \beta_3)^T$, and $\boldsymbol{\gamma}_1$ can be linked to $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3)^T$ through a self-apparent function $\boldsymbol{f}_1$. It is clear that the situation here fits into the general likelihood inference setting in Section 2.2.1, and Study 1 can be included in the analysis using the proposed approach. The analysis results are presented in Table 2.1.

Table 2.1: Meta-analysis in the presence of a study with missing covariate design

	Conventional Method			IPD Method			Proposed Method		
Parameters	Mean	SE	SEE	Mean	SE	SEE	Mean	SE	SEE
α_1	NA	NA	NA	-1.00	0.37	0.37	-1.00	0.37	0.37
α_2	0.01	0.86	0.86	0.01	0.51	0.51	0.01	0.51	0.51
α_3	1.01	0.64	0.65	1.01	0.46	0.46	1.01	0.46	0.47
β_1	0.98	0.73	0.73	0.99	0.51	0.50	0.99	0.51	0.50
β_2	2.00	0.17	0.16	2.00	0.16	0.15	2.00	0.16	0.15
β_3	-1.00	0.23	0.23	-1.00	0.20	0.20	-1.00	0.20	0.20

Remark: Mean = mean parameter estimates; SE = standard error; SEE = mean standard error estimates.

Table 2.1 shows that our proposed approach enables estimation of α_1 that is not estimable in Study 1. For all the slope parameters, our estimators are nearly unbiased, and the estimated standard errors are almost identical to the empirical standard errors. In comparison with the IPD method, our method yields virtually identical results in terms of point estimation and standard error estimation. This confirms that our method and the IPD method have similar numerical performance even when the sample sizes are moderate. On the other hand, Table 2.1 shows that the conventional method is deficient. It can not analyze Study 1 due to the heterogeneity, and it is not able to estimate the associated parameter α_1. The standard errors of the estimates for the study-specific parameters α_2 and α_3 and the common parameter β_1 are considerably larger than those obtained from our proposed method. This indicates a great loss of efficiency, resulting from the failure to fully utilize Study 1 and the correlation information.

As the second part of the simulation study, we let Study 1 have dichotomized responses, as described in Example 2.3. Specifically, we create binary responses d_{1j}

in such a way that $d_{1j} = 1$ if the observation $y_{1j} \geq 4$ and $d_{1j} = 0$ otherwise, for $j = 1, \ldots, n_1$. Then, we discard all original continuous responses y_{1j} and only keep the binary responses d_{1j} for analysis. This mimics the situation where a clinical center routinely reports "censored" outcomes instead of the original outcomes. In this case, as shown in Example 2, the model for Study 1 reduces to model (2.3). The estimable parameter $\boldsymbol{\gamma}_1 = ((\alpha_1 - 4)/\sigma_1, \beta_1/\sigma_1, \beta_2/\sigma_1, \beta_3/\sigma_1)^T$, and $\boldsymbol{\gamma}_1$ can be linked to $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \sigma_1)^T$ through a self-apparent function $\boldsymbol{f}_1$. Note that $\boldsymbol{f}_1$ is a non-linear function in this case. Again, the situation here fits into the general likelihood inference setting in Section 2.2.1, and Study 1 can be included in the analysis using our proposed approach. The analysis results are presented in Table 2.2.

Table 2.2: Meta-analysis in the presence of a study with dichotomized responses

	Conventional Method			IPD Method			Proposed Method			Dichotomization		
	Mean	SE	SEE	Mean	SE	SEE	Mean	SE	SEE	Mean	SE	SEE
α_1	NA	NA	NA	-0.96	0.63	0.60	-0.99	0.63	0.62	-1.13	0.73	0.72
α_2	-0.04	0.82	0.86	0.02	0.53	0.55	0.02	0.53	0.55	-0.11	0.76	0.76
α_3	1.05	0.64	0.65	1.02	0.50	0.50	1.02	0.50	0.51	0.91	0.67	0.67
β_1	0.99	0.72	0.73	1.00	0.63	0.64	0.98	0.63	0.65	1.05	0.84	0.84
β_2	1.99	0.17	0.16	2.00	0.16	0.16	1.99	0.16	0.16	2.08	0.31	0.28
β_3	-0.99	0.24	0.23	-1.00	0.21	0.21	-0.99	0.21	0.21	-1.05	0.36	0.33

Remark: Mean = mean parameter estimates; SE = standard error; SEE = mean standard error estimates.

Table 2.2 shows that our proposed method yields estimates for all the slope parameters, including α_1, and the numerical results are quite close to the IPD method. On the other hand, the conventional method can not utilize Study 1 due to the heterogeneity, and it is not able to estimate the associated parameter α_1. Moreover, it suffers substantial loss of efficiency in estimating some parameters, such as α_2, α_3 and

β_1. Detailed comparison between the conventional method, the IPD method and the proposed method can be summarized in a similar way to the first simulation study, and will not be repeated here. We also include in Table 2.2 the analysis results yielded from dichotomizing the continuous responses in Study 2 and 3. As discussed in Dominici and Parmigiani (2000), when combined inference is desired in the presence of continuous and dichotomous responses, a common practice is to dichotomize all continuous responses and then proceed as if all responses were binary. However, Table 2.2 shows that the dichotomization method has remarkably greater bias and standard error than the proposed method and the IPD method. This clearly indicates that dichotomization leads to substantial loss of information.

As the last part of the simulation study, we examine the performance of our proposed approach in the situation when we are given only variances estimates, rather than full covariance matrices estimates. For simplicity, we use the same setting as the first part of our simulation study but only assume the availability of $\hat{V}_i = \text{diag}\{\hat{\Sigma}_i\}$ instead of $\hat{\Sigma}_i$. In the implementation of our approach, we follow Example 2.4 and 2.5 in Section 2.3 to specify working correlation matrices. In particular, we consider the following three methods with a variety of implementation complexity: 1) independence method, i.e., using identity matrices as working correlation matrices; 2) design-driven method, i.e., using our "best guess" to set working correlation matrices based on the knowledge of the study designs; 3) data-driven method, i.e., using correlation matrices estimated from the observed data. More concretely, the design-driven method is implemented as follows. For Study 1 with model (2.2), we set two $n_1 \times 1$ vectors $\mathbf{1}_0 = (1,\ldots,1)^T$ and $\boldsymbol{x}_{1,0} = (0,1,\ldots,0,1)^T$. Then we let the off-diagonal entries of the 2×2 working correlation matrix $R_{1,0}$ be $r_{12} = r_{21} = \langle \mathbf{1}_0^{\perp}, \boldsymbol{x}_{1,0}^{\perp}\rangle / (\|\mathbf{1}_0^{\perp}\| \|\boldsymbol{x}_{1,0}^{\perp}\|)$. Here, $\langle\cdot,\cdot\rangle$ denotes the inner product, $\|\cdot\|$ denotes the Euclidean norm, and $\mathbf{1}_0^{\perp} = \mathbf{1}_0 - P(\mathbf{1}_0|\mathscr{L}(\boldsymbol{x}_{1,0}))$ is the residual after projecting $\mathbf{1}_0$ to the linear space $\mathscr{L}(\boldsymbol{x}_{1,0})$ spanned by $\boldsymbol{x}_{1,0}$. Note that $R_{1,0}$ is the true correlation matrix for $\hat{\boldsymbol{\gamma}}_1$ if $(\mathbf{1}_0, \boldsymbol{x}_{1,0})$ is the actual design matrix for model (2.2). For Study 2 and 3 with model (2.1), the working correlation matrices $R_{2,0}$ and

$R_{3,0}$ can be specified in a similar way. It is worth noting that this specification method is based on the prefixed design of randomized clinical trials, not the data observed after the experiment. Thus, $R_{i,0}$'s are considered as fixed as in Example 2.4. The data-driven method is implemented following Example 2.5. This method requires multiple studies that have the same correlation structure. For this method, we generate 10 independent copies for each of Study 1, 2 and 3 and this yields $K = 3 \times 10 = 30$ studies in total. The analysis results for all the three methods are presented in Table 2.3.

Table 2.3: Robust meta-analysis with a variety of working correlation matrices

	Independence		Design Driven		Data Driven	
Parameters	Bias	RE	Bias	RE	Bias	RE
α_1	0.00	0.97	0.00	1.00	0.00	0.97
α_2	-0.01	0.57	0.00	1.00	0.00	0.89
α_3	-0.02	0.72	-0.02	0.98	0.00	0.94
β_1	0.02	1.00	0.01	1.00	-0.01	0.98
β_2	0.01	0.94	0.00	1.00	0.00	0.95
β_3	-0.01	0.77	-0.01	1.00	0.00	0.95

Remark: RE = relative efficiency

Table 2.3 shows the bias and efficiency for $\hat{\boldsymbol{\theta}}_W$ obtained using working correlation matrices based on the independence method ($K = 3$), design-driven method ($K = 3$) and data-driven method ($K = 3 \times 10$). All the methods yield estimates for all the slope parameters, including α_1 that is not estimable in Study 1. Moreover, $\hat{\boldsymbol{\theta}}_W$ is unbiased whichever method is used. This numerically confirms our theoretical finding that the specification of working correlation matrices does not alter the consistency of our proposed estimator. Table 2.3 also presents the efficiency of $\hat{\boldsymbol{\theta}}_W$ relative to the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$. The independence method leads to substantial loss of efficiency

in estimating some parameters, such as α_2, α_3 and β_3. On the other hand, both the design-driven method and data-driven method are highly efficient. In particular, the design-driven method virtually does not have any efficiency loss. This implies that, for the self-designed studies, such as randomized clinical trials, the working correlation matrices derived from the design-driven method can be remarkably close to the true correlation matrices.

2.5 Real data example: flight landing data

Landing, as the last stage of a flight, is closely associated with both the safety and capacity concerns of an airport. A landing accident not only risks the passenger safety, but also disturbs the runway operation and reduces the airport capacity. According to Van Es (2005), among the most frequently reported flight landing accidents is the landing overrun, which occurs when a landing aircraft is unable to stop before the end of the runway. Van Es (2005) also pointed out that there appears to be a significant increase in overrun risk when an aircraft has long landing distance which refers to the distance between the beginning of the runway and the touch down point. Hence, it is of great importance to study how the landing distance can be impacted by the air-borne operation of an aircraft.

In our study, we obtain two data sets from FAA. One is for Airbus 321 with $n_1 = 6565$ observations and the other is for Boeing 737 with $n_2 = 15809$ observations. Each observation contains operational information of a flight during the landing period. To investigate how the landing distance is associated with the air-borne operation, we obtain the following linear model:

$$Y_{ij} = \alpha_i + \beta_1 X_{1ij} + \cdots + \beta_5 X_{5ij} + \beta_6 X_{6ij} + \beta_7 X_{5ij} X_{6ij} + \varepsilon_{ij}, \quad i = 1,2,\ j = 1,\ldots,n_i, \tag{2.23}$$

where Y_{ij} is the logarithm of landing distance for the j-th observation in the i-th data set ($i = 1$ for Airbus), $X_{1ij},\ldots,X_{6ij}$ are different measures of air-borne operation (e.g.,

height and speed) when the aircraft is passing over the beginning of the runway, and the noise variable $\varepsilon_{ij} \sim N(0, \sigma_i^2)$. The six measures included in model (2.23) are selected based on both expert opinions and statistical variable selection procedures. In the Airbus data set, the measure "flaps", i.e., the regressor X_6 in model (2.23), has a constant value, namely $X_{61j} \equiv 24$ for all j's. Thus, this study has a missing covariate design, and the estimable parameter $\boldsymbol{\gamma}_1 = (\alpha_1 + 24\beta_6, \beta_1, \ldots, \beta_4, \beta_5 + 24\beta_7)^T$. The situation is similar to Example 2.2.

Table 2.4 presents the analysis results for all intercept and slope parameters in model (2.23). Here, we include the results obtained from individual studies and different meta-analysis methods. Table 2.4 shows that the Airbus study does not provide estimates for the study-specific parameter α_1 and the slope parameters β_5, β_6 and β_7, due to the missing covariate design. Consequently, the conventional method can not utilize the Airbus study to improve the inference for those parameters. On the other hand, Table 2.4 shows that our proposed method provides estimates for all the parameters, including those that are not estimable in the Airbus study. Moreover, in comparison with the IPD method, our method yields almost identical point estimates and standard error estimates. This confirms once again that our method has similar numerical performance to the IPD method in the large sample setting. In Table 2.4, we also include under the column name "Robust" the analysis results obtained from our method given only variances estimates. More specifically, we use independence method to generate working correlation matrices. Table 2.4 shows that, in this situation, our method is still able to utilize both studies and yield estimates for all the parameters. For $\beta_1, \ldots, \beta_4$, although our robust method uses working independence assumption, the results are only slightly different from the IPD method. The underlying reason is that the corresponding regressors $X_1, \ldots, X_4$ are mildly correlated to others. But for $\alpha_1, \alpha_2, \beta_5, \ldots, \beta_7$, the differences are more noticeable because the corresponding regressors have stronger correlation with others.

Table 2.4: Meta-analysis of flight landing data from FAA – Part 1

	Individual Studies		Meta-Analysis Methods			
Parameters	Airbus 321	Boeing 737	Conventional	IPD	Proposed	Robust
α_1 (Airbus)	NA	NA	NA	4.10(0.15)	4.10(0.15)	3.83(0.20)
α_2 (Boeing)	NA	3.40(0.24)	NA	3.95(0.15)	3.95(0.15)	3.40(0.24)
β_1 (x_1)	-0.65(1.02)	-2.87(0.89)	-1.91(0.67)	-1.98(0.67)	-1.98(0.67)	-1.91(0.67)
β_2 (x_2)	7.83(0.13)	8.70(0.10)	8.38(0.08)	8.37(0.08)	8.37(0.08)	8.36(0.08)
β_3 (x_3)	2.27(0.10)	2.21(0.06)	2.23(0.05)	2.20(0.05)	2.19(0.05)	2.22(0.05)
β_4 (x_4)	-0.50(0.75)	-1.51(0.42)	-1.27(0.37)	-1.34(0.37)	-1.34(0.37)	-1.27(0.37)
β_5 (x_5)	NA	2.35(0.36)	NA	1.61(0.23)	1.61(0.23)	2.19(0.20)
β_6 (x_6)	NA	4.08(0.76)	NA	2.50(0.52)	2.50(0.52)	4.08(0.76)
β_7 ($x_5 : x_6$)	NA	-4.69(1.13)	NA	-2.40(0.76)	-2.40(0.76)	-5.08(0.83)

Remark: In the parentheses is the estimated standard error of the corresponding parameter estimate.

As a further study, we dichotomize the response variable in the Boeing study to mimic a possible situation where we have only binary evaluation of the landing distance, such as long-distance landing or not, rather than the actual landing distance. Specifically, we create binary responses d_{2j} in such a way that $d_{2j} = 1$ if $y_{2j} \geq \tau_2$ and $d_{2j} = 0$ otherwise, where τ_2 is the 90% quantile of the landing distance of Boeing aircraft. The observation of $d_{2j} = 1$ indicates long-distance landing of a flight. In our analysis, we suppose only the binary responses are available in the Boeing study. This situation is similar to Example 2.3. The estimable parameter $\boldsymbol{\gamma}_2 = ((\alpha_2 - \tau_2)/\sigma_2, \beta_1/\sigma_2, \ldots, \beta_7/\sigma_2)^T$.

Table 2.5 presents the analysis results for individual studies and meta-analysis results obtained from different methods. For individual studies, Table 2.5 shows that most of the slope parameters in model (2.23) are not estimable, due to the fact that the Airbus study has missing covariate design and the Boeing study has dichotomous

responses. In this situation, the conventional meta-analysis method is not able to combine any information at all. On the other hand, Table 2.5 shows that our method still provides estimates for all the intercept and slope parameters in model (2.23), including those that are not estimable in either study. For comparison purpose, we also provide the IPD estimates in Table 2.5. Clearly, both the point estimates and the corresponding standard error estimates from our method are quite close to those obtained from the IPD method. Moveover, given only variances estimates, our robust method is still able to combine the two studies and make inference for all the parameters. Table 2.5 shows that the results obtained using working independence correlation matrices remains close to the IPD estimates.

Table 2.5: Meta-analysis of flight landing data from FAA – Part 2

	Individual Studies		Meta-Analysis Methods			
Parameters	Airbus 321	Boeing 737	Conventional	IPD	Proposed	Robust
α_1 (Airbus)	NA	NA	NA	4.62(0.27)	4.63(0.27)	4.35(0.50)
α_2 (Boeing)	NA	NA	NA	4.47(0.27)	4.48(0.27)	3.57(0.65)
β_1 (x_1)	-0.65(1.02)	NA	NA	-0.97(0.89)	-0.93(0.89)	-1.37(0.91)
β_2 (x_2)	7.83(0.13)	NA	NA	7.67(0.12)	7.65(0.12)	7.83(0.13)
β_3 (x_3)	2.27(0.10)	NA	NA	2.22(0.08)	2.22(0.08)	2.26(0.09)
β_4 (x_4)	-0.50(0.75)	NA	NA	-0.25(0.58)	-0.18(0.58)	-0.22(0.59)
β_5 (x_5)	NA	NA	NA	1.12(0.38)	1.09(0.38)	1.81(0.49)
β_6 (x_6)	NA	NA	NA	0.16(1.00)	0.14(1.00)	1.93(2.05)
β_7 $(x_5 : x_6)$	NA	NA	NA	0.28(1.42)	0.37(1.43)	-3.52(2.06)

Remark: In the parentheses is the estimated standard error of the corresponding parameter estimate.

2.6 Discussion

In this chapter, we have proposed a meta-analysis approach based on combining confidence density functions. This approach only needs summary statistics from relevant

studies, but it is shown to be asymptotically as efficient as the IPD approach which requires individual-level data from all the studies. We also have shown that our propose approach adapts to a broad scope of heterogeneous studies. It enable us to incorporate indirect evidence in the analysis and achieve efficiency gain in the overall inference. Furthermore, we have established a robustness property of our proposed approach for the situation when the covariance estimates are misspecified. This property greatly enhances the potential of our approach in applications. All these methodological and theoretical developments are numerically confirmed in the stimulation studies and real data analysis.

The development in this chapter has important practical implications. Collecting IPD from original studies is widely regarded as the gold standard approach to meta-analysis. The IPD approach fully utilizes individual-level information, and it is efficient if the underlying model is correctly specified. Moreover, by accessing original data, one can enhance comparability among the studies with respect to inclusion/exclusion criteria, creation of subgroups, adjustments of covariates, and others (Lin and Zeng, 2010). Despite these obvious advantages, the majority of meta-analysis is not performed using the IPD approach (Sutton and Higgins, 2008). One practical limitation of carrying out an IPD analysis is that it requires individual-level data from all the studies included in the meta-analysis. It is well known that this requirement can not be easily fulfilled in practice. Furthermore, there is an ongoing debate on whether the benefits of using the IPD method can outweigh the tremendous cost of retrieving IPD from all relevant studies (Sutton and Higgins, 2008). Our development in this chapter clearly says that there is no need to retrieve IPD because the aforementioned benefits of the IPD approach can still be achieved using our proposed approach to analyze summary statistics. First of all, our approach of analyzing summary statistics is asymptotically as efficient as the IPD approach under a general likelihood inference framework. Second, since this framework adapts to a broad scope of heterogeneous studies, our results imply that, given only summary statistics, we can also

relax study inclusion criteria, create subgroups, and make adjustments of covariates in meta-analysis. These appealing properties of our approach have been seen numerically in our simulation and real data analysis.

Our proposed approach of combining confidence density functions provides a unified treatment for combining summary statistics. If the functions $\boldsymbol{f}_i$'s are all identical transformations of $\boldsymbol{\theta}$, our approach reduces to the approach of weighting the study-specific estimates using the inverse-covariance matrices as the weights. If the functions $\boldsymbol{f}_i$'s are all linear functions of $\boldsymbol{\theta}$, our approach is equivalent to the multivariate generalized least square approach proposed in Becker and Wu (2007). Clearly, Our approach subsumes the existing approaches of combining summary statistics. Furthermore, our approach is much broader in the sense that it allows the the functions $\boldsymbol{f}_i$'s to be any functions satisfying mild smoothness conditions. Even when $\boldsymbol{f}_i$'s are complex or irregular, our approach can still incorporate the corresponding studies in the analysis, and our theoretical results guarantee the asymptotic efficiency of our approach.

Clearly, our proposed approach provides a new alternative for *complex evidence synthesis*, which involves models that "incorporate evidence on multiple parameters and/or that specifically model data from different study designs" (Sutton and Higgins, 2008). In recent years, complex evidence synthesis has been gaining increasing interest, and one important development in this field is mixed treatment comparisons in randomized clinical trials. Most of the approaches developed in this field thus far are within the Bayesian framework (see, e.g., Ades and Sutton, 2006). Our proposed approach share the same accent as the Bayesian approaches in the sense that we also combine density functions. However, from the viewpoint of recent development on confidence distributions, confidence density functions use "sample-dependent density functions" to estimate the unknown but fixed parameters. This chapter have show that such "sample-dependent density functions" lead to desirable frequentist inference. Moreover, different from Bayesian inference procedures, our approach does not require specification of priors and Markov chain Monte Carlo procedures.

2.7 Appendix

2.7.1 Proofs for the results in Section 2.2.2

In this following, we provide technical details for the theoretical results in Section 2.2.2. First, we prove the asymptotic property in (2.10) for the IPD estimator. Second, we prove Lemma 2.1 which implies asymptotic equivalence between the IPD estimator and our proposed estimator. Then, the asymptotic properties of our proposed estimator, as shown in Theorem 2.1, hold automatically.

First, we prove (2.10) for the IPD estimator. Using Taylor expansion, we obtain

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\hat{\boldsymbol{\theta}}_{IPD}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) + \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log L(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{IPD} - \boldsymbol{\theta}) + O_p(1). \tag{2.24}$$

Notice that $\partial \log L(\hat{\boldsymbol{\theta}}_{IPD})/\partial \boldsymbol{\theta} = 0$ and it is easy to verify that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) = \sum_{i=1}^{K} J_i(\boldsymbol{\theta})^T \frac{\partial}{\partial \boldsymbol{\gamma}_i} \log L_i^*(\boldsymbol{\gamma}_i)$$

and

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log L(\boldsymbol{\theta}) = \sum_{i=1}^{K} J_i(\boldsymbol{\theta})^T \left\{ \frac{\partial^2}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^T} \log L_i^*(\boldsymbol{\gamma}_i) \right\} J_i(\boldsymbol{\theta}) = -\sum_{i=1}^{K} J_i(\boldsymbol{\theta})^T \Gamma_i(\boldsymbol{\gamma}_i) J_i(\boldsymbol{\theta}).$$

Plug the above results into Equation (2.24), and after some algebraic operations, we derive

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{IPD} - \boldsymbol{\theta}) = \left[\sum_{i=1}^{K} J_i(\boldsymbol{\theta})^T \left\{ \frac{\Gamma_i(\boldsymbol{\gamma}_i)}{n_i} \cdot \frac{n_i}{n} \right\} J_i(\boldsymbol{\theta}) \right]^{-1} \left[\sum_{i=1}^{K} J_i(\boldsymbol{\theta})^T \left\{ n_i^{-\frac{1}{2}} \frac{\partial}{\partial \boldsymbol{\gamma}_i} \log L_i^*(\boldsymbol{\gamma}_i) \right\} \left(\frac{n_i}{n} \right)^{\frac{1}{2}} \right] + o_p(1).$$

Following the conditions specified in Section 2.2.1, we can conclude that the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$ is asymptotically normally distributed as in (2.10).

Now, we prove Lemma 2.1. The first equation in (2.13) can be straightforwardly obtained by differentiation. We only show how to establish the second equation in (2.13). Since $\boldsymbol{s}_i(\boldsymbol{\theta}) = \partial \log L_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = J_i(\boldsymbol{\theta})^T \{\partial \log L_i^*(\boldsymbol{\gamma}_i)/\partial \boldsymbol{\gamma}_i\}$, we use Taylor

series to expand $\partial \log L_i^*(\boldsymbol{\gamma}_i)/\partial \boldsymbol{\gamma}_i$ around the consistent estimate $\hat{\boldsymbol{\gamma}}_i$ as follows

$$\frac{\partial}{\partial \boldsymbol{\gamma}_i} \log L_i^*(\boldsymbol{\gamma}_i) = \frac{\partial}{\partial \boldsymbol{\gamma}_i} \log L_i^*(\hat{\boldsymbol{\gamma}}_i) + \frac{\partial^2}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^T} \log L_i^*(\hat{\boldsymbol{\gamma}}_i)(\boldsymbol{\gamma}_i - \hat{\boldsymbol{\gamma}}_i) + o_p(1) = \hat{\Sigma}_i^{-1}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{f}_i(\boldsymbol{\theta})) + o_p(1).$$

Then, the second equation in (2.13) immediately follows. This completes the proof of Lemma 2.1.

In light of Equation (2.13), (2.14) and (2.15), our proposed estimator $\hat{\boldsymbol{\theta}}_{CD}$ have the same asymptotic properties as the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$. Hence, the results in Theorem 2.1 are implied by the established results.

2.7.2 Proofs for the results in Section 2.2.3

Without loss of generality, we assume $K = 2$. In this case, $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \boldsymbol{\beta}_{l\times 1}^T)^T$, and

$$J_1(\boldsymbol{\theta}) = J_1 = \begin{pmatrix} 1 & 0 & \mathbf{0}_{l\times 1}^T \\ 0 & 0 & \mathbf{1}_{l\times 1}^T \end{pmatrix}, \quad J_2(\boldsymbol{\theta}) = J_2 = \begin{pmatrix} 0 & 1 & \mathbf{0}_{l\times 1}^T \\ 0 & 0 & \mathbf{1}_{l\times 1}^T \end{pmatrix}.$$

From (2.7), the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{CD}$ is

$$\text{aVar}(\hat{\boldsymbol{\theta}}_{CD}) = \left(c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2\right)^{-1}.$$

Denote the inverses of the $(l+1)\times(l+1)$ matrices $c_1 I_1$ and $c_2 I_2$ by

$$(c_1 I_1)^{-1} = \begin{pmatrix} a_1 & \boldsymbol{b}_1^T \\ \boldsymbol{b}_1 & D_1 \end{pmatrix}, \quad (c_2 I_2)^{-1} = \begin{pmatrix} a_2 & \boldsymbol{b}_2^T \\ \boldsymbol{b}_2 & D_2 \end{pmatrix},$$

where D_1 and D_2 are $l \times l$ matrices. Under this setup, we prove Corollary 2.1 and Corollary 2.2 as follows.

First, we show $\text{aVar}(\hat{\alpha}_{i,CD}) \leq \text{aVar}(\hat{\alpha}_i)$ as in Corollary 2.1 for $i = 1$. It is easy to

see that $\text{aVar}(\hat{\alpha}_i) = a_1$, and

$$\text{aVar}(\hat{\alpha}_{1,CD}) = \left\{ \left(c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2 \right)^{-1} \right\}_{[1,1]},$$

where $M_{[1,1]}$ stands for the submatrix of M crossed by row 1 and column 1. Thus, it suffices to show that

$$\left\{ \left(c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2 \right)^{-1} \right\}_{[1,1]} \leq a_1. \tag{2.25}$$

By Lemma 2.2 in Appendix 2.7.3,

$$\left\{ \left(c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2 \right)^{-1} \right\}_{[1,1]} = (a_1 - \boldsymbol{b}_1^T D_1^{-1} \boldsymbol{b}_1) + \boldsymbol{b}_1^T D_1^{-1} (D_1^{-1} + D_2^{-1})^{-1} D_1^{-1} \boldsymbol{b}_1.$$

Using the results in Lemma 2.3 in Appendix 2.7.3, we obtain

$$\boldsymbol{b}_1^T D_1^{-1} (D_1^{-1} + D_2^{-1})^{-1} D_1^{-1} \boldsymbol{b}_1 \leq \boldsymbol{b}_1^T D_1^{-1} D_1 D_1^{-1} \boldsymbol{b}_1 = \boldsymbol{b}_1^T D_1^{-1} \boldsymbol{b}_1,$$

which leads to the establishment of (2.25). This completes the proof of Corollary 2.1.

Next, we show $\text{aVar}(\hat{\eta}_{CD}) \leq \text{aVar}(\hat{\eta}_{cvt})$ as in Corollary 2.2 for $\eta = g(\boldsymbol{\beta})$. For simplicity, we assume $\eta = g(\boldsymbol{\beta}) = \boldsymbol{\lambda}^T \boldsymbol{\beta}$, where $\boldsymbol{\lambda}$ is a l-dimensional vector. By Lemma 2.2 in Appendix 2.7.3,

$$\left\{ \left(c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2 \right)^{-1} \right\}_{[3:(l+2),3:(l+2)]} = (D_1^{-1} + D_2^{-1})^{-1}.$$

Thus, $\text{aVar}(\hat{\eta}_{CD}) = \boldsymbol{\lambda}^T (D_1^{-1} + D_2^{-1})^{-1} \boldsymbol{\lambda}$. On the other hand,

$$\text{aVar}(\hat{\eta}_{cvt}) = \left\{ (\boldsymbol{\lambda}^T D_1 \boldsymbol{\lambda})^{-1} + (\boldsymbol{\lambda}^T D_2 \boldsymbol{\lambda})^{-1} \right\}^{-1}.$$

It follows from Lemma 2.3 in Appendix 2.7.3 that $\text{aVar}(\hat{\eta}_{CD}) \leq \text{aVar}(\hat{\eta}_{cvt})$. This completes the proof of Corollary 2.2.

2.7.3 Some useful matrix results

Lemma 2.2. *Under the setup in Appendix 2.7.2, we have the following results:*

$$\left\{\left(c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2\right)^{-1}\right\}_{[1,1]} = (a_1 - \boldsymbol{b}_1^T D_1^{-1} \boldsymbol{b}_1) + \boldsymbol{b}_1^T D_1^{-1} (D_1^{-1} + D_2^{-1})^{-1} D_1^{-1} \boldsymbol{b}_1,$$

$$\left\{\left(c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2\right)^{-1}\right\}_{[3:(l+2),3:(l+2)]} = (D_1^{-1} + D_2^{-1})^{-1}.$$

Proof. Using the blockwise matrix inversion formula, we have

$$c_1 J_1^T I_1 J_1 = \begin{pmatrix} k_1 & 0 & -k_1 \boldsymbol{b}_1^T D_1^{-1} \\ 0 & 0 & \mathbf{0}^T \\ -k_1 D_1^{-1} \boldsymbol{b}_1 & \mathbf{0} & D_1^{-1} + k_1 D_1^{-1} \boldsymbol{b}_1 \boldsymbol{b}_1^T D_1^{-1} \end{pmatrix},$$

and similarly,

$$c_2 J_2^T I_2 J_2 = \begin{pmatrix} 0 & 0 & \mathbf{0}^T \\ 0 & k_2 & -k_2 \boldsymbol{b}_2^T D_2^{-1} \\ \mathbf{0} & -k_2 D_2^{-1} \boldsymbol{b}_2 & D_2^{-1} + k_2 D_2^{-1} \boldsymbol{b}_2 \boldsymbol{b}_2^T D_2^{-1} \end{pmatrix},$$

where $k_1 = 1/(a_1 - \boldsymbol{b}_1^T D_1^{-1} \boldsymbol{b}_1)$ and $k_2 = 1/(a_2 - \boldsymbol{b}_2^T D_2^{-1} \boldsymbol{b}_2)$. Therefore,

$$c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2 = \begin{pmatrix} k_1 & 0 & -k_1 \boldsymbol{b}_1^T D_1^{-1} \\ 0 & k_2 & -k_2 \boldsymbol{b}_2^T D_2^{-1} \\ -k_1 D_1^{-1} \boldsymbol{b}_1 & -k_2 D_2^{-1} \boldsymbol{b}_2 & D_1^{-1} + k_1 D_1^{-1} \boldsymbol{b}_1 \boldsymbol{b}_1^T D_1^{-1} + D_2^{-1} + k_2 D_2^{-1} \boldsymbol{b}_2 \boldsymbol{b}_2^T \end{pmatrix}.$$

Applying blockwise inversion formula to the upper-left block and the lower-right block of this matrix leads to the two desired equations. □

Lemma 2.3. *Suppose W_1 and W_2 are $q \times q$ positive definite matrices. Then, for any q-dimensional vector $\boldsymbol{v}$,*

$$\left(\boldsymbol{v}^T W_1^{-1} \boldsymbol{v}\right)^{-1} + \left(\boldsymbol{v}^T W_2^{-1} \boldsymbol{v}\right)^{-1} \leq \left\{\boldsymbol{v}^T (W_1 + W_2)^{-1} \boldsymbol{v}\right\}^{-1}. \tag{2.26}$$

This implies that $\boldsymbol{v}^T(W_1+W_2)^{-1}\boldsymbol{v} \le \boldsymbol{v}^T W_1^{-1}\boldsymbol{v}$.

Proof. Since $W_1 > 0$ and $W_2 > 0$, we can find a nonsingular matrix P such that $W_1 = P\text{diag}\{r_1,\ldots,r_q\}P^T$ and $W_2 = P\text{diag}\{u_1,\ldots,u_q\}P^T$, where $r_i > 0$ and $u_i > 0$ for all $i = 1,\ldots,q$. By redefining $\boldsymbol{v}$ as $P^{-1}\boldsymbol{v}$, it suffices to prove the lemma when $W_1 = \text{diag}\{r_1,\ldots,r_q\}$ and $W_2 = \text{diag}\{u_1,\ldots,u_q\}$. Denoting $\boldsymbol{v} = (v_1,\ldots,v_q)^T$, the inequality (2.26) becomes

$$\left(\sum_{i=1}^{q}\frac{v_i^2}{r_i}\right)^{-1} + \left(\sum_{i=1}^{q}\frac{v_i^2}{u_i}\right)^{-1} \le \left(\sum_{i=1}^{q}\frac{v_i^2}{r_i+u_i}\right)^{-1}.$$

After rearrangement, the above inequality can be equivalently written as

$$\sum_{i=1}^{q}\frac{v_i^2}{r_i}\sum_{j=1}^{q}\frac{v_j^2}{u_j} \ge \sum_{i=1}^{q}\frac{v_i^2}{r_i+u_i}\sum_{j=1}^{q}\frac{v_j^2(r_j+u_j)}{r_j u_j}.$$

Thus, it suffices to show that, for any i and j,

$$\frac{v_i^2}{r_i}\frac{v_j^2}{u_j} + \frac{v_j^2}{r_j}\frac{v_i^2}{u_i} \ge \frac{v_i^2}{r_i+u_i}\frac{v_j^2(r_j+u_j)}{r_j u_j} + \frac{v_j^2}{r_j+u_j}\frac{v_i^2(r_i+u_i)}{r_i u_i}.$$

With some algebraic simplification, we can show that the above inequality is equivalent to $r_i^2u_j^2 + r_j^2u_i^2 \ge 2r_ir_ju_iu_j$, which holds from the Cauchy-Schwartz inequality. □

Chapter 3

An Exact Meta-Analysis Approach and its Application to 2×2 Tables with Rare Events

3.1 Introduction

Another challenging problem in meta-analysis is *how to develop a general exact meta-analysis approach for discrete data, especially when the events of interest are rare and large sample theories do not apply*. Rare events data, for example, observing zero outcomes in binomial trials, may occur in either small-sample observational studies (e.g., Gastwirth, 1984) or large-sample trials with very low event rates (e.g., Nissen and Wolski, 2007). In the case of rare events, a single study is inadequate for drawing a reliable conclusion, but the conclusion can often be strengthened by using meta-analysis to synthesize results from a number of similar studies.

A challenging and recurrent situation in performing meta-analysis of rare event studies is that a non-negligible, sometimes even substantial, portion of the studies may observe no events in both treatment and control arms. These studies are generally referred to as *zero total event studies* (Sweeting, Sutton, and Lambert, 2004; Rücker, Schwarzer, Carpenter, and Olkin, 2009). For example, Nissen and Wolski (2007) performed a meta-analysis of clinical trials to examine the association between the diabetes drug Avandia (rosiglitazone) and the adverse events of myocardial infarction or cardiovascular death. Out of 48 studies included in their analysis, 10 studies do not observe any event of myocardial infarction and 25 studies do not observe any event of cardiovascular death. The conventional meta-analysis, including Mantel–Haenszel and Peto methods, either simply excludes such studies from the analysis (which is the

case in Nissen and Wolski, 2007), or applies continuity corrections to the 2×2 tables of these studies, as seen in most literature. The exclusion of any study raises the concern of possible loss of information, as noted in Tian et al. (2009) and Rücker et al. (2009). Intuitively, a zero total event study with 1000 cases and 1000 controls provides stronger evidence against any hypothesized effect than that with 10 cases and 10 controls. Hence, zero total event studies may reveal useful information through their sample sizes, and should not be discarded blindly. As for continuity corrections, Sweeting et al. (2004) provided compelling evidence showing that imputing arbitrary numbers to zero cells, depending on the numbers imputed, can result in very different conclusions. A natural question here is how to utilize all available data in meta-analysis without assigning arbitrary numbers to zero cells.

Recently, using the risk difference as the effect measure, Tian et al. (2009) developed an exact meta-analysis procedure that can utilize all available data without artificial corrections for zero events. Their approach combines confidence intervals for the risk difference obtained from each of the studies. But, for the odds ratio or risk ratio, an exact meta-analysis procedure that can utilize all available data without artificial corrections remains unavailable thus far. The difficulty stems from the fact that, in a zero total event study, the sample odds ratio or risk ratio is not well defined, and thus its corresponding point estimate or interval estimate is not readily available. Instead of working on point or interval estimates, this chapter develops a new meta-analysis approach by combining *p-value functions* (also known as *significance functions*; cf. Fraser, 1991) obtained from the exact tests associated with the individual studies. This is a general approach that can yield exact inference for any parameter of interest, including the odds ratio, risk ratio or risk difference. It uses all available studies, including zero total event studies, without continuity corrections. To our knowledge, this is the first successful attempt to have achieved this goal in the literature on exact meta-analysis of the odds ratio and risk ratio under a general setting. Here, the term "exact" refers to the use of exact distributions, rather than limiting distributions, in making inference

through a meta-analysis procedure.

A p-value function, is formed by computing p-values for a one-sided test with varying boundaries of the null hypothesis (e.g., Fraser, 1991). More specifically, consider a parameter of interest, say ψ. Let z denote the sample, and $p = p(\psi^*; z)$ denote a p-value computed based on a given test for the one-sided hypothesis testing $H_0 : \psi = \psi^*$ versus $H_1 : \psi > \psi^*$. The p-value $p = p(\psi^*; z)$ depends on both the sample z and the value of ψ^*. Given the sample z, as the value of ψ^* varies, $p(\psi^*) \equiv p(\psi^*; z)$ is a function on the parameter space of ψ. This sample-dependent function $p(\cdot)$ is called a p-value function. Under some mild conditions, a p-value function is typically a distribution function on the parameter space. From the viewpoint of the recent developments on the so-called confidence distributions, this p-value function can be viewed as a "distribution estimator" of the unknown parameter, in the sense that a sample-dependent distribution function, rather than a point or an interval, is used to estimate the parameter. The "distribution estimator" carries much more information than a point or an interval estimator, and it provides a useful device for combining information from multiple studies, see the review article by Xie and Singh (2012) and the references therein for recent developments on confidence distributions and their applications.

The proposed meta-analysis method in this chapter is fundamentally different from the combining of point estimates used in the conventional meta-analysis. Specifically, we combine the p-value functions using the combining framework developed in Xie et al. (2011) for confidence distributions. But, in contrast to Xie et al. (2011) which mainly focused on continuous data and large-sample studies, this chapter shows that the idea of combining "functions" can lead to exact inference for discrete and small-sample data as well. In fact, the idea of combining p-value functions yields a broad class of exact meta-analysis methods, in the sense that the p-value functions associated with each study are allowed to be obtained using *any* exact test. In the setting of rare event data where the data are discrete and the event rate is very low, exact tests are often conservative in the sense that the actual type I error rate is less than the nominal

rate, say α. In our proposed exact meta-analysis procedure, we provide an explicit formula for computing the actual type I error rate of the overall test based on those of the individual tests, as shown in Theorem 3.1 in Section 3.3.1. This formula makes it feasible to estimate the overall type I error rate directly from the observed data, and to devise adjustments of the p-value functions to further increase the power of the overall test.

To a certain extent, our approach may be viewed as a generalization of the classical combining p-values approach (see, e.g., Fisher, 1932; Stouffer et al., 1949 and others). However, unlike the classical approach which is to combine the observed p-values, our approach is to combine the entire p-value *functions*. Moreover, the classical approach uses only the equal weights in the combination, which is known to be inefficient in terms of preserving Fisher information. In comparison, our approach can afford flexible weights in the combination. In fact, we show in Section 3.5 that with suitably chosen weights our approach can substantially improve the small sample performance of combination, and, in the case when a large sample theory applies, the weighted combined estimator is asymptotically efficient.

We show in Sections 3.7 and 3.8 that the proposed method compares favorably with the existing methods in several aspects. First, most of the commonly used meta-analysis methods, Mantel–Haenszel and Peto methods included, are developed based on large sample theories. For rare event studies, these methods may lead to invalid inference even with moderately large sample sizes, as noted in Tian et al. (2009). Second, the proposed method involves only a simple explicit expression. Computationally, it is straightforward, especially compared with the conventional conditional inference approaches (e.g., Gart, 1970) in which the overall computational complexity increases exponentially, as the number of the studies increases. Third, also most important, in the presence of rare event studies, the proposed method can utilize all available data to perform exact meta-analysis for any parameter of interest, without using artificial continuity corrections for zero events. The method can capture the appreciable difference

between the effects from large and small studies, distinguishing, for example, a zero total event study with 1000 cases and 1000 controls from that with 10 cases and 10 controls. This clearly is not the case for the conventional approaches.

The rest of this chapter is organized as follows. We describe the new meta-analysis approach in Section 3.3.2, and study its small sample and large sample properties in Section 3.3 and Section 3.4, respectively. We provide in Section 3.5 useful guidelines for choosing appropriate weights to enhance the combination efficiency. In Section 3.6, we propose an adjustment of the p-value functions to increase the power of the overall test. Numerical studies using simulation and real data in Section 3.7 and Section 3.8, respectively, show that the proposed method outperforms the commonly used methods. Finally, we provide a discussion in Section 3.9. We point out that the proposed approach is applicable to general settings for exact inference on any common parameter in a series of independent studies, even though the motivating and illustrating examples throughout the chapter are in 2×2 tables.

3.2 Methodology

3.2.1 Problem setup

Consider K independent 2×2 tables formed by pairs of independent binomial random variables (X_i, Y_i) with sample sizes (n_i, m_i) and event rates (π_{1i}, π_{0i}), for $i = 1, \ldots, K$. Denote by x_i and y_i the observed numbers of events in the treatment and control arms of the i-th study, respectively. There are various effect measures for comparing the event rates π_{1i} and π_{0i}. Among the measures frequently used for comparing the event rates π_{1i} and π_{0i} are the odds ratio, risk ratio, and risk difference:

$$\mathrm{OR}_i = \frac{\pi_{1i}/(1-\pi_{1i})}{\pi_{0i}/(1-\pi_{0i})}, \quad \mathrm{RR}_i = \frac{\pi_{1i}}{\pi_{0i}}, \quad \mathrm{RD}_i = \pi_{1i} - \pi_{0i}, \quad i = 1, \ldots, K. \tag{3.1}$$

In meta-analysis and under a fixed effects setting, it is often assumed that an effect measure has a common value across all the studies, and it is of interest to make inference on this common parameter, denoted by ψ throughout this chapter. For example, Sweeting et al. (2004) reviewed meta-analysis methods for the common odds ratio $\psi = \mathrm{OR}_i$, for all $i = 1, \ldots, K$, and Bradburn et al. (2007) compared meta-analysis methods for the common risk difference $\psi = \mathrm{RD}_i$, for all $i = 1, \ldots, K$. Our goal is to develop an exact meta-analysis procedure for making inference on a common parameter ψ, being the common odds ratio, the common risk ratio or the common risk difference. Aside from the assumption of a common parameter ψ, we also assume that the fixed but unknown events rates π_{1i} and π_{0i} satisfy $0 < \pi_{1i} < 1$ and $0 < \pi_{0i} < 1$, and that they can be different from one study to another.

We are particularly interested in the setting where a significant portion of the sample entries (x_i, y_i) in the 2×2 tables are zeros. This occurs when the event rates (π_{1i}, π_{0i}) are very low or the sample sizes (n_i, m_i) are small. For a zero total event study where $x_i = y_i = 0$, the study-specific odds ratio estimate $\widehat{\mathrm{OR}}_i = \{x_i/(n_i - x_i)\}/\{y_i/(m_i - y_i)\}$ or risk ratio estimate $\widehat{\mathrm{RR}}_i = x_i/y_i$ is not well defined. Thus, zero total event studies can not be included in the conventional meta-analysis approaches that combine point estimates from individual studies, unless artificial corrections are applied to zero events. In Appendix, we briefly review some commonly used methods, including Mantel-Haenszel, Peto and classical conditional inference methods, for the odds ratio. From the formulas therein, it is easy to see that zero total event studies do not contribute to the inference in those methods. The same observation holds for the risk ratio.

3.2.2 The proposed exact meta-analysis approach

To make inference on the common parameter ψ in K independent 2×2 tables, we consider the following hypothesis testing

$$H_0 : \psi = \psi^* \text{ versus } H_1 : \psi > \psi^*, \tag{3.2}$$

where ψ^* is an arbitrary but fixed value on the parameter space. Suppose a p-value $p_i(\psi^*; x_i, y_i)$ is obtained based on an exact test from the i-th study. In the literature, various exact tests have been developed for testing the association in a 2×2 table. For example, Lydersen, Fagerland, and Laake (2009) recommended several exact tests, including Suissa and Shuster test (Suissa and Shuster, 1985), Fisher-Boschloo's test (Boschloo, 1970), and the mid-p adaptation of Fisher's exact test (Lancaster, 1961). More recently, Agresti and Min (2002) and Reiczigel, Abonyi-Tóth, and Singer (2008) proposed new exact tests for the odds ratio and risk ratio, respectively.

By varying the value of ψ^*, $p_i(\psi^*; x_i, y_i)$ becomes a p-value function. This p-value function $p_i(\psi^*; x_i, y_i)$ is obtainable regardless of whether the entries x_i and y_i are zeros or not. We propose to combine these p-value functions, denoted by $p_i(\cdot) \equiv p_i(\cdot; x_i, y_i)$, using the following recipe:

$$p_{(c)}(\psi) \equiv F_{(c)}\left[w_1 \Phi^{-1}\{p_1(\psi)\} + \cdots + w_K \Phi^{-1}\{p_K(\psi)\}\right]. \tag{3.3}$$

Here, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, $F_{(c)}(\cdot) = \Phi(\cdot/w_1) * \cdots * \Phi(\cdot/w_k)$ where $*$ stands for convolution, and w_i's are weights subject to $\sum_{i=1}^{K} w_i = 1$. As Xie et al. (2011) pointed out, the use of non-trivial weights makes the recipe (3.3) advantageous over the traditional p-value combination approaches, such as Stouffer's method (Stouffer et al., 1949). The latter is a special case of (3.3) with $w_i \propto 1$ for all $i = 1, \ldots, K$. In this chapter, we show that, by using informative weights, we can achieve asymptotic efficiency in the overall inference and

enhance small sample efficiency as well. A detailed discussion of the choice of weights is provided in Section 3.5.

We refer to $p_{(c)}(\psi)$ as the combined p-value function and use it for the overall inference on ψ. Specifically, we can use $M_n = p_{(c)}^{-1}(1/2)$ as a point estimator for ψ, and the intervals $(p_{(c)}^{-1}(\alpha),\infty)$ and $(p_{(c)}^{-1}(\alpha/2), p_{(c)}^{-1}(1-\alpha/2))$ as a $100(1-\alpha)\%$ one-sided and two-sided confidence intervals for ψ, respectively. Here $p_{(c)}^{-1}(\cdot)$ is the inverse function of $p_{(c)}(\cdot)$. Also, we can use $1-p_{(c)}(\psi^*)$ as the overall p-value for testing the hypotheses $H_0 : \psi = \psi^*$ versus $H_1 : \psi < \psi^*$, and $2\min\{p_{(c)}(\psi^*), 1-p_{(c)}(\psi^*)\}$ the overall p-value for testing the two-sided hypothesis $H_0 : \psi = \psi^*$ versus $H_1 : \psi \neq \psi^*$.

Singh et al. (2005) and Xie et al. (2011) developed a general framework to combine confidence distributions, and Xie et al. (2011) showed that this general framework can unify the classical p-value combination approaches and modern model-based meta-analysis approaches. In those developments, the elements for combining are required to be confidence distributions (or at least asymptotically). In the context of this chapter, this is equivalent to requiring that the statistic $p_i(\psi_0) \equiv p_i(\psi_0; X_i, Y_i)$ be U(0,1) distributed (or asymptotically) when ψ_0 is the true value of ψ. For some exact tests, such as the mid-p adaptation of Fisher's exact test, the statistic $p_i(\psi_0)$ is indeed U(0,1) distributed asymptotically as both $n_i \to \infty$ and $m_i \to \infty$. In this case, our approach can be viewed as a special case of Xie et al. (2011) for large sample inference. However, under the aforementioned asymptotic setting, the probability of having any zero entry in a 2×2 table is zero, and thus the asymptotic theory is not suitable for analyzing rare event data with zero events. Also, for small samples of discrete data, such as the case of rare events in 2×2 tables, the deviation of the distribution of $p_i(\psi_0)$ from U(0,1) distribution is not negligible. In fact, this deviation gauges precisely the loss of inference accuracy. In this chapter, we show that the general idea of combining "functions" described in Singh et al. (2005) and Xie et al. (2011) can still be used in the small sample setting, and we also justify the validity of using the combined function $p_{(c)}(\psi)$ for exact inference. More specifically, i) we show that the overall type I error rate can

be traced down to the the type I error rates associated with individual studies and it can be estimated readily from the data; ii) we show that zero total event studies have the effect of increasing uncertainty in the overall inference outcomes; iii) we provide a useful guideline for choosing the weights to increase the combination efficiency in finite sample settings; and finally iv) we propose an effective adjustment for individual p-value functions $p_i(\psi)$'s to reduce the conservatism of exact tests and thus increase the power of the overall test.

3.3 Small sample properties

3.3.1 The type I error rate

In the setting of rare events where the data are discrete and the event rates are very low, non-randomized test generally fails to achieve the nominal type I error rate. For exact tests, the actual type I error rate is typically lower than the nominal rate, resulting in some loss of test power. It is useful to investigate the actual type I error rate of the overall test derived from a meta-analysis procedure. For our proposed approach, we establish in this section an explicit formula for determining the actual type I error rate of the overall test from those of individual tests. We show that this formula makes it feasible to estimate the overall type I error rate directly from the observed data. We also show in Section 3.6 that this formula devises further adjustments of the p-value functions to increase the power of the overall test.

Let us define an overall type I error rate function for our proposed approach as follows:

$$R_{(c)}(s) = \Pr\{p_{(c)}(\psi_0) \leq s\}, \quad \text{for } 0 \leq s \leq 1,$$

which is the cumulative distribution function of the statistic $p_{(c)}(\psi_0)$. At $s = \alpha$, $R_{(c)}(\alpha)$ gives the overall type I error rate of the one-sided test $H_0 : \psi = \psi^*$ versus $H_1 : \psi > \psi^*$, when H_0 is true with $\psi^* = \psi_0$. The next theorem shows that the overall type I error rate

function $R_{(c)}(s)$ can be expressed explicitly in terms of the individual type I error rate functions $R_i(s) = \Pr\{p_i(\psi_0) \leq s\}$, $i = 1, \ldots, K$. This is possible because the combining formula in (3.3) is simple and explicit.

Theorem 3.1. *The overall type I error rate function $R_{(c)}(s)$ can be expressed as*

$$R_{(c)}(s) = s + \sum_{i=1}^{K} d_i(s)$$

where

$$d_i(s) = E\left(D_i\left[\Phi\left\{ \left(1 + \sum_{j \neq i} \frac{w_j^2}{w_i^2} \right)^{1/2} \Phi^{-1}(s) - \sum_{j \neq i} \frac{w_j}{w_i} \Phi^{-1}(B_{ij}) \right\} \right] \right). \tag{3.4}$$

Here, the functions $D_i(s) = R_i(s) - s$, $i = 1, \ldots, K$, and the expectation E is with respect to the random variables B_{ij} which are independent and of the following distributions: for any $0 \leq t \leq 1$, $\Pr(B_{ij} \leq t) = t$ if $j \leq i$, and $\Pr(B_{ij} \leq t) = R_i(t)$ if $j > i$.

Theorem 3.1 immediately yields the following corollary which shows that the overall deviation of the type I error rate $\{R_{(c)}(s) - s\}$ can be bounded using the bounds of individual deviations $\{R_i(s) - s\}$, $i = 1, \ldots, K$. Specifically, if the test of each study has a deflated type I error rate, the overall test will have a deflated type I error rate. Similarly, if the test of each study has an inflated type I error rate, the overall test will also have an inflated type I error rate.

Corollary 3.1. *Suppose there exists a set of fixed lower and upper bounds l_i and u_i such that*

$$l_i \leq R_i(s) - s \leq u_i, \quad i = 1, \ldots, K,$$

for any $0 \leq s \leq 1$. Then the overall type I error rate function $R_{(c)}(s)$ satisfies

$$\sum_{i=1}^{K} l_i \leq R_{(c)}(s) - s \leq \sum_{i=1}^{K} u_i,$$

for any $0 \le s \le 1$. *Specifically, if* $R_i(s) \le s$ *for* $0 \le s \le 1$ *and* $i = 1, \ldots, K$, *then* $R_{(c)}(s) \le s$. *Similarly, if* $R_i(s) \ge s$ *for* $0 \le s \le 1$ *and* $i = 1, \ldots, K$, *then* $R_{(c)}(s) \ge s$.

We can apply the results in Theorem 3.1 and Corollary 3.1 to evaluate the performance of our proposed approach and make further adjustments to improve the power of the overall inference. Theorem 3.1 shows that the overall type I error rate function can be traced down to the individual ones. Hence, if we can estimate the functions $R_i(s)$'s, then the estimate of $R_{(c)}(s)$ can be immediately derived from Theorem 3.1. In Section 3.5, we propose an empirical method for estimating π_{1i} and π_{0i}. This in turn allows us to estimate $R_i(s)$ and thus $R_{(c)}(s)$. If only the bounds for $\{R_i(s) - s\}$, $i = 1, \ldots, K$ are available, the bounds for $\{R_{(c)}(s) - s\}$ can be derived from Corollary 3.1. Furthermore, we can use the results here to improve the power of the overall test by tuning up the deflated type I error rates of the individual tests. More discussions on this point are given in Section 3.6. To the best of our knowledge, theoretical results similar to Theorem 3.1 are not known for Mantel–Haenszel, Peto or other conventional meta-analysis methods.

For the two-sided hypothesis testing $H_0 : \psi = \psi^*$ versus $H_1 : \psi \ne \psi^*$ where $\psi^* = \psi_0$, the actual type I error rate is

$$\Pr\left[2\min\{p_{(c)}(\psi_0), 1 - p_{(c)}(\psi_0)\} \le \alpha\right] = 1 - R_{(c)}(1-\alpha/2)^- + R_{(c)}(\alpha/2), \quad (3.5)$$

where $R_{(c)}(1-\alpha/2)^-$ is the limit of $R_{(c)}(s)$ when s approaches $(1-\alpha/2)$ from below. For the two-sided confidence interval $(p_{(c)}^{-1}(\alpha/2), p_{(c)}^{-1}(1-\alpha/2))$, the actual coverage probability is

$$\Pr\left\{\psi_0 \in (p_{(c)}^{-1}(\alpha/2), p_{(c)}^{-1}(1-\alpha/2))\right\} = R_{(c)}(1-\alpha/2)^- - R_{(c)}(\alpha/2). \quad (3.6)$$

Clearly, we can estimate the actual type I error rate (3.5) and the actual coverage probability (3.6) since they are expressed in terms of $R_{(c)}(s)$.

3.3.2 The effect of zero events

In this section, we discuss the general effect of including the studies with zero events in the analysis. We show that such studies increase the uncertainty in overall inference outcomes. This effect can be manifested using various exact tests for different effect measures, including the odds ratio and risk ratio. For convenience, we illustrate this effect using the mid-p adaptation of Fisher's exact test for the odds ratio. This test is advocated widely in the literature; see, e.g., Agresti (1992) (Section 8.1), Hwang and Yang (2001), Lydersen et al. (2009) and the references therein. Using this exact test, we can obtain the p-value function $p_i(\psi)$ for the odds ratio as follows:

$$p_i(\psi) \equiv p_i(\psi; x_i, y_i) = \Pr_\psi(X_i > x_i \mid T_i = t_i) + \frac{1}{2}\Pr_\psi(X_i = x_i \mid T_i = t_i), \quad i = 1, \ldots, K. \tag{3.7}$$

Here, conditional on $T_i = X_i + Y_i = t_i$, X_i follows the noncentral hypergeometric distribution: $\Pr_\psi(X_i = x \mid T_i = t_i) = \binom{n_i}{x}\binom{m_i}{t_i - x}\psi^x \big/ \sum_{v=L_i}^{U_i}\binom{n_i}{v}\binom{m_i}{t_i - v}\psi^v$, for $L_i \le x \le U_i$, where $L_i = \max(0, t_i - m_i)$ and $U_i = \min(n_i, t_i)$.

To facilitate the understanding of the combining effect, we first provide an intuitive interpretation of the p-value function (3.7). Generally speaking, since $p_i(\psi^*)$ is the p-value for right-tailed test $H_0 : \psi = \psi^*$ versus $H_1 : \psi > \psi^*$, a small value of $p_i(\psi^*)$ indicates low "support" for the null $H_0 : \psi = \psi^*$ but strong "support" for the alternative $H_1 : \psi > \psi^*$. Hence, the value of $p_i(\psi^*)$ indicates the "plausibility" that the true value of ψ is on the left side of ψ^*. Similar arguments can be found in the literature on inference using "distribution estimators", including, for example, Bayesian inference using a posterior distribution (see e.g. Gelman et al., 2004, Chapter 2), frequentist inference using a confidence distribution (see e.g. Xie and Singh, 2012, Section 3) and Fisher's fiducial inference using a fiducial distribution (see e.g. Kendall and Stuart, 1973, Chapter 21).

Figure 3.1 plots, at the log odds ratio scale, the p-value functions (3.7) for four representative types of studies: (a) both $x_i \neq 0$ and $y_i \neq 0$; (b) $x_i \neq 0$ and $y_i = 0$;

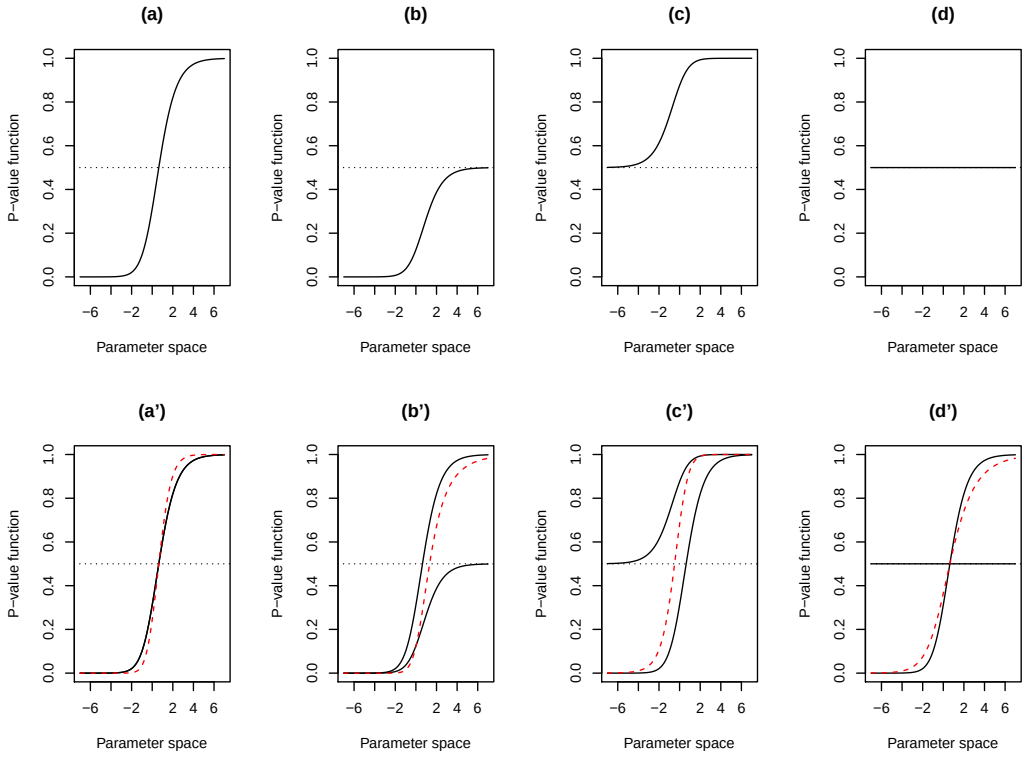

Figure 3.1: Individual p-value functions (upper row) and the combined p-value functions (lower row). The individual p-value functions illustrated in the upper row are for the cases: (a) $x_i = 2, y_i = 1$; (b) $x_i = 2, y_i = 0$; (c) $x_i = 0, y_i = 2$; and (d) $x_i = 0, y_i = 0$, all with the same sample sizes $n_i = m_i = 100$. The combined p-value functions (dashed curves) illustrated in the lower row are the resulting functions of combining an independent copy of the p-value function (a) with each of the p-value functions (a), (b), (c) and (d), using equal weights. The x-axes are in the logarithm scale.

(c) $x_i = 0$ and $y_i \neq 0$; and (d) both $x_i = 0$ and $y_i = 0$. In the case (a), the p-value function (3.7) is strictly increasing and forms a cumulative distribution function on the parameter space. For a given value ψ^* on the parameter space, if $p_i(\psi^*) < \alpha$, we reject the null $H_0 : \psi = \psi^*$ and favor the alternative $H_1 : \psi > \psi^*$. If $p_i(\psi^*) > 1 - \alpha$, we reject the null $H_0 : \psi = \psi^*$ and favor the alternative $H_1 : \psi < \psi^*$. The $(1 - \alpha)$ confidence interval $(\hat{\psi}_{\text{lo}}, \hat{\psi}_{\text{up}})$ for the true odds ratio is found by solving the equations $p_i(\hat{\psi}_{\text{lo}}) = \alpha/2$ and $p_i(\hat{\psi}_{\text{up}}) = 1 - \alpha/2$. In the case (b) with zero event in the control arm, the upper tail of the p-value function (3.7) degenerates to $1/2$. Consequently, the p-value function $p_i(\psi^*)$ is always less than $1/2$. Hence, we can not reject the null $H_0 : \psi = \psi^*$ and favor the alternative $H_1 : \psi < \psi^*$, for any ψ^*, and the upper end of the corresponding $(1 - \alpha)$ confidence interval is $\hat{\psi}_{\text{up}} = \infty$. These results are expected, because the observation $y_i = 0$ yields no information on how small the control event rate π_{0i} is, and thus how large the odds ratio ψ is. Similarly, in the case (c) with zero event in the treatment arm, the study yields no information on how small the treatment event rate π_{1i} is and thus how small the odds ratio ψ is. In this case, the lower tail of the p-value function (3.7) degenerates to $1/2$; the p-value function $p_i(\psi^*)$ is always greater than $1/2$; we can not reject the null $H_0 : \psi = \psi^*$ and favor the alternative $H_1 : \psi > \psi^*$, for any ψ^*; the lower end of the corresponding $(1 - \alpha)$ confidence interval is $\hat{\psi}_{\text{lo}} = 0$. In the case (d) with zero events in both control and treatment arms, the study yields no information on either how small or how large the odds ratio ψ is, which is reflected by a constant p-value function, namely $p_i(\psi^*) \equiv 1/2$, as shown in Figure 3.1(d).

In the remainder of this section, we show that the studies with zero events have the effect of increasing uncertainty in the overall inference outcomes. Without loss of generality, we consider combining $K = 2$ studies and illustrate the combining effects in the second row of Figure 3.1. Each figure of Figure 3.1(a') – (d') contains two solid curves: one is an identical copy of the solid curve shown in Figure 3.1(a) – (d), respectively; the other is an independent copy of the solid curve shown in Figure 3.1(a). The dashed curve in each figure plots the combined p-value function $p_{(c)}(\psi)$ resulting from

the combining of two studies represented by the two solid curves. Figure 3.1(a') shows that the conclusion is strengthened if two identical but independent studies are combined. This is reflected by the thinner tails of the combined p-value function on both sides. As a result, confidence intervals derived from the combined p-value function are tighter than those derived from each of the individual p-value functions. Figure 3.1(b') and (c') show that, if a zero event is observed in one and only one arm of a study, the tail of the combined p-value function $p_{(c)}(\psi)$ on the side corresponding to the non-zero event is fattened. As a result, the bound of a confidence interval corresponding to that side moves away from $\psi = 1$. Figure 3.1(d') shows that, when a zero total event study is included in the analysis, both tails of the combined p-value function $p_{(c)}(\psi)$ are fattened. To demonstrate the result in Figure 3.1(d') involving a zero total event study more explicitly, consider two independent p-value functions $p_1(\psi)$ and $p_2(\psi) \equiv 1/2$. Since $\Phi^{-1}\{p_2(\psi)\} \equiv \Phi^{-1}\{1/2\} = 0$, the combined p-value function becomes

$$p_{(c)}(\psi) = \Phi\left[\frac{w_1}{(w_1^2 + w_2^2)^{1/2}}\Phi^{-1}\{p_1(\psi)\}\right]. \tag{3.8}$$

It is clear that $p_1(\psi) > p_{(c)}(\psi) > 1/2$ when $p_1(\psi) > 1/2$, and $p_1(\psi) < p_{(c)}(\psi) < 1/2$ when $p_1(\psi) < 1/2$. Hence, both tails of $p_{(c)}(\psi)$ are fatter than those of $p_1(\psi)$, and confidence intervals derived from $p_{(c)}(\psi)$ are wider than those derived from $p_1(\psi)$. Thus, zero total event studies increase uncertainty in the overall inference outcomes, and subsequently any significant conclusion would be weakened if zero total event studies are incorporated in the analysis. This effect is also seen in our numerical studies in Section 3.8.

From the formula (3.8), the impact of a zero total event study on the overall inference is determined by its corresponding weight. In Section 3.5, we propose a weighting scheme that incorporates the sample size of a study in the weight formula, which distinguishes the impact of a zero total study with 1000 cases and 1000 controls from that with 10 cases and 10 controls.

3.4 Large sample properties

Note that our approach also applies to the general meta-analysis setting of 2×2 tables with no zero events. In this case, the large sample theory is often applicable. We now proceed to study the large sample properties of our proposed approach, which provides theoretical justification for our proposal in a large sample setting whenever it applies. More importantly, for rare events data, the result can help develop useful guidelines for choosing proper weights in finite sample settings, as shown in the next section.

Theorem 3.2 below states that, if the individual p-value functions $p_i(\psi)$, $i = 1, \ldots, K$, lead to accurate inference asymptotically, then so does the combined p-value function $p_{(c)}(\psi)$. In other words, the test derived from $p_{(c)}(\psi)$ achieves the nominal type I error rate asymptotically, and the confidence interval derived from $p_{(c)}(\psi)$ achieves the nominal coverage probability asymptotically.

Theorem 3.2. *If the individual type I error rate functions* $R_i(s) = \Pr\{p_i(\psi_0) \leq s\} \to s$ *for any* $0 \leq s \leq 1$ *and* $i = 1, \ldots, K$*, then the overall type I error rate function* $R_{(c)}(s) = \Pr\{p_{(c)}(\psi_0) \leq s\} \to s$ *for any* $0 \leq s \leq 1$.

Theorem 3.3 below states that, if the exact test associated with each study is equivalent to Wald test asymptotically (see, e.g., Fraser, 1991), then our approach can achieve asymptotic efficiency with suitably chosen weights.

Theorem 3.3. *Suppose that the p-value function* $p_i(\psi)$ *obtained from the exact test associated with the i-th study can be expressed as*

$$p_i(\psi) = \Phi\left[(\psi - \hat{\psi}_{i,MLE}) \Big/ \left\{\widehat{aVar(\hat{\psi}_{i,MLE})}\right\}^{1/2}\right] + o_p(1), \quad i = 1, \ldots, K, \qquad (3.9)$$

where $\hat{\psi}_{i,MLE}$ *is the maximum likelihood estimate (MLE) of* ψ *based on the i-th study, and* $\hat{\psi}_{i,MLE}$ *has the limiting variance* $aVar(\hat{\psi}_{i,MLE})$ *with the corresponding estimate* $\widehat{aVar(\hat{\psi}_{i,MLE})}$ *satisfying that the ratio* $\widehat{aVar(\hat{\psi}_{i,MLE})}/aVar(\hat{\psi}_{i,MLE})$ *converges to 1 in*

probability. Let the weights in the combining recipe (3.3) be

$$w_i \propto \{aVar(\hat{\psi}_{i,MLE})\}^{-1/2}, \quad i = 1, \ldots, K. \tag{3.10}$$

Then the median of the combined distribution function $p_{(c)}(\psi)$, namely $\hat{\psi}_c = p_{(c)}^{-1}(1/2)$, is consistent and asymptotically normally distributed as follows:

$$\left\{ \sum_{i=1}^{K} \frac{1}{aVar(\hat{\psi}_{i,MLE})} \right\}^{1/2} (\hat{\psi}_c - \psi_0) \to N(0,1). \tag{3.11}$$

The result above implies that our approach with the weights in (3.10) is as efficient as the maximum likelihood approach asymptotically. Specifically, the square of the normalizing constant in (3.11) satisfies that

$$\sum_{i=1}^{K} \frac{1}{\mathrm{aVar}(\hat{\psi}_{i,MLE})} = \frac{1}{\mathrm{aVar}(\hat{\psi}_{MLE})},$$

where $\hat{\psi}_{MLE}$ is the MLE obtained based on all the K studies (Lin and Zeng, 2010).

As an illustrative example for the statements in Theorems 3.2 and 3.3, we consider the p-value function $p_i(\psi)$ obtained from the mid-p adaptation of Fisher exact test in (3.7) for the odds ratio. We can show that $p_i(\psi_0; X_i, Y_i)$ converges to U(0,1) in distribution as both n_i and $m_i \to \infty$, provided that n_i/m_i is bounded away from 0 and ∞. Hence, under this condition, the combined p-value function $p_{(c)}(\psi)$ provides asymptotically accurate inference for the odds ratio. In addition, under the same condition, Breslow (1981) and Kou and Ying (1996) showed that the p-value function $p_i(\psi)$ can be expressed in the form of (3.9). Hence, the combined p-value function $p_{(c)}(\psi)$ with the weights in (3.10) leads to asymptotically efficient inference.

3.5 The choice of weights

In this section, we use the result in Theorem 3.3 to help develop proper weights for the combination in the finite sample setting. Specifically, we use the explicit formula of $\{\text{aVar}(\hat{\psi}_{i,MLE})\}^{-1/2}$ as the weights in the combining recipe (3.3). For example, Breslow (1981) showed that, when ψ is the common odds ratio, $\text{aVar}(\hat{\psi}_{i,MLE}) = \psi_0^2\left[\{n_i\pi_{1i}(1-\pi_{1i})\}^{-1} + \{m_i\pi_{0i}(1-\pi_{0i})\}^{-1}\right]$. Thus, we can use the weights

$$w_i \propto \left[\{n_i\pi_{1i}(1-\pi_{1i})\}^{-1} + \{m_i\pi_{0i}(1-\pi_{0i})\}^{-1}\right]^{-1/2}, \quad i = 1,\ldots,K, \tag{3.12}$$

to implement our approach for the odds ratio. This weighting scheme incorporates the sample sizes and the event rates of the studies, both of which are important factors in determining the amount of information contained in a study. For the risk difference, Tian et al. (2009) suggested to weight the studies based on the corresponding sample sizes. For the odds ratio, our simulation study (not reported here) shows that the weighting scheme in (3.12) substantially improves the efficiency over that solely based on the sample sizes.

In the weighting scheme, such as (3.12), π_{1i} and π_{0i} need to be estimated from the data. Clearly, the naive estimates of π_{1i} and π_{0i}, using the sample proportions, are not reliable in the setting of rare events. In the following, we propose a method to estimate π_{1i} and π_{0i} in the i-th study by borrowing information from the other studies. The idea here is similar to that in Efron (1996) as well as the Bayesian approaches using hierarchical models. Specifically, we make the working assumption that π_{0i} is a realization from a beta distribution $\text{beta}(\beta_1,\beta_2)$, noting that the beta distribution family is broad enough for capturing or approximating distributions of different shapes. The estimates of the parameters (β_1,β_2,ψ) are then obtained using the maximum likelihood method as follows:

$$(\hat{\beta}_1,\hat{\beta}_2,\hat{\psi}) = \arg\max_{(\beta_1,\beta_2,\psi)} \sum_{i=1}^{K} \log \int_0^1 f_\psi(x_i,y_i \mid \pi_{0i}) f_{\beta_1,\beta_2}(\pi_{0i})\mathrm{d}\pi_{0i}, \tag{3.13}$$

where $f_{\beta_1,\beta_2}(\pi_{0i}) = \pi_{0i}^{\beta_1-1}(1-\pi_{0i})^{\beta_2-1}/\int_0^1 \pi_{0i}^{\beta_1-1}(1-\pi_{0i})^{\beta_2-1}\mathrm{d}\pi_{0i}$, $f_\psi(x_i,y_i \mid \pi_{0i}) = c(x_i,y_i)\pi_{1i}^{x_i}(1-\pi_{1i})^{n_i-x_i}\ \pi_{0i}^{y_i}(1-\pi_{0i})^{m_i-y_i}$, and $\pi_{1i} = (\psi\pi_{0i})/(1-\pi_{0i}+\psi\pi_{0i})$ in the situation of a common odds ratio. We obtain the empirical conditional density of π_{0i}, namely $f_{\hat\beta_1,\hat\beta_2,\hat\psi}(\pi_{0i} \mid x_i,y_i) \propto f_{\hat\psi}(x_i,y_i \mid \pi_{0i}) f_{\hat\beta_1,\hat\beta_2}(\pi_{0i})$, by substituting the parameters (β_1,β_2,ψ) with their estimates $(\hat\beta_1,\hat\beta_2,\hat\psi)$. We then use the mean of this distribution, denoted by $\hat\pi_{0i}$, to estimate π_{0i} and the estimate of π_{1i} is $\hat\pi_{1i} = (\hat\psi\hat\pi_{0i})/(1-\hat\pi_{0i}+\hat\psi\hat\pi_{0i})$. This estimation method can apply to the situations of other common parameters, such as the risk ratio and others, with straightforward modification.

A particular difficult situation is when no events are observed in either one of the arms across all the studies; i.e., the situation when all entries $x_1 = x_2 = \ldots = x_K = 0$ or when all entries $y_1 = y_2 = \ldots = y_K = 0$. From the appendix in this chapter, we can see that Mantel–Haenszel method can not be applied in this case, unless continuity corrections for zero events are used. Our method does not have this problem, as the weight (3.12) can adapt to the case. For example, when all $x_i = 0$ $(i = 1,\ldots,K)$, we can show that $\hat\psi \to 0$ and $\hat\pi_{1i} \to 0$ $(i = 1,\ldots,K)$. In this situation, we use the limiting weights $\lim_{\psi\to 0}(w_i/\sum_{j=1}^K w_j)^2 = \{n_i\pi_{0i}/(1-\pi_{0i})\}/\{\sum_{j=1}^K n_j\pi_{0j}/(1-\pi_{0j})\}$ $(i = 1,\ldots,K)$. These limiting weights only depends on π_{0i} $(i = 1,\ldots,K)$, which can be estimated from the control arm as long as the observations y_i $(i = 1,\ldots,K)$ are not all zeros. The situation when all $y_i = 0$ $(i = 1,\ldots,K)$ can be handled similarly.

In our simulation studies in Section 3.7, we generate the event rate π_{0i} from uniform distributions U(0, ξ), where ξ is set to be some very small numbers. Clearly, such uniform distributions do not belong to the beta distribution family. Nevertheless, our simulation results show that the estimation method proposed above still performs well. We can view the working beta distribution assumption simply as a catalyst for borrowing strength from the other studies. In particular, if $x_i = y_i = 0$ in the i-th study, our estimation method still yields positive estimates of π_{1i} and π_{0i} if the other studies observe non-zero events. The magnitude of the non-zero estimates in a zero total event study is determined jointly by 1) the information borrowed from the other studies,

roughly speaking, the "average level" of the event rates in the other studies; and 2) the information provided by the i-th study itself, namely the sample sizes n_i and m_i. Our numerical results show that the estimates of π_{1i} and π_{0i} for a zero total event study with 1000 cases and 1000 controls are generally much smaller than those with 10 cases and 10 controls. This agrees with the fact that the true values of π_{1i} and π_{0i} for a zero total event study with 1000 cases and 1000 controls are more likely to be closer to zero than those with 10 cases and 10 controls.

3.6 Beta adjustment for individual p-value functions

It is well known that the exact test is often conservative for discrete data when the event rates are very low or sample sizes are not sufficiently large. In our context, this conservatism means that the actual type I error rate of the test associated with each study, namely $\Pr\{p_i(\psi_0) \leq \alpha\}$, can be far less than the nominal level α. This phenomenon is illustrated in Figure 3.2, where the cumulative distribution function of $p_i(\psi_0)$ is plotted using the solid curve. It is clear that the left tail of this curve is much lower than the 45 degree line, indicating that the actual type I error rate is much smaller than the nominal one. The conservatism of the exact tests associated with individual studies would inevitably pass on to the overall test, resulting in loss of the overall testing power. As shown in Section 3.3.1, our combining approach enables us to trace the overall type I error rate down to the type I error rates associated with each individual studies. This suggests the idea of improving the power of the overall test by reducing the conservatism of individual tests. In this section, we explore this idea and propose a simple adjustment for the individual p-value functions in (3.7). The proposed adjustment approach is quite general, as it is applicable to any exact tests as well as to settings beyond 2×2 tables. Note that our adjustment is imposed on the entire p-value function, which is different from that in Boschloo (1970) and Crans and Shuster (2008) where the significance level α of Fisher exact test is raised to $\alpha + \varepsilon$, $\varepsilon \geq 0$ only for the

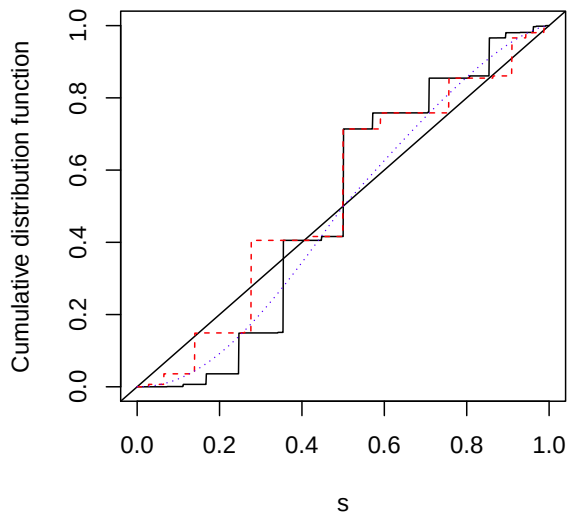

Figure 3.2: The cumulative distribution function $\Pr\{p_{i,\lambda}(\psi_0) \le s\}$ before ($\lambda = 0$; solid curve) and after ($\lambda = 0.4$; dashed curve) the beta adjustment proposed in Section 3.6. The illustration is for the case where $\psi_0 = 1$ and $\pi_{0i} = 0.05$, using one of the 10 studies in Table 8 of Gastwirth (1984) with sample sizes $n_i = 17$ and $m_i = 7$. The corresponding beta adjustment function $G_i(s;\lambda = 0.4)$ (dotted) is illustrated as well.

fixed null value of the parameter.

As illustrated in Figure 3.2, the cumulative distribution function curve of $p_i(\psi_0)$ is roughly S-shaped. We propose to use two beta distribution functions to reduce the deviation between the S-shaped curve and the 45 degree line. Specifically, let $F_{beta}(x;\beta_1,\beta_2)$ be the cumulative distribution function of the beta distribution with the parameters β_1 and β_2. We define beta adjustment functions as follows

$$G_i(x;\lambda) = \begin{cases} F_{beta}\left\{x; 1+\frac{\lambda}{m_i\pi_{0i}(1-\pi_{0i})}, 1+\frac{\lambda}{m_i\pi_{0i}(1-\pi_{0i})}\right\} & \text{if } x \le 1/2, \\ F_{beta}\left\{x; 1+\frac{\lambda}{n_i\pi_{1i}(1-\pi_{1i})}, 1+\frac{\lambda}{n_i\pi_{1i}(1-\pi_{1i})}\right\} & \text{if } x > 1/2, \end{cases} \quad \text{for } i = 1,\ldots,K, \tag{3.14}$$

where $\lambda \ge 0$ is a tuning parameter. Then, the p-value functions (3.7) can be adjusted

as

$$p_{i,\lambda}(\psi) = G_i\{p_i(\psi);\lambda\}, \quad i = 1,\ldots,K. \tag{3.15}$$

Figure 3.2 illustrates a $G_i(x;\lambda)$ function using a dotted curve and the p-value function after adjustment using a dashed curve. We can see that the tail part of the dashed curve is generally closer to the 45 degree line, in comparison with the solid curve. We can also mathematically show that, for any $0 \le \alpha \le 1$,

$$\Pr\{p_{i,\lambda}(\psi_0) \le \alpha/2\} \ge \Pr\{p_i(\psi_0) \le \alpha/2\}, \quad i = 1,\ldots,K,$$

and

$$\Pr\{p_{i,\lambda}(\psi_0) \ge 1-\alpha/2\} \ge \Pr\{p_i(\psi_0) \ge 1-\alpha/2\}, \quad i = 1,\ldots,K.$$

As a result, the power of individual tests is increased, and so is the power of the overall test. Note that the function $G_i(x;\lambda)$ in (3.14) converges to x as both $n_i \to \infty$ and $m_i \to \infty$. In other words, asymptotically, $G_i(x;\lambda)$ becomes an identical transformation. Thus, $p_{i,\lambda}(\psi)$ and $p_i(\psi)$ are equivalent asymptotically, and their difference is noticeable only in small sample settings. The proposed adjustment involves a tuning parameter λ. In our numerical studies in Section 3.7, we use $\lambda = 0.4$ by trial and error. It appears that the choice of $\lambda = 0.4$ generally works well. We may also use a more formal cross validation technique to help select appropriate value of λ.

Denote by $p_{(c),\lambda}(\psi)$ the combined p-value function obtained by combining the adjusted p-value functions $p_{i,\lambda}(\psi)$, $i = 1,\ldots,K$, following the same combining recipe (3.3). We can make inference from $p_{(c),\lambda}(\psi)$ in the same way as from $p_{(c)}(\psi)$. This function $p_{(c),\lambda}(\psi)$ has all the desirable properties established for $p_{(c)}(\psi)$ in Section 3.3 and Section 3.4. Specifically, a formula for the overall type I error rate function can be derived in a similar way to Theorem 3.1. This enables us to estimate the overall type I error rate from the observed data, and monitor the adjustment effect for suitably tuning the parameter λ to avoid over-adjustment. Moreover, $p_{(c),\lambda}(\psi)$ and $p_{(c)}(\psi)$

are asymptotically equivalent as $n_i \to \infty$ and $m_i \to \infty$, $i = 1, \ldots, K$, because of the asymptotic equivalence of $p_{i,\lambda}(\psi)$ and $p_i(\psi)$. The large sample properties established in Theorems 3.2 and 3.3 also hold for the inference using $p_{(c),\lambda}(\psi)$.

3.7 Simulation studies

To examine the performance of the proposed meta-analysis approach, we carry out simulation studies in the setting of rare events. The simulated data structure follows two real data sets: the Avandia data (Nissen and Wolski, 2007, Table 3) and the promotion data (Gastwirth, 1984, Table 8). In our simulation studies we assume that the common parameter of interest is the odds ratio, which is the effect measure used in the original analysis by Nissen and Wolski (2007). For comparison purposes, we include in the simulation studies Mantel–Haenszel and Peto methods, in view of their popularity in applications and the recommendation in Bradburn et al. (2007). We also include the classical conditional inference method (Gart, 1970) but using the likelihood ratio chi-square test to ease the computational burden. To facilitate our comparison, the formulas of these methods are provided in Appendix. In all, seven methods are included in our numerical analysis. They are: (i) the proposed meta-analysis approach combining the p-value functions (3.7) ($\circ$); (ii) the proposed meta-analysis approach combining the beta-adjusted p-value functions (3.15) with the tuning parameter $\lambda = 0.4$ ($\triangle$); (iii) Mantel–Haenszel method without continuity corrections ($+$); (iv) Mantel–Haenszel method with 0.5 imputation for the studies with zero events in either arm ($\times$); (v) Peto method without continuity corrections ($\diamond$); (vi) Peto method with 0.5 imputation for the studies with zero events in either arm ($\bigtriangledown$); (vii) the classical conditional inference method ($\boxtimes$).

In the first simulation study, we independently generate $K = 48$ studies using the same sample sizes as those in the Avandia data. The median of the sample sizes $\{n_1, \ldots, n_K\}$ and $\{m_1, \ldots, m_K\}$ are 222 and 142, respectively. For the i-th study, the

event rate π_{0i} in the control arm is generated from a uniform distribution U(0,ξ). Here, ξ is set to be a small number, such as 0.005, 0.01, 0.05 and 0.1, to ensure low event rate. The event rate in the other arm is determined by $\text{logit}(\pi_{1i}) = \log(\psi) + \text{logit}(\pi_{0i})$ for a fixed odds ratio ψ ranging from 1 to 10. The data (x_i, y_i) are generated using the binomial model described in Section 3.2.1. Because of the low event rates, we can observe a substantial portion of the studies with $x_i = y_i = 0$. This simulation setting is similar to those in Bradburn et al. (2007) and Tian et al. (2009).

Figure 3.3(a) presents the empirical coverage probability of 95% confidence intervals obtained from the aforementioned methods when ξ =0.01. It is clear in Figure 3.3(a) that the coverage probabilities of Mantel–Haenszel method (iv), Peto methods (v) and (vi) decrease quickly as the true odds ratio increases away from one. Only the proposed methods (i) and (ii), Mantel–Haenszel method (iii) and the classical conditional inference method (vii) can yield confidence intervals with adequate coverage probability. These methods also achieve comparable power for testing the hypothesis $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$, as shown in Figure 3.3(b).

In the second simulation study, we repeat the same simulation procedure but use the data structure of the promotion data. This data set consists of 10 studies, and the median of the sample sizes $\{n_1, \ldots, n_K\}$ and $\{m_1, \ldots, m_K\}$ are 25 and 9, respectively. With this data structure, there is a non-negligible chance that zero events are observed in one arm for all the simulated studies, as seen in the real promotion data. In this situation, Mantel–Haenszel method is not applicable unless continuity corrections are applied, and thus Mantel–Haenszel method (iii) can not be applied in our second simulation study.

The analysis results for our second simulation study are shown in Figure 3.3(c)–(d) for $\pi_{0i} \sim$ U(0,0.05). Figure 3.3(c) shows that Mantel–Haenszel method (iv) has a severe coverage problem with very low coverage probability, and the classical conditional inference method (vii), though improved over (iv), generally still has a lower

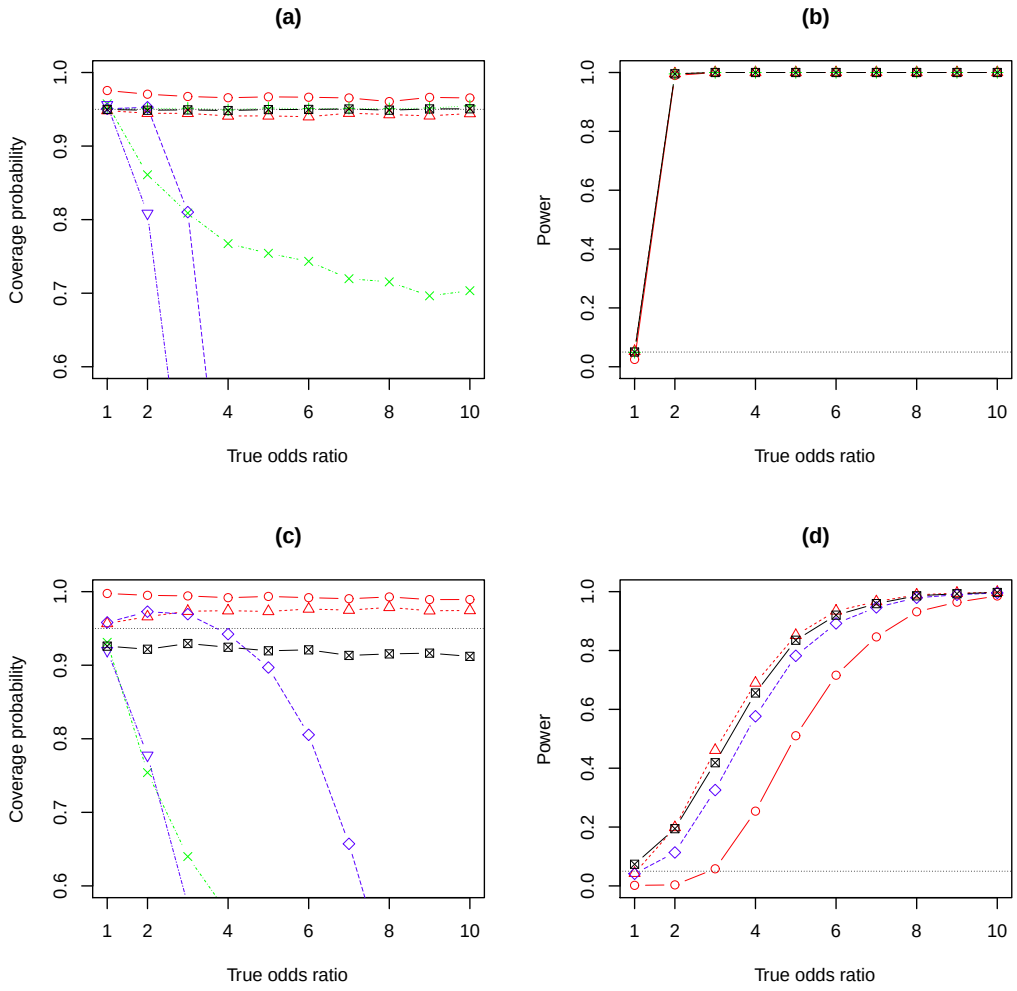

Figure 3.3: Empirical coverage probability of 95% confidence intervals and empirical power of testing $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$, for the odds ratios between 1 and 10. The empirical results are calculated based on 10000 data sets simulated from the structures of the Avandia data (a)-(b) and the promotion data (c)-(d). The baseline event rate π_{0i}, $i = 1, \ldots, K$, are generated from U(0, 0.01) and U(0, 0.05) for illustrations (a)-(b) and (c)-(d), respectively. The methods illustrated are: (i) the proposed method combining p-value functions (3.7) ($\circ$); (ii) the proposed method combining beta adjusted p-value functions with tuning parameter $\lambda = 0.4$ ($\triangle$); (iii) Mantel–Haenszel method without continuity corrections ($+$); (iv) Mantel–Haenszel method with 0.5 continuity corrections for the studies with zero event in either arm ($\times$); (v) Peto method without continuity corrections ($\diamond$); (vi) Peto method with 0.5 continuity corrections for the studies with zero event in either arm ($\bigtriangledown$); and (vii) the classical conditional inference method ($\boxtimes$).

coverage than the nominal value. For Peto method (v), the coverage probability is adequate when the true odds ratio ψ is less than or equal to 4, but falls below 90% when ψ is greater than 5, and decreases further to about 80% when $\psi = 6$. Continuity corrections make the coverage problems of Peto method much worse. These observations are consistent with the findings in Bradburn et al. (2007). Based on their extensive simulation studies, Bradburn et al. (2007) concluded that, when event rates are very low, Peto method (v) has the best confidence interval coverage and is the most powerful method among the commonly used methods, provided the sample sizes of two arms are not severely unbalanced and the true odds ratio is not very large. On the other hand, Figure 3.3(c) shows that the proposed methods (i) and (ii) maintain adequate coverage probability consistently for all the odds ratios in the range of the figure. We also see in Figure 3.3(c) that the proposed method (i) is quite conservative, which is reflected by its exceedingly high empirical coverage probability. As explained in Section 3.6, this conservatism is largely due to the conservatism of individual exact tests. We note that similar conservatism is also seen in the numerical studies for the exact meta-analysis approach proposed for the risk difference in Tian et al. (2009). Figure 3.3(d) shows that the conservatism is greatly reduced in our proposed methods (ii) by combining the adjusted p-value functions. In fact, Figure 3.3(d) shows that the proposed method (ii) is more powerful than Peto method (v). In summary, our proposed methods (ii) completely outperforms Peto method (v) (which was recommended in Bradburn et al. (2007)), and yields both higher coverage probability and higher power simultaneously.

Finally, we present in Table 3.1 the estimated coverage probability of the confidence intervals obtained from the proposed methods (i) and (ii), all in the setting of our first simulation study. As shown in Section 3.3.1, our proposed meta-analysis approach enables us to estimate the actual coverage probability (3.6). To evaluate the accuracy of such estimation, we compare the estimated coverage probability with the actual coverage probability for the 95% confidence intervals. In Table 3.1, the values of the actual coverage probability listed in Part I and Part II correspond to the points in Figure 3.3(a)

denoted by ($\circ$) and ($\triangle$), respectively. More specifically, the actual coverage probability presented here is

$$\int_0^{\xi}\cdots\int_0^{\xi}\{R_{(c)}(0.975)^{-}-R_{(c)}(0.025)\}f_1(\pi_{01})\cdots f_K(\pi_{0K})\,\mathrm{d}\pi_{01}\cdots\mathrm{d}\pi_{0K}, \qquad (3.16)$$

where $f_i(\pi_{0i})=1/\xi$ for $0\le\pi_{0i}\le\xi$ is the density function of the U(0,ξ) distribution. The estimated coverage probability is the counterpart of (3.16), with $R_{(c)}(\cdot)$ replaced by the estimate $\hat{R}_{(c)}(\cdot)$. Part I of Table 3.1 shows that, for the proposed method (i), the absolute difference between the estimated and actual coverage probability is never greater than 0.5% for the listed odds ratios. Part II of Table 3.1 shows that, for the proposed method (ii), the difference is slightly higher, but no greater than 1.5%. This suggests that perhaps the estimation is sufficiently accurate for practical uses.

Table 3.1: Estimated coverage probability of 95% confidence interval

Part I. The proposed method (i)

True odds ratio	1	2	3	4	5	6	7	8	9	10
Actual coverage (%)	97.5	97.0	96.7	96.6	96.7	96.6	96.5	96.0	96.6	96.5
Estimated coverage (%)	97.8	97.1	96.8	96.7	96.6	96.6	96.5	96.5	96.5	96.4
Absolute difference (%)	0.3	0.1	0.1	0.1	0.1	0.0	0.0	0.5	0.1	0.1

Part II. The proposed method (ii)

True odds ratio	1	2	3	4	5	6	7	8	9	10
Actual coverage (%)	94.8	94.5	94.4	94.1	94.1	94.0	94.5	94.3	94.1	94.4
Estimated coverage (%)	96.0	95.7	95.6	95.5	95.4	95.4	95.4	95.3	95.3	95.3
Absolute difference (%)	1.2	1.2	1.2	1.4	1.3	1.4	0.9	1.0	1.2	0.9

3.8 Real data examples

3.8.1 Avandia data

The Avandia data set (Nissen and Wolski, 2007) consists of 48 large randomized clinical trials, and it is used to examine whether the diabetes drug Avandia is associated with myocardial infarction or cardiovascular death. Out of the 48 studies, 10 studies observe no event of myocardial infarction and 25 studies observe no event of cardiovascular death. Nissen and Wolski (2007) excluded these studies from the analysis. Based on Peto method (Peto method (v) in our simulation study), they obtained for the endpoint myocardial infarction a 95% confidence interval of (1.031, 1.979) and a p-value of 0.032 for the odds ratio. Hence, Nissen and Wolski (2007) concluded that Avandia is significantly associated with myocardial infarction. For the endpoint cardiovascular death, they obtained a 95% confidence interval of (0.980, 2.744) and a p-value of 0.060.

Table 3.2 summarizes the analysis results obtained from the seven methods considered in the previous section. For myocardial infarction, from the proposed method (i), we obtain a 95% confidence interval of (0.972, 2.001) and a p-value of 0.071. The estimated coverage probability (3.6) is 97.3%, indicating that the results here may be conservative. From the proposed method (ii), we obtained a 95% confidence interval of (1.037, 2.004) and a p-value of 0.029, and the estimated coverage probability is 96.1%. For cardiovascular death, from the proposed method (i), we obtain a 95% confidence interval of (0.765, 2.965) and a p-value of 0.252. The estimated coverage probability (3.6) is 98.5%. From the proposed method (ii), we obtained a 95% confidence interval of (0.956, 2.981) and a p-value of 0.073, and the estimated coverage probability is 96.5%. In our analysis, we have utilized all available data, including zero total event studies.

Table 3.2 shows that Mantel–Haenszel method (iii) yields a p-value of 0.033 and Peto method (v) yields a p-value of 0.032, both of which are significant at $\alpha = 0.05$

significance level, for myocardial infarction. But both methods exclude all zero total event studies from the analysis. Table 3.2 also shows that, after applying continuity corrections to zero events, Mantel–Haenszel method (iv) yields a p-value of 0.163 and Peto method (vi) yields a p-value of 0.158. Neither of the results is significant even at $\alpha = 0.1$ significance level. Clearly, for Mantel–Haenszel and Peto methods, the exclusion or inclusion of zero total event studies lead to contradictory conclusions. This contradiction has generated confusion in practice (see, e.g., Diamond et al., 2007). The results are also consistent with the finding reported in Sweeting et al. (2004) that imputation to zero events can result in very different conclusions, depending on the numbers imputed.

Table 3.2: Analysis results of Avandia data and promotion data

	Avandia data				Promotion data	
	Myocardial infarction		Cardiovascular death			
	95% CI	P	95% CI	P	95% CI	P
Proposed method (i)	(0.972, 2.001)	0.071	(0.765, 2.965)	0.252	(0.842, ∞)	0.080
Proposed method (ii)	(1.037, 2.004)	0.029	(0.956, 2.981)	0.073	-	-
Mantel–Haenszel	(1.029, 1.978)	0.033	(0.984, 2.930)	0.057	-	-
Mantel–Haenszel-CC	(0.919, 1.647)	0.163	(0.760, 1.689)	0.541	(0.738, 5.396)	0.174
Peto	(1.031, 1.979)	0.032	(0.980, 2.744)	0.060	(1.522, 12.86)	0.006
Peto-CC	(0.921, 1.659)	0.158	(0.761, 1.690)	0.538	(0.776, 4.270)	0.168
Conditional inference	(1.030, 1.979)	0.032	(0.984, 2.880)	0.058	(3.731, ∞)	0.000

Remark: CI = Confidence interval; P = p-value for hypothesis testing $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$; CC = Continuity corrections to zero events using 0.5 imputation.

For the proposed approach, we have shown in Section 3.3.2 that zero total event studies have the effect of increasing uncertainty in the overall inference outcomes. To illustrate this effect numerically, we intentionally remove all zero total event studies and analyze the Avandia data again. For myocardial infarction, we obtain 95% confidence intervals of (0.978, 1.994) and (1.040, 1.996) from the proposed method (i) and (ii), respectively. Both intervals are slightly narrower than those obtained by analyzing the entire data.

3.8.2 Promotion data

The promotion data set (Gastwirth, 1984, Table 8) consists of 10 small observational surveys, and it is used to examine the difference between white and black employees in promotion rates. A special feature of this data set is that no event of promotions is observed in the arm of black employees among all the studies. In this situation, Mantel–Haenszel point estimate is undefined for the odds ratio. With 0.5 continuity corrections, Mantel–Haenszel method (iv) yields a 95% confidence interval of (0.738, 5.396) and a p-value of 0.174. This is a statistically insignificant result even at α =0.1 significance level. Using the proposed method (i), we obtained a 95% confidence interval of (0.842, ∞) and a p-value of 0.080. Analysis results of other methods are reported in Table 3.2. Note that the proposed method (i) and the classical conditional inference method yield infinity as the upper end of the confidence interval for the odds ratio. In fact, the promotion data set, with no event in the arm of black employees across all the studies, does not provide any evidence for rejecting the hypothesis $H_0 : \psi = \psi^*$ and favoring $H_1 : \psi < \psi^*$, for any value of ψ^*. Thus, it is proper to use infinity as the upper end of the confidence interval. In contrast, Table 3.2 shows that both Mantel–Haenszel and Peto methods yield finite confidence intervals. This is because both methods obtain Wald-type intervals by computing point estimates plus/minus a constant times estimated standard errors. But finite confidence intervals are inappropriate for the case of the promotion data. Apparently, both Mantel–Haenszel and Peto methods have artificially

placed an upper bounds on the odds ratio. Our proposed approach and the classical conditional inference method do not have this problem.

3.9 Discussion

In this chapter, we have proposed an exact meta-analysis approach by combining the p-value functions obtained from the exact tests associated with individual studies. This approach is fundamentally different from conventional meta-analysis approaches. In the setting of rare events, this approach can incorporate zero total event studies in the analysis without applying artificial continuity corrections to zero events. We have shown both theoretically and numerically that the proposed combined p-value function can be used for making exact inference. In particular, the overall type I error rate can be expressed explicitly in terms of the type I error rates associated with individual studies. This enables us to estimate the overall type I error rate from the data, and also to devise further adjustment of individual p-value functions to improve the power of the overall test. We also use an asymptotic result to provide useful guidelines for choosing appropriate weights to improve the combination efficiency. Through numerical studies using simulated and real data, we show that our proposed meta-analysis approach is superior to Mantel-Haenszel, Peto and the classical conditional inference methods in the setting of rare events.

The proposed approach may be seen as a generalization of the classical approach of combining p-values, but it combines p-value functions rather than point estimators or the observed p-values. The idea of combining p-value functions can lead to a broad class of exact meta-analysis approaches, in the sense that the individual p-value functions in the combining recipe (3.3) can be obtained based on any exact test, including those recommended by Lydersen et al. (2009), and those proposed by Agresti and Min (2002) and Reiczigel et al. (2008). In fact, the p-value functions are allowed to be derived from different exact test procedures among the studies. Moreover, the $\Phi(\cdot)$

function in the combining recipe (3.3) can be replaced by any continuous monotonic function, as seen in (Singh et al., 2005; Xie et al., 2011). This further broadens the class of the proposed exact meta-analysis approaches. For the mid-p adaptation of Fish exact test illustrated in this chapter, we have shown that the use of $\Phi(\cdot)$ in (3.3) can yield asymptotically efficient inference and good small sample performance. Note that this asymptotic efficiency result (in terms of preserving Fisher information) does not hold for other common choices of the continuous monotonic functions listed in Singh et al. (2005) and Xie et al. (2011), including those used in the classical Fisher approach of combining p-values.

We also observe that our proposed meta-analysis approach provides a computationally feasible way for extending an exact test for a single study to multiple studies. The overall computational expense of our approach amounts to the summation of the individual computational expense associated with each of the studies. In other words, the overall computational complexity is at the order of $O(K)$. Hence, the development in this chapter provides a feasible solution to the computational question raised in Agresti and Min (2002). Specifically, Agresti and Min (2002) pointed out a challenging problem in extending their proposed exact test for a single 2×2 table to multiple 2×2 tables. That is, the computational complexity increases exponentially as the number of the studies increases, namely at the order of $O(g^K)$. Here g is the number of grid values in a greed search. Consequently, it would be problematic to implement their proposed test even when K is moderate. In fact, this computational problem prevails in almost all exact test methods when they are extended to multiple 2×2 tables. To this end, Mehta, Patel, and Gray (1985) proposed a computing algorithm that applies to a conditional exact test for the common odds ratio. But their approach does not apply to either unconditional exact tests or other effect measures, such as the risk ratio, risk difference and others. Our proposed approach, on the other hand, clearly provides a simple but unified framework for extending an exact test for a single study to multiple studies. In fact, our approach can be implemented simply by compiling the existing

computer programs for computing the p-values for a single study. For running our approach with the Avandia data with $K = 48$, including inverting the p-value functions, it takes less than 30 seconds on a Dell PC using the R code which is available on the first author's website.

We stress that, although the presentation in this chapter focuses on the problems in meta-analysis of 2×2 tables with zero events, our approach applies to any general meta-analysis setting with discrete or/and continuous data.

3.10 Appendix

3.10.1 Review of traditional methods for meta-analysis of 2×2 tables

The Mantel-Haenszel method uses the following formula to estimate the common odds ratio

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^{K} x_i(m_i - y_i)/(n_i + m_i)}{\sum_{i=1}^{K} y_i(n_i - x_i)/(n_i + m_i)}.$$

The variance estimate of $\log(\hat{\psi}_{MH})$, proposed by Robins, Breslow, and Greenland (1986), is

$$\mathrm{Var}\{\log(\hat{\psi}_{MH})\} = \frac{\sum_{i=1}^{K} P_i R_i}{2(\sum_{i=1}^{K} R_i)^2} + \frac{\sum_{i=1}^{K} (P_i S_i + Q_i R_i)}{2(\sum_{i=1}^{K} R_i)(\sum_{i=1}^{K} S_i)} + \frac{\sum_{i=1}^{K} Q_i S_i}{2(\sum_{i=1}^{K} S_i)^2},$$

where $P_i = (x_i + m_i - y_i)/(n_i + m_i), Q_i = (n_i - x_i + y_i)/(n_i + m_i), R_i = x_i(m_i - y_i)/(n_i + m_i)$, and $S_i = y_i(n_i - x_i)/(n_i + m_i)$ $(i = 1, \ldots, K)$.

The Peto method uses the following to make inference about the common odds ratio

$$\log(\hat{\psi}_{Peto}) = \frac{\sum_{i=1}^{K} (O_i - E_i)}{\sum_{i=1}^{K} V_i} \quad \text{and} \quad \mathrm{Var}\{\log(\hat{\psi}_{Peto})\} = \frac{1}{\sum_{i=1}^{K} V_i},$$

where $O_i = x_i, E_i = n_i(x_i + y_i)/(n_i + m_i)$ and $V_i = n_i m_i (x_i + y_i)(n_i + m_i - x_i - y_i)/\{(n_i + m_i)^2 (n_i + m_i - 1)\}$ $(i = 1, \ldots, K)$.

The classical conditional inference method is based on the multiplied conditional likelihood function $L(\psi) = \prod_{i=1}^{K} \Pr_{\psi}(X_i = x_i \mid T_i = t_i)$, which can yield an η-level confidence interval

$$\left\{ \psi : 2\log \frac{\sup_{\psi} L(\psi)}{L(\psi)} \leq d_{\eta} \right\}.$$

Here d_{η} is the η-th percentile of the χ^2 distribution with one degree of freedom (Andersen, 1971).

3.10.2 Proofs

Proof of Theorem 3.1. Define random variables B_{ij} $(i = 1, \ldots, K; j = 1, \ldots, K)$ as seen in Theorem 3.1. By sequentially conditioning on $p_i(\psi_0)$ $(i = 1, \ldots, K)$, we can establish that

$$\begin{aligned}
& \Pr\{p_{(c)}(\psi_0) \leq s\} \\
= \; & E\left\{ \Pr\left(\Phi\left[\left(\sum_{i=1}^{K} w_i^2 \right)^{-1/2} \sum_{i=1}^{K} w_i \Phi^{-1}\{p_i(\psi_0)\} \right] \leq s \;\middle|\; p_2(\psi_0), \ldots, p_k(\psi_0) \right) \right\} \\
= \; & E\left(\Pr\left[p_1(\psi_0) \leq \Phi\left\{ \left(1 + \sum_{j \neq 1} \frac{w_j^2}{w_1^2} \right)^{1/2} \Phi^{-1}(s) - \sum_{j \neq 1} \frac{w_j}{w_1} \Phi^{-1}(B_{1j}) \right\} \;\middle|\; p_2(\psi_0), \ldots, p_k(\psi_0) \right] \right) \\
= \; & E\left(\Pr\left[U_1 \leq \Phi\left\{ \left(1 + \sum_{j \neq 1} \frac{w_j^2}{w_1^2} \right)^{1/2} \Phi^{-1}(s) - \sum_{j \neq 1} \frac{w_j}{w_1} \Phi^{-1}(B_{1j}) \right\} \;\middle|\; p_2(\psi_0), \ldots, p_k(\psi_0) \right] \right) \\
& \quad + E\left(D_1 \left[\Phi\left\{ \left(1 + \sum_{j \neq 1} \frac{w_j^2}{w_1^2} \right)^{1/2} \Phi^{-1}(s) - \sum_{j \neq 1} \frac{w_j}{w_1} \Phi^{-1}(B_{1j}) \right\} \right] \right) \\
= \; & \Pr\left[\Phi\left\{ \left(\sum_{i=1}^{K} w_i^2 \right)^{-1/2} \sum_{i=1}^{K} w_i \Phi^{-1}(B_{1i}) \right\} \leq s \right] + d_1(s) \\
= \; & \Pr\left[\Phi\left\{ \left(\sum_{i=1}^{K} w_i^2 \right)^{-1/2} \sum_{i=1}^{K} w_i \Phi^{-1}(B_{Ki}) \right\} \leq s \right] + \sum_{i=1}^{K} d_i(s) \\
= \; & s + \sum_{i=1}^{K} d_i(s)
\end{aligned}$$

This completes the proof. □

Proof of Theorem 3.2. This is a direct result of Theorem 3.1 and Corollary 3.1. □

Proof of Theorem 3.3. It is easy to show that

$$p_{(c)}(\psi) = \Phi\left[\frac{1}{\left\{\sum_{i=1}^{K} w_i^2\right\}^{1/2}} \sum_{i=1}^{K} w_i \Phi^{-1}\{p_i(\psi)\}\right] = \Phi\left[\left\{\sum_{i=1}^{K} \frac{1}{\widehat{\text{aVar}}(\hat{\psi}_{i,MLE})}\right\}^{1/2} (\psi - \hat{\psi}_c)\right] + o(1),$$

where

$$\hat{\psi}_c = \left\{\sum_{i=1}^{K} \frac{\hat{\psi}_{i,MLE}}{\widehat{\text{aVar}}(\hat{\psi}_{i,MLE})}\right\} \Bigg/ \left\{\sum_{i=1}^{K} \frac{1}{\widehat{\text{aVar}}(\hat{\psi}_{i,MLE})}\right\}.$$

The result of Theorem 3.3 then follows. □

Chapter 4
Concluding Remarks

In this dissertation, we have developed a new statistical methodology for combining information from independent studies. Our development is based on the combination of confidence distributions. This is fundamentally different from the combination of point estimates in the conventional meta-analysis approaches. We have shown that the proposed methodology has desirable inferential properties in two recurrent settings. More specifically, 1) in the presence of heterogeneous studies, we have shown that the proposed approach in Chapter 2 can integrate indirect evidence and achieves asymptotic efficiency; 2) for rare event studies, we have shown that the proposed approach in Chapter 3 can yield exact inference and incorporate all available data in the analysis without using artificial continuity corrections for zero events. The established properties have been demonstrated in our numerical studies.

The general idea of combining confidence distributions as a way to synthesize information was first proposed in Singh et al. (2005). This idea was further developed in Xie et al. (2011). They showed that the combining of confidence distributions provides a unifying framework for meta-analysis and leads to the development of robust meta-analysis methods. The developments in this dissertation further demonstrate that confidence distribution is a useful inferential tool in meta-analysis. Here, we stress that our developments utilize a key feature of confidence distribution, that is, it uses a distribution function on the parameter space as an "information carrier". Since a distribution function can carry wealthy information, much more than a point or interval estimator, the combining of confidence distribution can potentially preserve more information in the process of evidence synthesis. More concretely, in Chapter 2, we have shown that

the multiparameter confidence distribution can carry correlation information, and this enables us to achieve asymptotically efficient inference for those parameters that are not even estimable in some of the studies. In Chapter 3, we have shown that p-value functions (which can be viewed as confidence distributions in asymptotic sense) can inherit the "exact" properties of the tests that are based on the exact distributions of the underlying test statistics, and this enables us to make an overall exact inference for the common parameter of the studies. On the other hand, point estimates generally can not preserve "distributional information", such as correlation or exactness. Consequently, the developments in Chapter 2 and Chapter 3 can not be easily achieved in the conventional meta-analysis framework of combining point estimates.

To close this dissertation, we remark that the general idea of combining confidence distributions can be potentially used to solve other open problems in meta-analysis.

Appendix A
Notions of Confidence Distribution

A confidence distribution (CD) is a sample-dependent distribution function on the parameter space satisfying certain requirements in terms of repeated sampling performance. It is a pure frequentist concept and can be used to estimate an unknown parameter. In the following, we list the formal definitions of CD for 1) single parameter case (Schweder and Hjort, 2002; Singh et al., 2005); and 2) multiparameter case (Singh et al., 2007). A comprehensive review for the modern interpretation and application of this concept is provided in Xie and Singh (2012).

Definition 1 (Single-parameter CD)**.** A function $H_n(\cdot) = H_n(\boldsymbol{X}_n, \cdot)$ on $\mathscr{X} \times \Theta \to [0,1]$ is called a CD for a parameter θ if (i) For each given sample set $\boldsymbol{X}_n$ in the sample space $\mathscr{X}$, $H_n(\cdot)$ is a continuous cumulative distribution function in the parameter space Θ; (ii) At the true parameter value $\theta = \theta_0$, $H_n(\theta_0) = H_n(\boldsymbol{X}_n, \theta_0)$, as a function of the sample set $\boldsymbol{X}_n$, has a uniform distribution $U(0,1)$.

The function $H_n(\cdot)$ is called an asymptotic confidence distribution (aCD) if (ii) is replaced by (ii)': at $\theta = \theta_0$, $H_n(\theta_0) \to U(0,1)$ as $n \to \infty$ and the continuity requirement on $H_n(\cdot)$ is dropped.

In other words, a CD is a function, defined on both the sample space and the parameter space, that satisfies two requirements. The requirement (i) is simply that, for a given sample, it is a distribution function on the parameter space. The requirement (ii) imposes a restriction to this sample-dependent distribution function so that inference drawn from it has desirable frequentist properties.

For the multiparameter case, the definition of CD is not unique due to the multi-dimensionality. Singh et al. (2007) provided two definitions, namely l-CD and c-CD, from different perspectives. The l-CD is defined as a sample-dependent multivariate distribution satisfying certain requirement in terms of its marginal distributions. The c-CD is defined as a sample-dependent multivariate distribution satisfying certain requirement in terms of its "central region".

Definition 2 (Multiparameter l-CD)**.** A function $H_n(\cdot) = H_n(\boldsymbol{X}_n, \cdot)$ on the parameter space $\Theta \subseteq \mathbb{R}^p$ is a (asymptotic) CD for the p-dimensional vector parameter $\boldsymbol{\theta}$ in the *linear* sense if (i) it is a probability distribution function over the parameter space Θ for any fixed sample $\boldsymbol{X}_n$; and (ii) for any $p \times 1$ vector $\boldsymbol{\lambda}$, the conditional distribution of $\boldsymbol{\lambda}'\boldsymbol{\xi}_n$ given $\boldsymbol{X}_n$ is a (asymptotic) CD for $\boldsymbol{\lambda}'\boldsymbol{\theta}$, where the $p \times 1$ random vector $\boldsymbol{\xi}_n$ has the distribution $H_n(\cdot)$ given $\boldsymbol{X}_n$.

In other words, an l-CD for a multi-dimensional parameter is a sample-dependent multivariate distribution on the parameter space such that its marginal distribution is a CD for the corresponding (one-dimensional) component of the parameter.

Definition 3 (Multiparameter c-CD)**.** A function $H_n(\cdot) = H_n(\boldsymbol{X}_n, \cdot)$ on the parameter space $\Theta \subseteq \mathbb{R}^p$ is a (asymptotic) CD for the p-dimensional vector parameter $\boldsymbol{\theta}$ in the *circular* sense if (i) it is a probability distribution function over the parameter space Θ for any fixed sample $\boldsymbol{X}_n$; and (ii) for any $0 < \alpha < 1$, the $100(1-\alpha)\%$ central region of $H_n(\cdot)$ is a confidence region for $\boldsymbol{\theta}$ with (asymptotic) coverage probability $1-\alpha$.

In other words, an c-CD for a multi-dimensional parameter is a sample-dependent multivariate distribution on the parameter space such that its "central region" is a confidence region for the parameter. Such "central region" can be defined using the notion of data depth (see, e.g., Liu et al., 1999).

Bibliography

Ades, A. E. and Sutton, A. J. (2006), "Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches," *Journal of the Royal Statistical Society: Series A*, 169, 5–35.

Agresti, A. (1992), "A survey of exact inference for contingency tables," *Statist. Sci.*, 7, 131–53.

Agresti, A. and Min, Y. (2002), "Unconditional small-sample confidence intervals for the odds ratio," *Biostatistics*, 3, 379–386.

Andersen, E. B. (1971), "The asymptotic distribution of conditional likelihood ratio tests," *J. Am. Statist. Assoc.*, 630–633.

Becker, B. J. and Wu, M. J. (2007), "The synthesis of regression slopes in meta-analysis," *Statistical Science*, 22, 414–429.

Boschloo, R. D. (1970), "Raised conditional level of significance for the $2\times$ 2-table when testing the equality of two probabilities," *Statistica Neerlandica*, 24, 1–9.

Bradburn, M. J., Deeks, J. J., Berlin, J. A., and Localio, A. R. (2007), "Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events," *Statist. Med.*, 26, 53–77.

Breslow, N. (1981), "Odds ratio estimators when the data are sparse," *Biometrika*, 68, 73–84.

Cox, D. R. and Reid, N. (2004), "A note on pseudolikelihood constructed from marginal densities," *Biometrika*, 91, 729–737.

Crans, G. G. and Shuster, J. J. (2008), "How conservative is Fisher's exact test? A quantitative evaluation of the two-sample comparative binomial trial," *Statist. Med.*, 27, 3598–611.

Diamond, G. A., Bax, L., and Kaul, S. (2007), "Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death," *Ann. Intern. Med.*, 147, 578–81.

Dominici, F. and Parmigiani, G. (2000), "Combining studies with continuous and dichotomous responses: a latent-variables approach," in *Meta-analysis in Medicine and Health Policy*, eds. Stangl, D. K. and Berry, D. A., New York: Marcel Dekker, pp. 105–125.

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

— (1996), "Empirical Bayes methods for combining likelihoods," *J. Am. Statist. Assoc.*, 91, 538–50.

— (1998), "R. A. Fisher in the 21st Century," *Statistical Science*, 13, 95–114.

Fahrmeir, L. and Tutz, G. (2001), *Multivariate statistical modelling based on generalized linear models*, Springer Verlag.

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

Fraser, D. A. S. (1991), "Statistical inference: Likelihood to significance," *J. Am. Statist. Assoc.*, 86, 258–65.

Gart, J. J. (1970), "Point and interval estimation of the common odds ratio in the combination of $2\times$ 2 tables with fixed marginals," *Biometrika*, 57, 471–475.

Gastwirth, J. L. (1984), "Statistical methods for analyzing claims of employment discrimination," *Industr. Labor Relat. Rev.*, 38, 75–86.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Boca Raton: Chapman & Hall/CRC, 2nd ed.

Hwang, J. T. G. and Yang, M. C. (2001), "An optimality theory for mid p-values in $2\times$ 2 contingency tables," *Statist. Sinica*, 11, 807–26.

Kendall, M. and Stuart, A. (1973), *The advanced theory of statistics*, vol. 2, London: Griffin, 3rd ed.

Kou, S. G. and Ying, Z. (1996), "Asymptotics for a $2\times$ 2 table with fixed margins," *Statist. Sinica*, 6, 809–29.

Lancaster, H. O. (1961), "Significance tests in discrete distributions," *J. Am. Statist. Assoc.*, 56, 223–34.

Lehmann, E. and Casella, G. (1998), *Theory of point estimation*, Springer Verlag.

Liang, K. Y. and Zeger, S. L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22.

Lin, D. Y. and Zeng, D. (2010), "On the relative efficiency of using summary statistics versus individual-level data in meta-analysis," *Biometrika*, 97, 321–332.

Liu, R., Parelius, J., and Singh, K. (1999), "Multivariate analysis by data depth: descriptive statistics, graphics and inference," *Annals of Statistics*, 27, 783–840.

Lydersen, S., Fagerland, M. W., and Laake, P. (2009), "Recommended tests for association in $2\times$ 2 tables," *Statist. Med.*, 28, 1159–75.

Mathew, T. and Nordstrom, K. (1999), "On the equivalence of meta-analysis using literature and using individual patient data," *Biometrics*, 55, 1221–1223.

Mehta, C. R., Patel, N. R., and Gray, R. (1985), "Computing an exact confidence interval for the common odds ratio in several 2 $\times$ 2 contingency tables," *J. Am. Statist. Assoc.*, 80, 969–973.

Murphy, S. A. and van der Vaart, A. W. (2000), "On profile likelihood," *Journal of the American Statistical Association*, 95, 449–465.

Nissen, S. E. and Wolski, K. (2007), "Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes," *N. Engl. J. Med.*, 356, 2457–71.

Normand, S. L. T. (1999), "Meta-analysis: formulating, evaluating, combining, and reporting," *Statist. Med.*, 18, 321–59.

Olkin, I. and Sampson, A. (1998), "Comparison of meta-analysis versus analysis of variance of individual patient data," *Biometrics*, 317–322.

Reiczigel, J., Abonyi-Tóth, Z., and Singer, J. (2008), "An exact confidence set for two binomial proportions and exact unconditional confidence intervals for the difference and ratio of proportions," *Computational Statistics & Data Analysis*, 52, 5046–5053.

Robins, J., Breslow, N., and Greenland, S. (1986), "Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models," *Biometrics*, 42, 311–23.

Rücker, G., Schwarzer, G., Carpenter, J., and Olkin, I. (2009), "Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells," *Statist. Med.*, 28, 721–38.

Schweder, T. and Hjort, N. L. (2002), "Confidence and likelihood," *Scand. J. Statist.*, 29, 309–32.

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, 2982–2999.

Singh, K., Xie, M., and Strawderman, W. E. (2005), "Combining information from independent sources through confidence distributions," *Ann. Statist.*, 159–83.

— (2007), "Confidence distribution (CD): distribution estimator of a parameter," in *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond. IMS Lecture Notes-Monograph Series*, eds. Liu, R., Strawderman, W., and Zhang, C.-H., Beachwood, Ohio: Institute of Mathematical Statistics, vol. 54, pp. 132–50.

Stouffer, S. A., DeVinney, L. C., and Suchmen, E. A. (1949), *Adjustment During Army Life*, vol. 1, Princeton: Princeton University Press.

Suissa, S. and Shuster, J. J. (1985), "Exact unconditional sample sizes for the $2\times$ 2 binomial trial," *J. R. Statist. Soc. A*, 317–327.

Sutton, A. J. and Higgins, J. P. T. (2008), "Recent developments in meta-analysis," *Statist. Med.*, 27, 625–50.

Sweeting, M. J., Sutton, A. J., and Lambert, P. C. (2004), "What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data," *Statist. Med.*, 23, 1351–75.

Tian, L., Cai, T., Pfeffer, M. A., Piankov, N., Cremieux, P. Y., and Wei, L. J. (2009), "Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent $2\times$ 2 tables with all available data but without artificial continuity correction," *Biostatistics*, 10, 275–81.

Tian, L., Wang, R., Cai, T., and Wei, L. J. (2010), "The highest confidence density region and its usage for joint inferences about constrained parameters," *Biometrics*, 67, 604–610.

Van Es, G. (2005), "Running out of runway. Analysis of 35 years of landing overrun accidents." Tech. Rep. NLR-TP-2005-498, National Aerospace Laboratory.

Whitehead, A., Bailey, A. J., and Elbourne, D. (1999), "Combining summaries of binary outcomes with those of continuous outcomes in a meta-analysis." *Journal of Biopharmaceutical Statistics*, 9, 1–16.

Xie, M. and Singh, K. (2012), "Confidence distribution, the frequentist distribution estimator of a parameter — a Review," *Int. Statist. Rev.*, Accepted.

Xie, M., Singh, K., and Strawderman, W. E. (2011), "Confidence distributions and a unifying framework for meta-analysis," *J. Am. Statist. Assoc.*, 106, 320–33.

Xie, M. and Yang, Y. (2003), "Asymptotics for generalized estimating equations with large cluster sizes," *Annals of Statistics*, 31, 310–347.

Vita

Dungang Liu

2006	B.S. in Statistics, University of Science and Technology of China, Hefei, P.R. China
2006-2007	Graduate Fellow, Department of Statistics and Biostatistics, Rutgers University
2007-2012	Teaching Assistant/Research Assistant, Department of Statistics and Biostatistics, Rutgers University