RECENT DEVELOPMENTS IN COMPLEX META-ANALYSIS UTILIZING THE CONFIDENCE DISTRIBUTION APPROACH

BY YANG JIAO

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

Min-ge Xie, Eun-Young Mun

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2017

ABSTRACT OF THE DISSERTATION

RECENT DEVELOPMENTS IN COMPLEX META-ANALYSIS UTILIZING THE CONFIDENCE DISTRIBUTION APPROACH

by Yang Jiao

Dissertation Director: Min-ge Xie, Eun-Young Mun

In recent years, a new information combining method that combines confidence distributions has been demonstrated as a powerful statistical inferential tool. One can draw most types of frequentist inference based on confidence distributions (Singh et al., 2007). The fact that a confidence distribution contains a wealth of information can be leveraged for synthesizing information from multiple studies (Singh et al., 2005). Xie et al. (2012) showed that by applying appropriate recipes when combining data, the confidence distribution approach (hereinafter referred to as the CD approach) can subsume most of the classical meta-analysis methods within a unified meta-analysis framework. For a comprehensive review of the CD approach and recent developments, see Xie and Singh (2013).

This dissertation extends the existing meta-analysis methods via combining confidence distributions to overcome two challenges. First, most of the existing data situations for which the CD approach has been examined have been to combine continuous data. Therefore, we demonstrate a new CD method for discrete data and apply it to combine 2 x 2 tables from disparate sources. Second, as a major extension of the recent efforts on

drawing joint inference for multiple related parameters from different studies through combining multivariate confidence distributions (Liu et al., 2015; Yang et al., 2014), we propose a CD based three-stage synthesis method to combine 13 parameters from individual participant level data of 14 clinical studies.

The first part of this dissertation focuses on how to apply the CD approach to make exact inferences on 2×2 tables that may involve rare events. While most conventional methods rely on large sample approximations, many 2×2 tables derived in medical fields may have very limited total sample sizes, in which case the use of asymptotic based approaches may lead to misleading conclusions. In addition, we also consider the situations where study total sample size is large, but with zero observed events in one or both treatment arms in a 2×2 table. This can happen in drug safety studies where zero or rare cases of adverse effects are observed in large samples of patients. The new CD method provides an exact inference and does not rely on large sample approximations. In addition, by incorporating prior information, the proposed CD approach can deal with zero events more systematically in contrast to a typical approach adding a small constant (e.g., 0.5) to empty cells. This new approach accounts for various data sampling schemes and can readily be generalized to most of the risk metrics used for 2×2 tables.

The second part of this dissertation focuses on how to synthesize multiple parameters from various studies with heterogeneous designs and partial information. Such a data situation is quite typical for synthesis of clinical studies. For instance, in our motivating data example, individual participant data from Project INTEGRATE were obtained from 24 clinical trials aimed at examining the efficacy of brief alcohol interventions to reduce excessive alcohol use and to prevent harm among college students. Despite having similar objectives among these trials, they differed in terms of the interventions evaluated, covariates assessed, follow-up schedules, among others. With the existing methods, one may have to limit the analysis to a subset of trials with all covariates or to a subset of covariates for a reduced model, either of which excludes partially available data, resulting in an important loss of information. The new CD-based method

can efficiently combine all studies with all the covariates, thus minimizing the information loss that would have occurred under the existing synthesis methods. Moreover, compared to the existing multivariate CD approach proposed by Yang et al. (2014), the current work extends it to random-effects meta-analysis models and to a complex model requiring synthesis of a large number of parameters.

Acknowledgements

Dedication

To my family

List of Tables

List of Figures

Table of Contents

Chapter 1

Introduction

Meta-analysis can be broadly defined as any quantitative methods aimed at synthesizing information from multiple, independent sources. The first known use of a meta-analytic method was by Karl Pearson in 1904. Pearson examined the association between inoculation and mortality from typhoid among soldiers who had volunteered for inoculation against typhoid for their deployment in various places across the British Empire (O'Rourke, 2007). Despite its deep root, meta-analysis as a method did not take off until 1976 when the term "meta-analysis" was coined by Gene Glass as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the finding.". Meta-analysis has since caught on in many substantive fields, especially in medicine. Recent methodological advances in the field of meta-analysis include methods utilizing individual participant data (IPD) and complex meta-analysis methods, such as multivariate meta-analysis and meta-regression. In particular, Xie et al. (2012) proposed a broad unifying approach to combining data from multiple studies. This unifying approach combines confidence distributions from multiple sources in which a confidence distribution (CD) is a sample-dependent distribution function that can represent confidence intervals of all levels for a parameter of interest. Accordingly, a CD contains much more information about data than a point or an interval estimator. This CD-based approach can encompass most of the classical methods, including those from model-based approaches (i.e., fixed-effects and random-effects meta-analysis models using Maximum Likelihood estimator or Bayesian estimator) under one inclusive theoretical umbrella.

Two recent developments for the CD approach to meta-analysis are as follows:

– Liu et al. (2014): demonstrated exact inference for a meta-analysis of 2×2 tables from clinical trials through combining p-value functions;

– Yang et al. (2014): jointly analyzed multiple parameters from individual studies through combining multivariate CD functions;

This dissertation is aimed at extending the two approaches above as follows:

- For Liu et al. (2014): p-value functions (Fraser, 1991) obtained from the exact tests associated with individual studies were used as the combining vehicle. However, Fisher's exact test is based on the assumption that both margins of a 2×2 table are fixed. This assumption is rarely met in actual data because most of the 2×2 tables are observed under different data situations. In this dissertation, we propose the confidence distribution approach for 2×2 tables that accounts for the actual data generation process. One particular challenge is the discrete nature of data, resulting in the confidence distributions that are not unique. Moreover, to deal with the case of zero observed events, we incorporate prior information when deriving confidence distributions instead of utilizing the ad-hoc 0.5 artificial corrections. Finally, we adopt the same combining recipe used in Liu et al. (2014) to combine confidence distributions from individual studies.

- For Yang et al. (2014): As pointed out in Jackson et al. (2011), under the multivariate meta-analysis setting, the assumption that there is no between-study variation in any endpoints in a vector of parameters is implausible. So we extend the existing multivariate CD approach for fixed-effects meta-analysis(Liu et al., 2015; Yang et al., 2014) to random-effects meta-analysis model. Furthermore, in our real data application, the dimension of the parameter vector to be synthesized is much higher than the simulation studies conducted previously. Aside from the methodology developments, both of these modifications requires much more efficient implementation from a computation perspective. In addition, instead of combining effect sizes for multiple interventions as endpoints, we combine regression model parameters directly to facilitate model based inference. This will be

3

illustrated in greater details at a later point.

This dissertation organizes each chapter as a self-contained paper. Specifically, Chapter 2 can be considered as an extension from Liu et al. (2014). We propose several ways to construct confidence distributions of the log odds-ratio for a 2×2 table under different sampling schemes; compare them in simulation studies; and subsequently apply these methods to synthesize information from multiple sources in the context of meta-analysis. Real data analysis and simulation studies showed that the proposed CD approach often outperforms classical methods such as Wald method. Sensitivity analysis also showed that the proposed approach performs consistently well even when prior information was mis-specified. Because the proposed method can easily be generalized to other commonly used measures of association for two-way contingency tables and also because it takes into account different sampling schemes, the proposed CD approach may provide a general framework for 2×2 table inference. Chapter 3 builds upon Yang et al. (2014) and provides a CD-based, three-stage synthesis method to combine multiple related parameters in a multivariate random-effects meta-analysis of IPD from clinical trials. This approach may provide a methodological solution to handle systematic study-level missing data when synthesizing relatively high dimensional data from studies with heterogeneous designs. We also report data from sensitivity analysis to investigate the robustness of the proposed method. Chapter 4 contains some concluding remarks.

Chapter 2

A Confidence Distribution Approach for Exact Inference on 2×2 Tables with Applications to Rare Event Data

In this chapter we discuss the challenges of drawing inference from 2×2 tables and propose a confidence distribution (CD) approach to overcome the challenges. The proposed method may be useful especially for studies on rare events because it applies readily to 2×2 tables with zero observed counts without any artificial corrections. It can explicitly account for the sampling scheme utilized and can easily be generated to any metrics, thus providing an unified framework for analyzing 2×2 tables. As a demonstration, we apply the proposed CD approach to combine information from multiple 2×2 tables in a meta-analysis context to investigate drug safety.

2.1   Introduction

Two way contingency tables have been widely used in various scientific fields to study relationships between two binary variables. Most of the existing inferences for 2×2 tables rely heavily on large sample approximations, which may not necessarily produce valid inference when sample size is small. In such cases, Fisher's exact test (Fisher, 1956) is usually advised. However, its applicability has long been questioned because most 2×2 tables are not collected under the sampling constrains that both marginal totals are fixed. Also, Fisher's test is known to be overly conservative due to high discreteness of the hypergeometric distribution, especially with small samples. In this article, we propose to use a confidence distribution (CD) approach to analyze 2×2 tables. Specifically, we consider confidence intervals of log odds-ratio as an inferential

tool. Our approach does not rely on limiting distributions. Thus, it can be considered as exact inference (Agresti, 2003). In addition, the sampling method under which a 2×2 table was generated is reflected in the method, which may also be applicable to any other metrics for the analysis of 2×2 tables. From these perspectives, the CD approach to 2×2 table inference provides a unified analytical framework.

Loosely speaking, a CD function is often referred to as a sample-dependent distribution function that can represent the confidence intervals of all levels for a parameter of interest (see, e.g, Cox (1958); Efron (1993); Xie and Singh (2013)). It is a distribution estimator developed under the pure frequentist framework, as a counterpart to Bayesian posterior. One special case that is well known and extremely popular in modern statistics is Efron's bootstrap distribution (Efron, 1998), albeit the concept of CD is much broader. Given sample data $\mathbf{x} \sim F_\theta(X)$, denote the CD function for parameter $\theta \in \Theta$ as $H(\cdot) = H(\cdot, \mathbf{x})$, we can construct a random variable ξ such that conditional on the sample $\mathbf{x}$, ξ has the distribution $H(\cdot)$. We call ξ a CD-random variable (Singh et al., 2007) and its density as CD density.

To further elaborate the CD concept, we provide two examples, one for parameter from a continuous distribution and the other from a discrete distribution, the latter of which is more closely related to our problem setting but more challenging.

Example 1. Assume $X \sim \mathcal{N}(\mu, 1)$, we can re-express X as

$$X = \mu + U, \ \text{ where } U \sim \mathcal{N}(0, 1),$$

which is known as data generating equation Hannig et al. (2016). Inverting the above equation and plugging in the observed sample $X = x$, we have

$$\mu = x - u,$$

where u is an unobserved realization of U. If we estimate u by u^* that is randomly drawn from $\mathcal{N}(0, 1)$, then

$$\mu^* = x - u^*$$

can be considered as an estimate of μ. By repeating this process many times, we obtain a distribution estimator $\mu^* \sim \mathcal{N}(x, 1)$. In this case, μ^* is a CD-random variable and $\mathcal{N}(x, 1)$ is CD function for μ.

Similar idea can be applied to discrete distributions except that a CD function is not unique anymore. See the following example.

Example 2. The data generating equation for a Bernoulli distributed random variable $X \sim \text{Bernoulli}(p)$ is

$$X = I(U \leq p), \quad \text{where } U \sim \text{U}(0, 1).$$

Given an observed $X = x$, By inverting the data generating equation we obtain

$$\begin{cases} u \leq p \leq 1 & \text{if } x = 1 \\ 0 \leq p < u & \text{if } x = 0 \end{cases},$$

where u is an unobserved realization of U. Consider a random sample u^* from $\text{U}(0, 1)$ as an estimate for u, we obtain a CD-random variable for p. However, due to the intrinsic discrete nature of Bernoulli distribution, inverting the data generating equation results in an inequality. For instance, if we observe $x = 1$, the CD-random variable for the lower bound follows $\text{U}(0, 1)$ and the one for upper bound follows a distribution with all the probability mass at a value of 1. We will discuss this in depth in later sections.

Given the CD for a parameter of interest, we can perform almost all types of frequentist inference. For instance, consider $i.i.d$ sample $x_1 \cdots x_n$, where $x_i \sim \mathcal{N}(\mu, 1)$ for all i. To make inference on μ, we normally use $\bar{x}_n = \sum_{i=1}^{n} x_i/n$ as a point estimate or $(\bar{x}_n - 1.96/\sqrt{n}, \bar{x}_n + 1.96/\sqrt{n})$ as a 95% confidence interval. In comparison, under the CD approach, we use a CD for μ, which is $H(\mu) = \Phi(\sqrt{n}(\mu - \bar{x}_n))$ to conduct frequentist inference. For example, $(H^{-1}(\alpha/2), H^{-1}(1-\alpha/2)) = (\bar{x}_n + \Phi^{-1}(\alpha/2)/\sqrt{n}, \bar{x}_n + \Phi^{-1}(1-\alpha/2)/\sqrt{n})$ gives a $(1-\alpha)100\%$ confidence interval for μ, for any $0 \leq \alpha \leq 1$. In addition, $\bar{x}_n$, mean/median of $H(\mu)$, provides a point estimate for μ. The tail mass $H(b) = \Phi(\sqrt{n}(b - \bar{x}_n))$ yields a p-value for the one-sided hypothesis test $H_0 : \theta \leq b$ versus $H_a : \theta > b$.

Note that the CD functions mentioned above are also the fiducial distributions in Fisher's fiducial inference (see, e.g., Fisher (1935); Hannig (2009); Hannig et al. (2016)). In essence, both CD and fiducial approaches share the same goal in common in that both are aimed at providing a distribution estimator for parameters of interest via capturing all information that any given data contain about these parameters. It is not surprising that many fiducial distributions satisfy the conditions required for being a CD function. In general, the relationship between the fiducial distribution and the CD function is similar to the one between MLE and consistent estimator, where the latter is a much broader concept (see Fraser, 2011). We highlight the fact that although the idea originates from the fiducial approach, all of our interpretations and inferences reside within the frequentist framework and do not require any fiducial reasoning.

Throughout the article, we assume a 2×2 table with the following layout,

	Events	Non-events	
Exposure	X_{11}	X_{12}	n_1
Non-exposure	X_{21}	X_{22}	n_2
	m_1	m_2	N

We use X_{ij} and X_{ij}^* to indicate a random variable and its random copy, respectively, while x_{ij} and x_{ij}^* denote their corresponding realizations, for $i, j \in \{1, 2\}$. The column variable represents an outcome of interest, such as cure and non-cure, and the row variables is the explanatory variable, such as assigned treatment groups. In this paper, we consider drawing inference based on confidence intervals for log odds-ratio.

The rest of this paper is organized as follows. In Section 2.2 we derive CD functions of the log odds-ratio for 2×2 tables collected under all possible sampling schemes. We discuss how to construct confidence intervals through CD functions since it is not obvious for discrete distributions. In order to apply the proposed method to studies with rare events, we discuss how to incorporate prior information when analyzing 2×2 tables containing cells having zero counts. In Section 2.3, we conduct a simulation study to compare performance of the proposed CD method with two traditional approaches, namely confidence intervals from Wald method and Fisher's exact test. We also present

the sensitivity analysis to investigate robustness of the CD method under a rare events setting when prior information is mis-specified. In Section 2.4, we apply this novel approach to a real data set in a meta-analysis context to draw inference on drug safety. Finally, we conclude the paper with discussion in Section 2.5.

2.2   Methodology

In subsection 2.2.1—2.2.4, we derive CD functions of log odds-ratios in 2×2 tables under different sampling schemes. We then recommend two ways to construct confidence intervals through lower and upper CD functions for a log odds-ratio in subsection 2.2.5. To conclude this section, we explain how to handle 2×2 tables with zero counts using the CD approach.

2.2.1   Confidence distributions for the log odds-ratio from 2×2 tables without any sampling constraints

When the sampling scheme does not impose any constraints on marginal totals of a 2×2 table, cell counts X_{ij} are independent Poisson random variables denoted as

$$X_{ij} \sim P_{\lambda_{ij}}, \tag{2.1}$$

for $i, j \in \{1, 2\}$, where λ_{ij} is the Poisson mean parameter. Though rarely seen in reality, this can be the case for observational studies where data collection process is terminated arbitrarily. We refer to such a sampling scenario as Poisson sampling. By definition, the log odds-ratio θ is

$$\theta = \log(\lambda_{11}) - \log(\lambda_{12}) - \log(\lambda_{21}) + \log(\lambda_{22}). \tag{2.2}$$

In order to obtain the CD function for θ, we first derive CD function for the Poisson mean parameter.

Theorem 1. *If $X \sim P_\lambda$, then CD-random variables for λ follows a mixture distribution of $\Gamma(x, 1)$ and $\Gamma(x + 1, 1)$, where $\Gamma(\cdot)$ denotes Gamma distribution.*

Proof. Proof. It is well known that X and λ can be interpreted as the observed and expected number of times that an event occurred in a given time interval, respectively. For simplicity, let's assume it is the unit interval $[0,1]$. Then the inter-arrival times between two consecutive events satisfies $U_i \sim \text{Exp}(\lambda)$, for $i = 1, \cdots, X$, where $\text{Exp}(\cdot)$ denotes exponential distribution. And the arrival time of *X-th* event is defined as $T_X = \sum_{i=1}^{X} U_i \sim \Gamma(X, \lambda)$. It follows that

$$X \sim P_\lambda \Leftrightarrow T_X \leq 1 < T_{X+1}$$
$$\Leftrightarrow \frac{E_X}{\lambda} \leq 1 < \frac{E_{X+1}}{\lambda} \text{ , where } E_X \triangleq \Gamma(X, 1)$$
$$\Leftrightarrow E_X \leq \lambda < E_X + E_X', \text{ where } E_X' \sim \text{Exp}(1). \tag{2.3}$$

This shows that we have mapped the sample space $X \in \mathbb{N}$ to the parameter space $\left\{ \lambda : E_X \leq \lambda < E_X + E_X' \right\}$. Given the observed data $X = x$,

$$E_x \leq \lambda < E_x + E_x'. \tag{2.4}$$

Since E_x and E' are unobserved quantities with known distributions, we can estimate them by generating their corresponding random copies E_x^* and E'^*. By repeating this many times and denoting CD-random variable of λ as R_λ, we have

$$E_x^* \leq R_\lambda < E_x^* + E_x'^*. \tag{2.5}$$

If we define random variables for lower and upper bounds in (2.5) as $R_{\lambda,\text{L}}$ and $R_{\lambda,\text{U}}$, then $R_{\lambda,\text{L}} \sim \Gamma(x, 1)$ and $R_{\lambda,\text{U}} \sim \Gamma(x+1, 1)$ and any mixtures of $R_{\lambda,\text{L}}$ and $R_{\lambda,\text{U}}$ can be considered as CD-random variables for λ. ∎

Remark 1. Alternatively, we can obtain CD functions for λ by directly working on the cumulative distribution function of X. Define $H_x(\lambda) \triangleq P(X \leq x)$ as a sample dependent function on λ, and its inverse w.r.t. λ as $H_x^{-1}(\alpha)$, where $0 \leq \alpha \leq 1$. Then

$$\{X = k\} \Longleftrightarrow \{H_{k-1}(\lambda) < U \leq H_k(\lambda), \text{with } U \sim \text{Uniform}(0, 1)\}$$
$$\Longleftrightarrow \left\{ H_{k-1}^{-1}(U) < \lambda \leq H_k^{-1}(U) \right\} \triangleq \lambda_L < \lambda \leq \lambda_U. \tag{2.6}$$

Where (2.6) follows the fact that $H_x(\lambda)$ is a decreasing function in λ. Then cumulative distribution function of λ_L satisfies

$$G_{\lambda_L}(t) = P(\lambda_L \leq t)$$

$$= P\left(U \geq H_{k-1}(t)\right) \qquad (H_{k-1}(\lambda) \text{ is decreasing in } \lambda)$$

$$= 1 - \sum_{x=0}^{k-1} \frac{t^x}{x!} e^{-t}, \tag{2.7}$$

which is exactly the cumulative distribution function of $\Gamma(x, 1)$. Similarly, λ_U follows $\Gamma(x+1, 1)$. This is a more standard way of obtaining a CD function, and it can be particularly useful when there are no special distributional properties to exploit.

Apply theorem 1 to a 2×2 table collected from Poisson sampling, we have

$$E_{ij}^* \leq R_{ij} < E_{ij}^* + E_{ij}^{'*}, \tag{2.8}$$

with $E_{ij}^* \sim \Gamma(x_{ij}, 1)$ and $E_{ij}^{'*} \sim \text{Exp}(1)$, where E_{ij}^*, $E_{ij}^{'*}$ are mutually independent for $i, j \in \{1, 2\}$. Since the sampling is completely unconstrained, it is also reasonable to consider the sum of four cells $X = \sum_{i,j \in \{1,2\}} X_{ij} \sim P_{\sum_{i,j \in \{1,2\}} \lambda_{ij}}$, Therefore, we have

$$\sum_{i,j} E_{ij}^* \leq \sum_{i,j} R_{ij} < \sum_{i,j} E_{ij}^* + E^{'*}, \tag{2.9}$$

for $i, j \in \{1, 2\}$, where $E^{'*} \sim \text{Exp}(1)$. Combine conditions (2.8) and (2.9) and define the CD random variable for the log odds-ratio as $R_\theta \triangleq \log(R_{\lambda_{11}}) - \log(R_{\lambda_{12}}) - \log(R_{\lambda_{21}}) + \log(R_{\lambda_{22}})$. Consequently,

$$E_\theta^* - \log\left(1 + \frac{E^{'*}}{\min(E_{12}^*, E_{21}^*)}\right) \leq R_\theta < E_\theta^* + \log\left(1 + \frac{E^{'*}}{\min(E_{11}^*, E_{22}^*)}\right) \tag{2.10}$$

Where $E_\theta^* = \log(E_{11}^*) - \log(E_{12}^*) - \log(E_{21}^*) + \log(E_{22}^*)$. If any of $R_{ij} = 0$, one or both of the bounds is equal to $\pm\infty$, respectively.

2.2.2 Confidence distributions for the log odds-ratio from 2×2 tables with sampling conditional on one margin

Binomial sampling is one of the most popular sampling schemes in scientific research, where either row or column marginal totals are fixed beforehand. For instance, a

randomized clinical trial or cohort study have row totals fixed in advance, while case-control has column totals fixed.

Without loss of generality, let's assume the row totals are fixed, it follows directly that X_{11} and X_{21} are independent binomial random variables denoted as $X_{11} \sim \text{Bin}(n_1, p_1)$ and $X_{21} \sim \text{Bin}(n_2, p_2)$, where $p_i = \lambda_{i1}/(\lambda_{i1} + \lambda_{i2})$, for $i \in \{1, 2\}$. The log odds-ratio can then be expressed as

$$\theta = \log(p_1) - \log(1 - p_1) - \log(p_2) + \log(1 - p_2).$$

To derive CD function for θ, we first present CD function for parameter p in a binomial distribution.

Theorem 2. *If $X \sim Bin(n, p)$, then a CD random variable for p follows a mixture distribution of $B(x + 1, n - x)$ where $B(x, n - x + 1)$, $B(\cdot, \cdot)$ denotes Beta distribution.*

Proof. Proof By definition, we can write $X = \sum_{k=1}^{n} B_k$, where B_k are Bernoulli distributed random variables with $P(B_k = 1) = p = 1 - P(B_k = 0)$. As we demonstrated earlier, for each trial k,

$$\begin{cases} U_k \leq p \leq 1 & \text{if } B_k = 1 \\ 0 \leq p < U_k & \text{if } B_k = 0 \end{cases}$$

, with $U_k \sim \text{U}(0, 1)$. With all n trials taken into considerations,

$$\max\{U_k, \text{ s.t. } B_k = 1\} \leq p < \min\{U_k, \text{ s.t. } B_k = 0\} \Rightarrow U_{(X)} \leq p < U_{(X+1)}, \quad (2.11)$$

where $U_{(\cdot)}$ indicates the order statistics of U_k for $k = 1, \cdots, n$. This shows that we have mapped the sample space $X \in \mathbb{N}$ to the parameter space $\{p : U_{(X)} \leq p < U_{(X+1)}\}$. Given the observed data $X = x$, a CD random variable R_p of p satisfies

$$U^*_{(x)} \leq R_p < U^*_{(x+1)}. \quad (2.12)$$

It is well known that $U^*_{(x)} \sim \text{B}(x, n - x + 1)$ and $U^*_{(x+1)} \sim \text{B}(x + 1, n - x)$. So CD function for p is a mixture of $\text{B}(x, n - x + 1)$ and $\text{B}(x + 1, n - x)$. ∎

Apply theorem 2 to a 2×2 table with fixed row totals, we have

$$U^*_{(x_{i1})} \leq R_{p_i} < U^*_{(x_{i1}+1)}, \tag{2.13}$$

for $i \in \{1, 2\}$. Use the notation in Section 2.1 and take advantage of the relation between gamma and beta distribution, $E^*_{i1}/(E'^*_i + E^*_{i1} + E^*_{i2})$ and $(E^*_{i1} + E'^*_i)/(E'^*_i + E^*_{i1} + E^*_{i2})$ can be shown to have the same distribution as $U^*_{(x_{i1})}$ and $U^*_{(x_{i1}+1)}$, respectively. So (2.13) can be rewritten as

$$\frac{E^*_{i1}}{E'^*_i + E^*_{i1} + E^*_{i2}} \leq R_{p_i} \leq \frac{E^*_{i1} + E'^*_i}{E'^*_i + E^*_{i1} + E^*_{i2}}, \tag{2.14}$$

for $i \in \{1, 2\}$. A simple calculation shows that the CD random variable for θ satisfies

$$E^*_\theta - \log\left(1 + \frac{E'^*_1}{E^*_{12}}\right) - \log\left(1 + \frac{E'^*_2}{E^*_{21}}\right) \leq R_\theta$$

$$\leq E^*_\theta + \log\left(1 + \frac{E'^*_1}{E^*_{11}}\right) + \log\left(1 + \frac{E'^*_2}{E^*_{22}}\right). \tag{2.15}$$

If any of $E^*_{ij} = 0$, one or both of the bounds is equal to $\pm\infty$, respectively.

Remark 2. Notice that (2.15) can also be obtained from the condition (2.8) in unconstrained case with additional constraints

$$\sum_{j=1}^{2} R^*_{ij} \leq E'^*_i + \sum_{j=1}^{2} E^*_{ij},$$

for $i \in \{1, 2\}$. This change reflects the fact that we are conditioning on the observed value of $X_{i1} + X_{i2} = n_i$ and, therefore it is not necessary to pool all cells as in Poisson sampling.

2.2.3 Confidence distributions for the log odds-ratio from 2×2 tables with sampling conditional on the grand total

Under the Poisson model (2.1), when the grand total is fixed, $\mathbf{X} = (X_{11}, X_{12}, X_{21}, X_{22}) \sim \text{multinomial}(n, p_{11}, p_{12}, p_{21}, p_{22})$ with $p_{ij} = \lambda_{ij}/\sum_{k,l \in \{1,2\}} \lambda_{kl}$ for $i, j \in \{1, 2\}$. This sampling happens often in cross-sectional studies where only total number of participants is predetermined and neither column nor row margins are

known. We call this scheme as multinomial sampling. The log odds-ratio in this case is defined as

$$\theta = \log(p_{11}) - \log(p_{12}) - \log(p_{21}) + \log(p_{22}). \tag{2.16}$$

Similarly, we present CD function for $\mathbf{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ first.

Theorem 3. If $\mathbf{X} = (X_1, \ldots, X_k)$ follows multinomial $(n, p_1, \ldots, p_k)$ with $\sum_{i=1}^{k} p_i = 1$, then a CD random variable for $\mathbf{p} = (p_1, \ldots, p_k)$ satisfies

$$\left\{ \mathbf{p} : W_i \le p_i \le 1 - \sum_{j \ne i}^{n} W_i, i = 1, \ldots, k \right\}$$

where $\mathbf{W} = (W_0, W_1, \cdots, W_k)$ follows Dirichlet$(1, X_1, \cdots, X_k)$

Proof. Proof. First, analogous to binomial distribution scenario, by considering each p_i separately, we have $V_i \le p_i \le 1$ where $V_i \overset{i.i.d}{\sim} \text{B}(X_i, n - X_i + 1)$. This is because V_i has the same distribution as $U_{(X_i)}$ where $U_i \overset{i.i.d}{\sim} \text{U}(0,1)$ with $i = 1, \cdots, n$. In addition, since $\sum_{i=1}^{k} p_i = 1$ we have $\sum_{i=1}^{k} V_i \le 1$. In summary, each p_i satisfies

$$V_i \le p_i \le 1 - \sum_{j \ne i} V_j. \tag{2.17}$$

Consider the following random vector W whose distribution is taken as a conditional distribution

$$\mathbf{W} = (W_0, W_1, \cdots, W_k) \triangleq (1 - V_1 - \cdots - V_k, V_1, \cdots, V_k) | \{V_1 + \cdots + V_k \le 1\}.$$

Then $f_{\mathbf{W}}$, the density of $\mathbf{W}$, can be shown as

$$f_{\mathbf{W}}(\mathbf{w}) \propto \prod_{j=0}^{k} X_j w_j^{X_j - 1},$$

where $X_0 = 1$. By definition it follows Dirichlet$(1, X_1, \cdots, X_k)$. ∎

Follow the notations introduced before, by leveraging the relationship between Gamma and Dirichlet distributions, we can show $(E'^*, E_{11}^*, \cdots, E_{22}^*)/(E'^* + E_{11}^* + \cdots + E_{22}^*)$ has the same distribution as $\mathbf{W}$. So replacing all the V_i in (2.17) by E_{ij}^*, for realization $\mathbf{x} = (x_1, \ldots, x_k)$ of $\mathbf{X}$, it follows

$$\frac{E_{ij}^*}{E'^* + E_{11}^* + E_{12}^* + E_{21}^* + E_{22}^*} \le R_{p_{ij}} \le \frac{E_{ij}^* + E'^*}{E'^* + E_{11}^* + E_{12}^* + E_{21}^* + E_{22}^*}, \tag{2.18}$$

for $i, j \in \{1, 2\}$. Since $\sum_{k,l} p_{kl} = 1$, it is reasonable to assume $\sum_{k,l} R_{p_{kl}} \leq 1$. Basically, this yields the same conditions for R_θ as in the Poisson sampling, so the resulting CD function for θ in this case is identical to (2.10).

2.2.4 Confidence distributions for the log odds-ratio from 2×2 tables with sampling conditional on both margins

Though hardly seen in practice, a classic example of such sampling scheme is Fisher's tea tasting experiment (Fisher, 1956). In this setting, assume only X_{11} can vary after conditional on both marginal totals, we have

$$P(X_{11} = x | m_1, m_2, n_1, n_2) = \frac{\binom{n_1}{x}\binom{n_2}{m_1-x}e^{\theta x}}{\sum_{k=l}^{u}\binom{n_1}{k}\binom{n_2}{m_1-k}e^{\theta k}},$$

where $l = \max(0, x_{11} - x_{22})$, $u = \min(n_1, m_1)$. This is Fisher's non-central hypergeometric distribution and, when $\theta = 0$, it reduces to a hypergeometric distribution. The log odds-ratio in this case is θ itself. And we may call such sampling scheme hypergeometric sampling.

Similar to remark 1, we derive a CD function for θ by directly inverting the above conditional distribution. Define $H_x(\theta) = P_\theta(X_{11} \leq x)$ as a sample dependent function on θ and its inverse w.r.t. θ as $H_x^{-1}(\alpha)$ where $0 \leq \alpha \leq 1$, then

$$\{\text{Observing } X_{11} = x\} \iff \{H_{x-1}(\theta) < U \leq H_x(\theta), \text{with } U \sim U(0,1)\}$$

$$\iff \left\{H_{x-1}^{-1}(U) < \theta \leq H_x^{-1}(U)\right\} \triangleq \theta_L < \theta \leq \theta_U, \qquad (2.19)$$

where (2.19) follows the fact that $H_x(\theta)$ is a decreasing function in θ. The cumulative distribution function of θ_L can then be obtained as

$$G_{\theta_L}(t) = P(H_{x_{11}-1}^{-1}(U) \leq t)$$

$$= P(U \geq H_{x_{11}-1}(t)) \qquad (H(\cdot) \text{ is decreasing in t})$$

$$= \frac{\sum_{k=x_{11}}^{u}\binom{n_1}{k}\binom{n_2}{m_1-k}e^{tk}}{\sum_{k=l}^{u}\binom{n_1}{k}\binom{n_2}{m_1-k}e^{tk}} \qquad (2.20)$$

and similarly,

$$G_{\theta_U}(t) = \frac{\sum_{k=x_{11}+1}^{u}\binom{n_1}{k}\binom{n_2}{m_1-k}e^{tk}}{\sum_{k=l}^{u}\binom{n_1}{k}\binom{n_2}{m_1-k}e^{tk}}, \qquad (2.21)$$

Any mixtures distributions between θ_L and θ_U is a CD function for θ.

2.2.5 Constructing confidence intervals using CD functions

So far, we have obtained lower and upper CD functions for the parameter of interest in a discrete distribution, now we discuss two ways of constructing confidence intervals from them. The first one is to use $[q\{\alpha/2, \mathrm{CD_{lower}}\}, q\{1-\alpha/2, \mathrm{CD_{upper}}\}]$, we call it raw method since we merely take quantiles from both lower and upper CD functions without melding them into a single one. The second way is to meld a CD function first through a 50-50 mixture between the lower and upper densities, i.e., $\mathrm{CD_{meld}} = b \times \mathrm{CD_{lower}} + (1-b) \times \mathrm{CD_{upper}}$, where $b \sim \mathrm{Bernoulli}(0.5)$, and then construct the $1-\alpha$ confidence interval from the melded CD in a usual way as $[q\{\alpha/2, \mathrm{CD_{meld}}\}, q\{1-\alpha/2, \mathrm{CD_{meld}}\}]$. We may also refer to it as a half-corrected version. See Figure i as an illustration.

Regardless of how we meld a single CD function from the lower and upper CDs, all CD functions should converge to the true distribution when sample size goes to infinity. Their corresponding asymptotic properties are guaranteed and have extensively been studied (e.g., see theorem 3 in Hannig (2009)).

2.2.6 Inference on 2×2 table with zero events

When there are observed zero events in a 2×2 table, the corresponding $E_{ij}^* = 0$. This would leads to infinite confidence bounds on the log odds-ratio for (2.10) and (2.15). We must keep this bound when using conservative (raw) version of the proposed CD approach. However, with the melded version we can use a weak a-prior to replace $E_{ij}^* = 0$ in order to obtain a non infinite confidence bound for the parameter of interest.

Let r_{ij} denotes the lower bound on the expected number of observed counts. This is often available. For instance, in mortality studies that usually suffer from low observed event counts, we have $r_{ij} = n_i p_0$, where n_i is the number of participants in a control or treatment arm, p_0 is the background mortality rate, i.e., the rate of people dying during the course of the study from unrelated causes, which is usually well known.
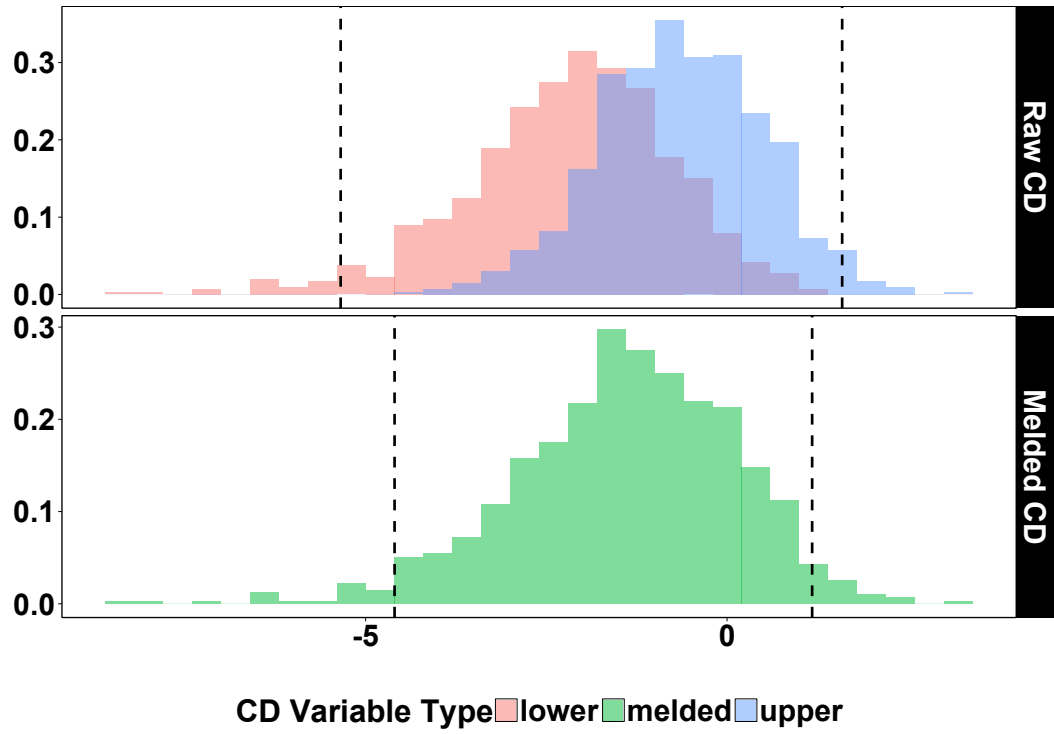
Figure i: CD densities and the corresponding 95% confidence intervals for data $(x_{11}, x_{12}, x_{21}, x_{22}) = (1, 9, 3, 7)$. In the upper panel, lower and upper CD densities are plotted along with lower 2.5% and upper 97.5% quantile from lower and upper CD densities, respectively. The bottom panel shows the half-corrected melded version of the CD density and its associated 95% confidence interval.

Then, whenever $x_{ij} = 0$, we replace $E_{ij}^* = 0$ with $\tilde{E}_{ij}^* \sim \Gamma(r_{ij}, 1)$.

Effectively, with such modifications, when generating samples for confidence bounds (2.10) developed under Poisson or multinomial sampling, we use $r_{ij} = \min(n_i p_0, 1/4)$, we also need to replace $E^{'*}$ with $\tilde{E}^{'*} \sim \Gamma(1 - \sum_{\{ij:x_{ij}=0\}} r_{ij}, 1)$. For confidence bounds (2.15) obtained from binomial sampling, we use $r_{ij} = \min(n_i p_0, 1/2)$ and substitute $E_i^{'*}$ with $\tilde{E}_i^{'*} \sim \Gamma(1 - \sum_{\{j:x_{ij}=0\}} r_{ij}, 1)$. There is no update for the zero events case to the melded version for the hypergeometric sampling scheme.

2.3 Simulation Studies

We now proceed to examine the performance of our approach through simulation studies. First, we evaluate actual coverages and median widths of the 95% confidence bounds constructed from the proposed approach and two classical methods, Wald and Fisher's exact method, for 2×2 tables that were simulated under different sampling schemes. Second, we conduct sensitivity analysis to investigate robustness of the proposed method against mis-specified prior information for 2×2 tables collected under binomial sampling with low event rates.

2.3.1 Simulations for the comparison of three different methods

The simulation setting for the first analysis is described as below. Within each sampling scheme, we generate 200 2×2 balanced tables (i.e., expected row totals are the same), with an expected study total sample size of 20 for the unconstrained sampling and with exactly 20 for all other sampling methods. For Poisson case, λ_{11} varies from 1 to 9, and the rest of the parameters are set as $\lambda_{22} = 10 - \lambda_{12} = 10 - \lambda_{21} = \lambda_{11}$ so that the true log odds-ratio $= 2\log(\lambda_{11}/(10 - \lambda_{11}))$ is a monotonic function of λ_{11}. Similarly, for multinomial sampling, we have p_{11} ranging from 0.01 to 0.49 with $p_{22} = 0.5 - p_{12} = 0.5 - p_{21} = p_{11}$. For the binomial case, p_1 ranges from 0.01 to 0.99 with $p_1 = 1 - p_2$. We directly vary odds-ratio in the hypergeometric sampling scenario. Subsequently, we use the original parameters in each sampling scheme rather than log

odds-ratio to track performance since this gives us a clearer picture when rare events arise. In terms of results, all the methods are compared regardless of under which sampling scenario data were simulated. Finally, in the case of cell with zero observed counts, when generating CD-random variable samples for melded confidence intervals, we apply prior information as $r_{ij} = \min(\lambda_{ij}, 1/4)$ for Poisson; $r_{ij} = \min(20p_{ij}, 1/4)$ for multinomial; and $r_{ij} = \min(10p_{ij}, 1/2)$ for binomial sampling to accommodate different simulation settings.

Simulation results are summarized in Figures ii to v and Tables i to iv. In general, we observe similar patterns across all sampling schemes. First, between different CD approaches, we see that when the sampling method used to construct confidence intervals is aligned with the one used to generate data, the corresponding confidence intervals perform the best in terms of both empirical coverage and median width. Second, for the CD approach, raw versions of confidence interval are always more conservative than the melded versions, meaning that their empirical coverages are always greater than the nominal ones, while the melded versions can guarantee coverage on average and they tend to have a narrower width. Third, despite the fact that CD based confidence intervals behave similar to the ones built from traditional methods in the middle range of the parameter space, their performance was clearly better when parameters fell on boundaries of their domain. Finally, across all sampling schemes, the melded version of confidence intervals from binomial sampling performed consistently well. Considering the binomial sampling scheme is widely seen in scientific research, we recommend confidence intervals constructed under binomial sampling when the actual sampling scheme is unknown.

2.3.2 Simulation for sensitivity analysis

As discussed in Section 2.5, when there are zero observed cells in a 2×2 table, we adopt a weak prior for the melded versions of the CD based confidence interval in order to obtain finite bounds. Here, we perform a sensitivity analysis to investigate the impact of prior used on the constructed confidence intervals. The simulation generates
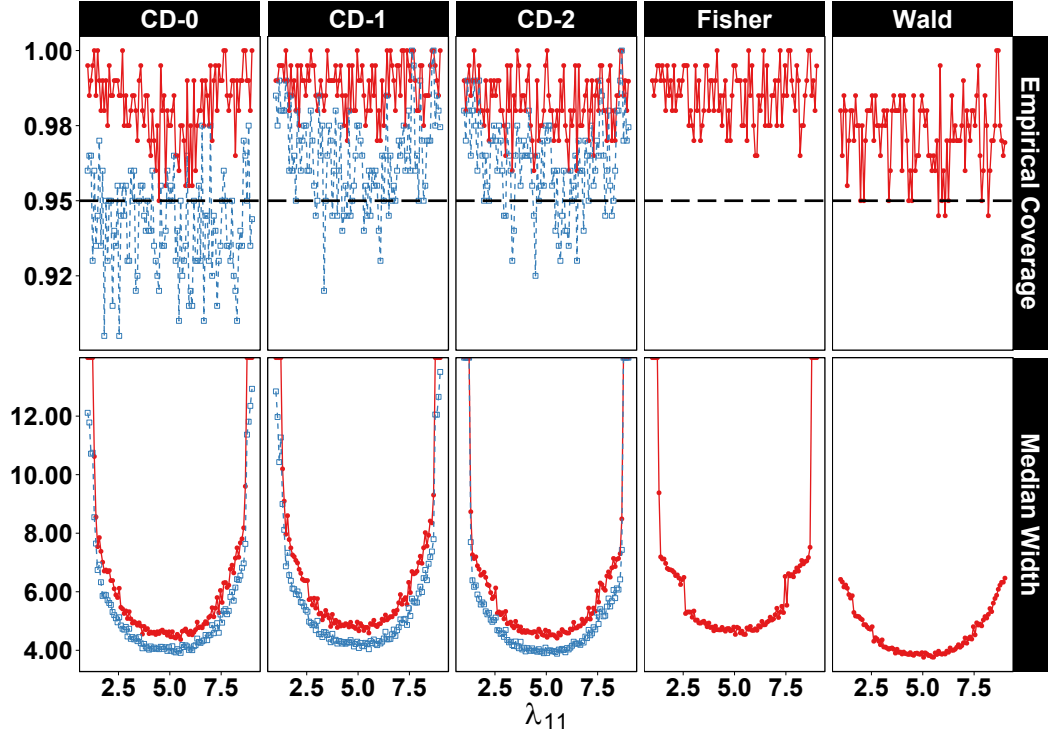
Figure ii: Empirical coverage (top) and median width (bottom) of 95% confidence intervals constructed using different methods for data generated under Poisson sampling. X-axis is the expected counts for X_{11}. Original distribution parameters are set up as $\lambda_{11} = \lambda_{22} = 10 - \lambda_{12} = 10 - \lambda_{21}$ so that log odds-ratio is monotonic in λ_{11}, where λ_{ij} is the mean parameter for X_{ij} in a 2×2 table. Results are calculated based on 200 simulated 2×2 tables with an expected total sample size of 20. Plots for different methods are shown in columns. "CD-0," "CD-1," and "CD-2" indicate CD approaches under Poisson/multinomial, binomial and hypergeometric sampling, respectively. "Fisher" and "Wald" denote confidence intervals using Fisher's exact test and Wald method. For CD based methods, red solid lines connected with ($\bullet$) are confidence intervals using the raw method described in the text, and the blue dashed lines connected with ($\square$) are confidence intervals using the half-corrected method.

Table i: Empirical coverage and median width of 95% confidence intervals for data generated under Poisson sampling

Method	Empirical Coverage	Median Width	Empirical Coverage	Median Width
	$\lambda_{11} = 1$ (LOR $= -4$)		$\lambda_{11} = 5$ (LOR $= 0$)	
CD-0.raw	1.00	Inf	0.98	4.46
CD-0.hc	0.96	12.11	0.96	3.98
CD-1.raw	0.99	Inf	0.99	4.67
CD-1.hc	0.99	12.85	0.98	4.08
CD-2.raw	0.99	Inf	1.00	4.44
CD-2.hc	0.98	Inf	0.97	3.91
Wald	0.98	6.42	0.98	3.76
Fisher	0.99	Inf	1.00	4.54

Note: λ_{11} is the Poisson mean parameter for X_{11}. Two log odds-ratio (LOR) values are shown, one in the middle range and the other close to the boundary of the parameter space of λ_{11} for a rare events scenario. CD-0, CD-1, and CD-2 are CD methods derived under Poisson/multinomial, binomial and hypergeometric sampling, respectively. Both raw and half-corrected (hc) versions described in the article are presented. "Inf" stands for infinity.

Table ii: Empirical coverage and median width of 95% confidence intervals for data generated under multinomial sampling

Method	Empirical Coverage	Median Width	Empirical Coverage	Median Width
	$p_{11} = 0.01$ (LOR $= -9$)		$p_{11} = 0.25$ (LOR $= -2$)	
CD0 (raw)	1.00	Inf	0.97	4.46
CD0 (hc)	0.99	18.04	0.95	3.92
CD1 (raw)	0.99	Inf	0.98	4.68
CD1 (hc)	0.99	10.73	0.96	4.16
CD2 (raw)	0.99	Inf	0.99	4.41
CD2 (hc)	0.97	Inf	0.96	3.83
Wald	0.95	8.04	0.97	3.70
Fisher	0.99	Inf	0.99	4.46

Note: p_{11} is the probability parameter for X_{11} in a multinomial distribution. Two log odds-ratio (LOR) values are shown, one in the middle range and one close to boundary of parameter space of p_{11} for rare events scenario. CD-0, CD-1 and CD-2 are CD methods derived under Poisson/multinomial, binomial and hypergeometric sampling, respectively. Both raw and half-corrected (hc) versions described in the article are presented. "Inf" stands for infinity.
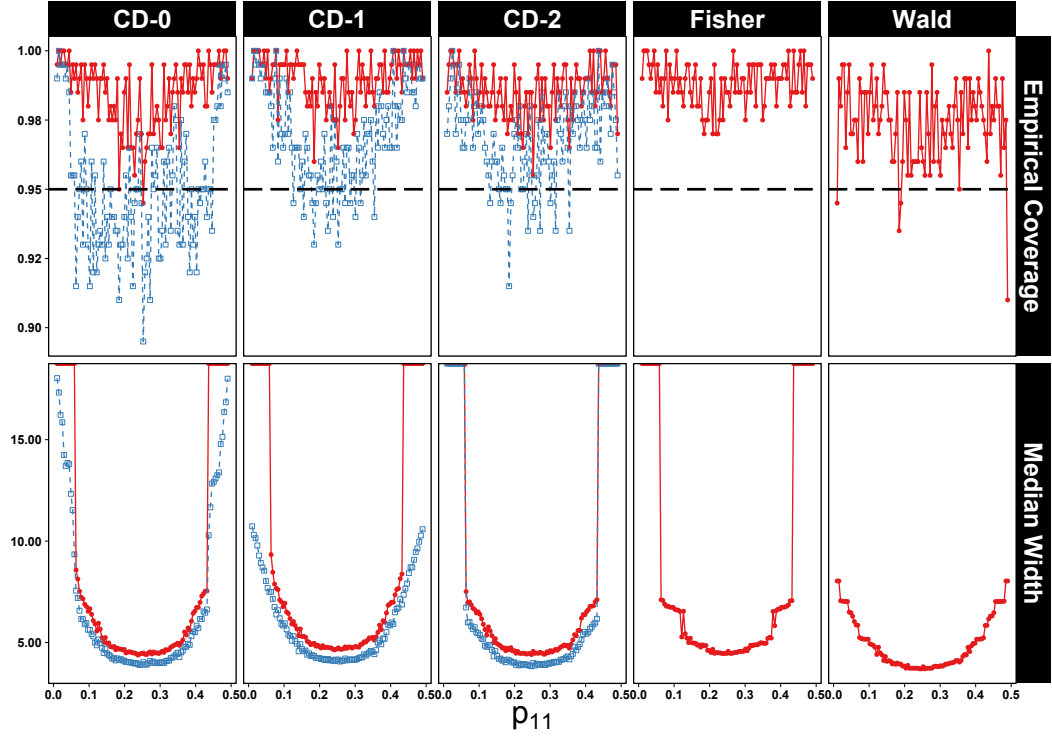
Figure iii: Empirical coverage (top) and median width (bottom) of 95% confidence intervals constructed for different methods for data generated under multinomial sampling. X-axis is the probability parameter for X_{11}. Original distribution parameters are set up as $p_{11} = p_{22} = 0.5 - p_{12} = 0.5 - p_{21}$ so that log odds-ratio is monotonic in p_{11}, where p_{ij} is the probability parameter for X_{ij} in a multinomial distribution. Results are calculated based on 200 simulated 2×2 tables with sample size of 20. Plots for different methods are shown in columns. "CD-0," "CD-1," and "CD-2" indicate CD approaches under Poisson/multinomial, binomial and hypergeometric sampling, respectively. "Fisher" and "Wald" denote confidence intervals using Fisher's exact test and Wald method. For CD based methods, red solid lines connected with ($\bullet$) are confidence intervals using the raw method described in the text, and the blue dashed lines connected with ($\square$) are confidence intervals using the half-corrected method.
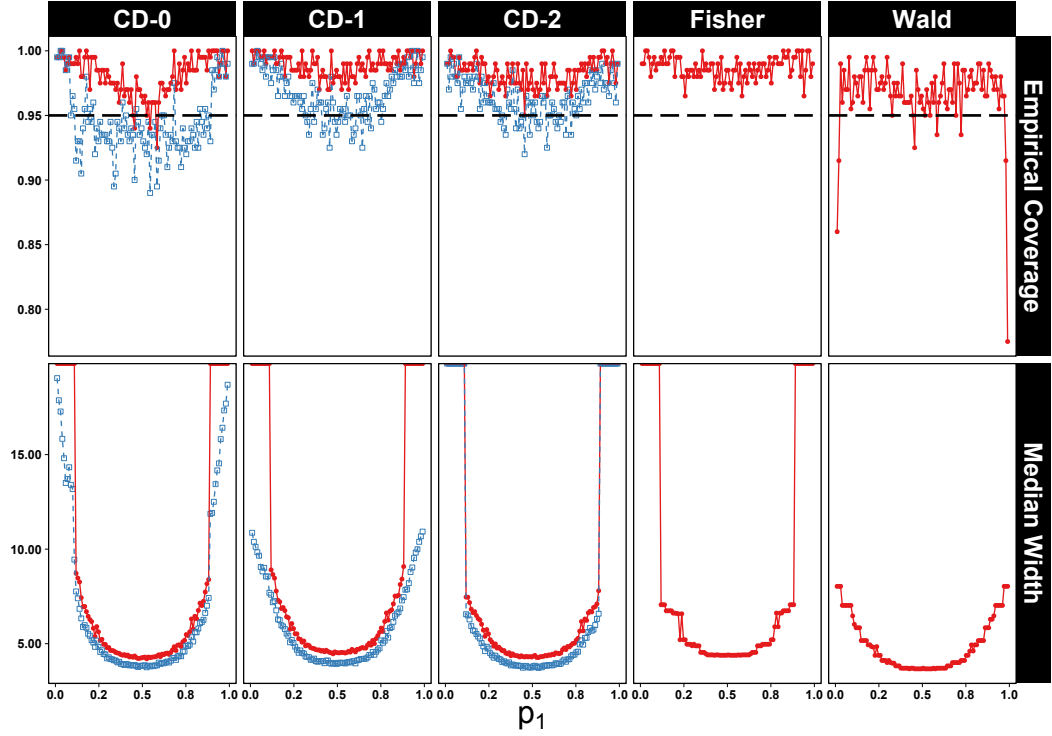
Figure iv: Empirical coverage (top) and median width (bottom) of 95% confidence intervals constructed for different methods for data generated under binomial sampling. X-axis is the probability parameter for X_{11}. Original distribution parameters are set up as $p_1 = 1 - p_2$ so that log odds-ratio is monotonic in p_1, where p_i is the probability parameter for X_{i1} in a binomial distribution. Results are calculated based on 200 simulated 2×2 tables with a total sample size of 20. Plots for different methods are shown in columns. "CD-0," "CD-1," and "CD-2" indicate CD approaches under Poisson/multinomial, binomial and hypergeometric sampling, respectively. "Fisher" and "Wald" denote confidence intervals using Fisher's exact test and Wald method. For CD based methods, red solid lines connected with ($\bullet$) are confidence intervals using the raw method described in the text, and the blue dashed lines connected with ($\square$) are confidence intervals using the half-corrected method.

Table iii: Empirical coverage and median width of 95% confidence intervals for data generated under binomial sampling

Method	Empirical Coverage	Median Width	Empirical Coverage	Median Width
	$p_1 = 0.01$ (LOR $= -9$)		$p_1 = 0.5$ (LOR $= 0$)	
CD-0 (raw)	1.00	Inf	0.96	4.28
CD-0 (hc)	1.00	19.05	0.93	3.83
CD-1 (raw)	1.00	Inf	0.98	4.51
CD-1 (hc)	0.99	10.87	0.95	3.96
CD-2 (raw)	0.99	Inf	0.97	4.34
CD-2 (hc)	0.99	Inf	0.95	3.81
Wald	0.86	8.03	0.96	3.70
Fisher	0.99	Inf	0.99	4.41

Note: p_1 is the probability parameter for X_{11} in a binomial distribution. Two log odds-ratio (LOR) are shown, one in the middle range and the other close to boundary of the parameter space of p_1 for a rare events scenario. CD-0, CD-1, and CD-2 are CD methods derived under Poisson/multinomial, binomial and hypergeometric sampling, respectively. Both raw and half-corrected (hc) versions described in the article are presented. "Inf" stands for infinity.

Table iv: Empirical coverage and median width of 95% confidence intervals for data generated under hypergeometric sampling

Method	Empirical Coverage	Median Width	Empirical Coverage	Median Width
	LOR $= -5$		LOR $= 0$	
CD-0 (raw)	1.00	Inf	0.96	4.05
CD-0 (hc)	0.99	17.10	0.95	3.70
CD-1 (raw)	1.00	Inf	0.98	4.40
CD-1 (hc)	1.00	10.11	0.96	3.89
CD-2 (raw)	0.99	Inf	0.97	4.10
CD-2 (hc)	0.99	Inf	0.97	3.58
Wald	0.97	8.03	0.97	3.58
Fisher	1.00	Inf	0.97	4.23

Note: Two log odds-ratio (LOR) are shown, one in the middle range and the other close to boundary of its parameter space for a rare events scenario. CD-0, CD-1, and CD-2 are CD methods derived under Poisson/multinomial, binomial and hypergeometric sampling, respectively. Both raw and half-corrected (hc) versions described in the article are presented. "Inf" stands for infinity.
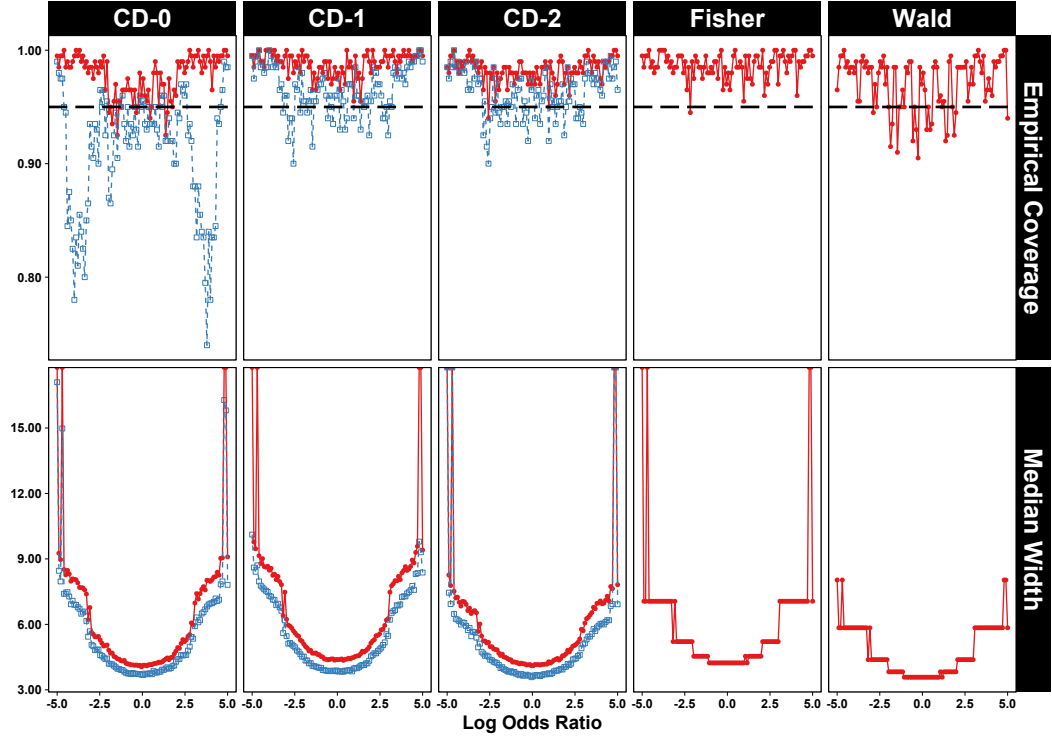
Figure v: Empirical coverage (top) and median width (bottom) of 95% confidence intervals constructed for different methods for data generated under hypergeometric sampling. Results are calculated based on 200 simulated 2×2 tables with a total sample size of 20. Plots for different methods are shown in columns. "CD-0," "CD-1," and "CD-2" indicate CD approaches under Poisson/multinomial, binomial and hypergeometric sampling, respectively. "Fisher" and "Wald" denote confidence intervals using Fisher's exact test and Wald method. For CD based methods, red solid lines connected with ($\bullet$) are confidence intervals using the raw method described in the text, and the blue dashed lines connected with ($\square$) are confidence intervals using the half-corrected method.

samples under binomial sampling, and the methods considered are melded versions of the CD-based confidence intervals for Poisson/multinomial and binomial sampling.

To resemble large size clinical trials with low rates of event, we simulate data sets with a fixed total sample size of 1000 for both control and treatment arms. One thousand 2×2 tables are generated for each of three different true log odds-ratios, -1, 0 and 1. To ensure low counts in the events cells X_{11} and X_{21}, first, the probability of the events of interest for a control group, p_2, is randomly generated from $\mathrm{U}(0.001, 0.005)$, and p_1 for a treatment group is then calculated as $\mathrm{logit}(p_1) = \log(\text{odds-ratio}) + \mathrm{logit}(p_2)$. In addition, we set the true global background events rate $p_0 = 10^{-4}$ and impose that rate on the simulated 2×2 tables by generating events cells X_{i1} as the sum of samples from $\mathrm{Bin}(1000, p_i)$ and $\mathrm{Bin}(1000, p_0)$, for $i \in \{1, 2\}$. Then, for each generated table, we construct the melded confidence intervals under background rates $\tilde{p}_0$ ranging from $0.01p_0$ to their upper bound, depending on the method used.

Figures vi and vii show the results for sensitivity analysis. In Figure vi, we see that both melded confidence intervals can achieve nominal coverage, on average,e across different $\tilde{p}_0$ as well as true log odds-ratios. Meanwhile, Figure vii indicates that the median width is relatively stable for different $\tilde{p}_0$, although it decreases when true log odds-ratio increases, which is reasonable given our simulation setting. In addition, analyses with sample sizes of 100 and 10000 are also conducted and we observe similar themes. In summary, it is clear that neither empirical coverage nor median width is influenced too much by the prior information used. This demonstrates robustness of the proposed CD methods when analyzing 2×2 tables with zero cell counts.

2.4   Real Data Analysis

Rare events data can often be seen in large-sample clinical trials with extremely low events rates. In such cases, a single study may be inadequate for drawing reliable conclusions, yet it usually can be strengthened by using meta-analysis to synthesize information from multiple similar studies. In this section, we apply the proposed CD
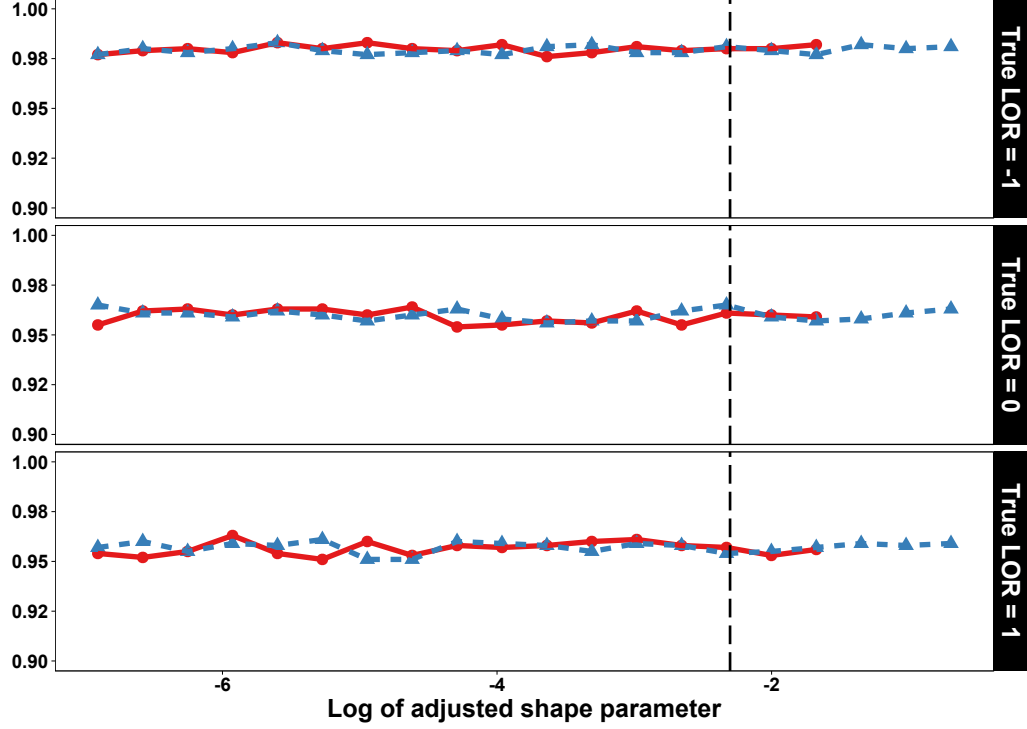
Figure vi: Empirical coverage of 95% confidence intervals constructed under Poisson/multinomial and binomial sampling. X-axis shows adjusted Gamma shape parameters values $\tilde{r}_{ij} = 1000\tilde{p}_0$ in log scale. A vertical line indicates a true value for the adjusted shape parameter used to generate data. Panels from top to bottom are shown for three different true log odds-ratios. Red solid lines connected with (●) and the blue dashed lines connected with (▲) denote the melded version of confidence intervals from data generated under Poisson/multinomial and binomial sampling, respectively.
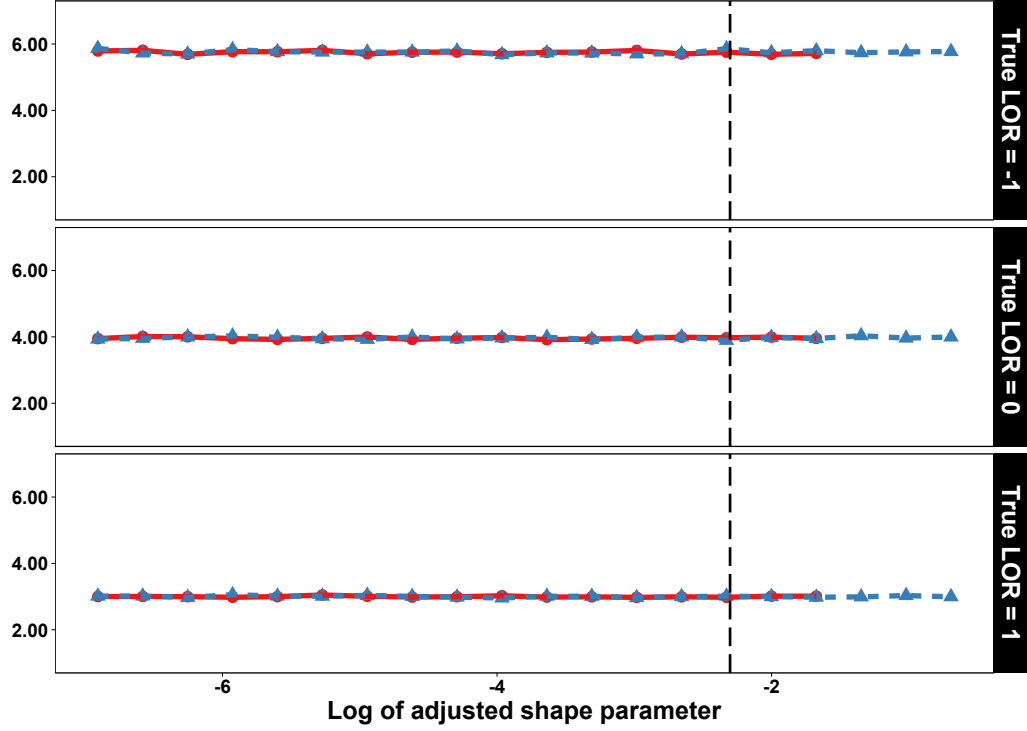
Figure vii: Median width of 95% confidence intervals constructed under Poisson/multinomial and binomial sampling. X-axis shows adjusted Gamma shape parameters values $\tilde{r}_{ij} = 1000\tilde{p}_0$ in log scale. A vertical lines indicates true value of adjusted shape parameter used to generate data. Panels from top to bottom are shown for three different true log odds-ratios. Red solid lines connected with ($\bullet$) and the blue dashed lines connected with ($\blacktriangle$) denote the melded version of confidence intervals from data generated under Poisson/multinomial and binomial sampling, respectively.

approach in the context of meta-analysis to a real data set from Nissen and Wolski (2007) for drug safety evaluation. The data contains 48 clinical trials that were used to assess the risk of myocardial infarction and death from cardiovascular causes associated with the drug Avandia. Several methods are considered in our analysis, namely, the well known Mantel-Haenszel and Peto's methods both with and without artificial correction for zero events studies, the approaches by combining p-value functions and our proposed CD functions, using the combining recipe discussed in Liu et al. (2014). In addition, we also conduct a small simulation study to validate the performance of the proposed method.

Table v: Analysis result of the Avandia data

method	Myocardial infarction		Cardiovascular death	
	95% CI	p-value	95% CI	p-value
Peto	(1.031,1.979)	0.032	(0.980,2.744)	0.060
Peto-0.5	(0.921,1.659)	0.158	(0.760,1.690)	0.538
MH	(1.029,1.978)	0.033	(0.983,2.929)	0.057
MH-0.5	(0.919,1.647)	0.163	(0.759,1.689)	0.541
Liu et al.	(0.972,2.001)	0.071	(0.765,2.965)	0.252
proposed CD	(0.975,2.013)	0.067	(0.758,2.904)	0.257

Note: CI denotes confidence interval; method names ending with "0.5" indicates 0.5 is added to zero cells; Liu et al. is the exact meta-analysis method by combing *p*-value functions discussed in Liu et al. (2014); proposed CD is the method by combining melded version of CD densities derived from binomial sampling.

2.4.1 Real data analysis results

Table v presents meta-analysis results under various methods for both odds-ratios of myocardial infarction and death due to cardiovascular causes. For the endpoint of myocardial infarction, when there is no artificial correction for zero cells, neither confidence intervals obtained from Mantel-Haenszel (MH) nor Peto's (Peto) method contains 1, indicating a significant association between Avandia and myocardial infarction at 95% confidence level. However, after applying the 0.5 correction to zero events, both confidence intervals yield conclusions of no associations at 95% confidence level. Such observations imply that the use of corrections to zero events may result in contradictory conclusions, which has been discussed extensively in J Sweeting et al. (2004). In

comparison, combining the proposed melded CDs from binomial sampling performed similarly to the exact analysis by combining p-value functions using the combining recipe discussed in Liu et al. (2014), although the former produced narrower confidence intervals.

2.4.2   Meta-analysis simulations

In this simulation, we generate meta-analysis data sets that mimic the structure of Avandia data and compare the behaviors of different approaches across a range of true odds-ratios. binomial sampling is used to generate data for individual studies and the row marginal totals in each study are set to the numbers as those in the Avandia data. The 2×2 table for study i is formed by pairs of independent binomial random variables $(X_{11,i}, X_{21,i})$ with $X_{11,i} \sim \text{Bin}(n_{1,i}, p_{1,i})$ and $X_{21,i} \sim \text{Bin}(n_{2,i}, p_{2,i})$, where first and second rows denote treatment and control groups, respectively. To ensure a low event rate, $p_{2,i}$ in the control arm is generated from a uniform distribution $U(0, 0.01)$. Consequently, the event rate in the treatment arm is determined by $\text{logit}(p_{1,i}) = \log(\text{odds-ratio}) + \text{logit}(p_{2,i})$ for a fixed odds-ratio ranging from 1 to 10.

Figure viii shows the empirical coverage and median width of constructed 95% confidence intervals for the combined odds-ratio. Coverage of Mantel-Haenszel method with artificial corrections to zero events , Peto's method with and without artificial corrections to zero events drops quickly as the true odds-ratio increases. On the other hand, both CD-based approach (combining the proposed CD densities derived from binomial sampling and combining p-value functions discussed in Liu et al. (2014)) and Mantel-Haenszel method without correction to zero events yield confidence intervals with proper coverage. In terms of median width, these three methods behave similarly.

2.5   Discussion

In this article, we have proposed a new CD based approach to draw exact inferences on 2×2 tables. As mentioned in Agresti (2007), an inferential procedure is valid only to the
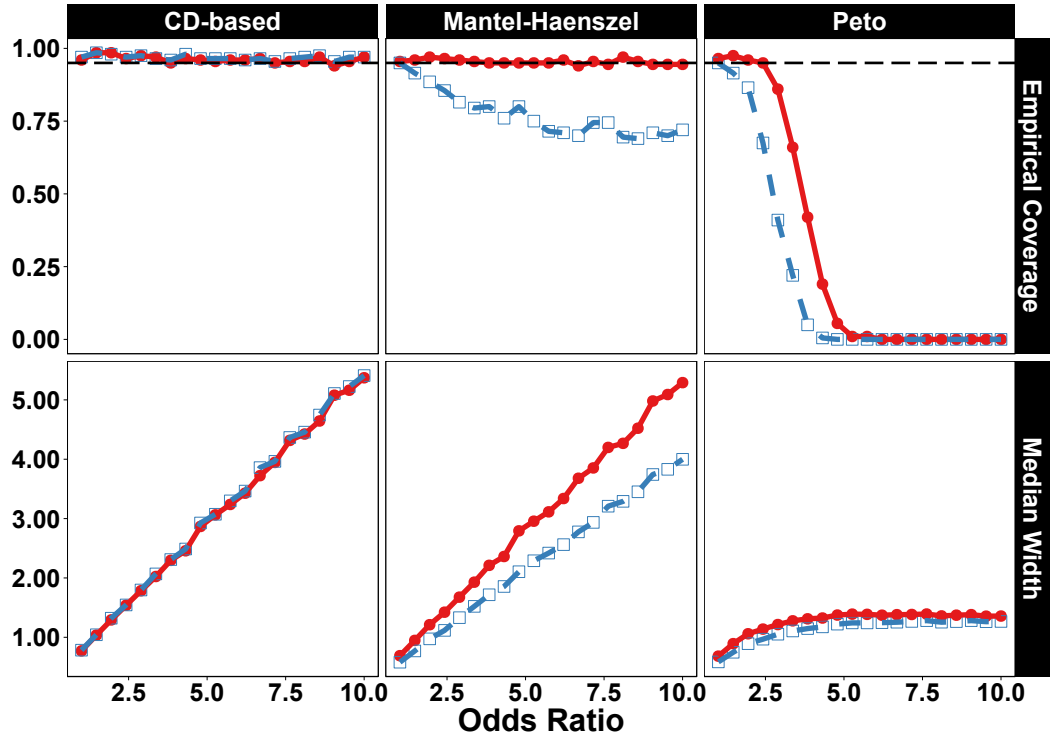
Figure viii: Empirical coverage (top) and median width (bottom) of 95% confidence intervals for odds-ratios ranging between 1 and 10. Results are calculated based on 2000 data sets simulated based on Avandia data. Plots for three different methods are shown in columns. For the "CD-based" column, red solid lines connected with ($\bullet$) and blue dashed lines connected with ($\square$) indicate results from combing proposed CD densities using binomial sampling (melded version) and from combing p-value functions proposed in Liu et al. (2014), respectively. For both "Mantel-Haenszel" and "Peto" columns, red solid lines connected with ($\bullet$) and blue dashed lines connected with ($\square$) indicate the corresponding method without and with the addition of 0.5 to the cells of the 2×2 tables for a zero event, respectively.

extent that sampling assumptions upon which it is based are met. Our simulations agree with this viewpoint as we observe that within each sampling scenario the confidence intervals from the CD approach performs best when the underlying assumption it based off aligned with the one used to generate data. To this point, our approach may be particularly appealing since sampling schemes are directly taken into account when deriving inferential procedures under the CD framework. Furthermore, even the current paper focuses on inferences for odds-ratio and log odds-ratio, the general applicability of the approach to any commonly used measure in a 2×2 table is obvious. By providing such an unified framework under the CD umbrella, our approach offers great flexibilities in terms of sampling schemes, outcome of interest and even inferential procedures.

Performance of the CD approach has been extensively investigated through numerical studies for small sample as well as rare events settings and its large sample properties has already been proved in Hannig (2013). Simulation studies show that our approach is favorable in comparison to other traditional methods like asymptotic based Wald method and Fisher's exact test. Especially, for binomial sampling, the most widely seen sampling method in scientific research, Wald method fails to maintain nominal coverage when event rate is low and Fisher's exact test is too conservative. In contrast, our melded version of CD approach with binomial sampling achieves nominal coverage while maintaining a narrower median width. For 2×2 tables with observed zero event cells, we also performed sensitivity analysis to examine behaviors of the proposed method when prior information is mis-specified. Based on our simulation, we find that the proposed approach is relatively robust to the incorrect prior information applied.

As mentioned, no unique CD exists for parameter of interest for discrete data. We study two approaches to construct $(1 - \alpha)$ confidence intervals from the lower and upper CD densities, one with lower $\alpha/2$ quantile in lower CD function and upper $\alpha/2$ quantile in upper CD function, we call it a raw method, and another one with middle $1 - \alpha$ quantiles from the 50-50 mixture of lower and upper CD densities and we call it a half-corrected melded version. Results from simulation study suggest, within each sampling scheme, the raw version is more conservative, meaning its actual coverage

is usually greater than the nominal one. On average, the melded version can achieve nominal coverage across all values of parameter of interest, and its confidence interval is narrower than the raw method. In addition, across different sampling schemes, we see that the melded confidence interval using binomial sampling performs well consistently in terms of both empirical coverage and median width. Based on these observations, in general, we advise to use the half-corrected melded version of the CD approach when analyzing 2×2 tables, and pick the melded version from binomial sampling when the underlying data generating process is unknown.

We also apply the proposed approach in context of meta-analysis on Avandia data to study the drug safety development where events of interest are rarely seen. Our simulation results are similar to those reported in Liu et al. (2014). The traditional methods like Peto's and Mantel-Haenszel method with and without artificial corrections can yield conflicting conclusions. In addition, exact analyses like combining our CD or p-value functions provide similar results, although we recommend our method since it gives narrower confidence intervals. As for future research, it is possible to study other ways to meld a single CD function from lower and upper CD densities. In addition, the current method may be extended to analyze N-way contingency tables.

Chapter 3

Multivariate Random-effects Meta-analysis of Individual Participant Data from Clinical Trials with Heterogeneous Designs and Partial Information

In this chapter, we propose a novel multivariate random-effects meta-analysis model to analyze individual participant data (IPD) from trials with heterogeneous designs and partial information. The proposed model is fitted through multivariate CD method. We subsequently apply the approach to data from project INTEGRATE.

3.1 Introduction

Meta-analysis is a well-established statistical procedure for quantitatively synthesizing evidence from independent studies (Norman, 1999). In recent years, meta-analysis has increasingly been discussed as an important research method to strengthen statistical inference (Ioannidis, 2005), although meta-analysis reviews can generate divergent and confusing answers (Ioannidis, 2010). The number of publications on meta-analysis has also grown exponentially (Cheung, 2015; Sutton and Higgins, 2008), including those featuring individual participant data (IPD; Riley et al. (2010)). Meta-analysis of IPD, compared to meta-analysis of aggregate data (AD), can expand the scope of possible investigations and produce clinically more meaningful results. When IPD are available, one can ensure that the same model is applied across studies. In addition, it is possible to address more complex research questions with more appropriate and sophisticated models. IPD across all studies can be combined in a one-stage meta-analysis in advanced models that correctly reflect the hierarchical data structure. Alternatively, IPD can

be analyzed sequentially for each study in the first stage, generating multiple related coefficients, which are subsequently combined in the second stage of a two-stage analysis through the use of multivariate meta-analysis models that take into account within-study and between-study correlations (Jackson et al., 2011; Raudenbush and Bryk, 2002). In the past, IPD have been used primarily to derive AD for two-stage analyses Simmonds et al. (2005). More recently, however, IPD have been utilized in one-stage, two-stage, or combined meta-analysis methods (Simmonds et al., 2015).

Meta-analysis of IPD (also called integrative data analysis in the social and behavioral sciences; Hussong et al. (2013); Mun et al. (2015a)) provides unparalleled flexibility for analysts. More can be done to actually check and correct data and harmonize different measures across different studies, compared to what is feasible for meta-analysis of AD. For example, with IPD, a commensurate metric can be established across studies via advanced latent variable modeling approaches using item-level IPD, which is critical for the inclusion of informative psychological covariates and behavioral outcome variables in an analysis (Bauer and Hussong, 2009; Curran et al., 2008, 2014; Huo et al., 2015). At the same time, it is not always possible to obtain commensurate scores (see Mun et al. (2015a)). In extreme data situations, no study may provide chains to link different measures across studies (no between-study overlap; see Siddique et al. (2015)). As research questions get more complex, study-level missing data inevitably go up because answers to these questions require multiple variables to consider and also because many clinical studies differ in key study features, including their designs, populations, measures, or settings (Hofer and Piccinin, 2009; Simmonds and Higgins, 2007). For example, with respect to different measures, a commonly used screening equipment in medical settings may not always be available in other locations. Newer and better screening tools may replace the older ones over time. In psychological research, some studies may use a multi-informant and multi-method measurement approach for the assessment of psychological and behavioral variables, which may not be feasible for other studies with limited resources. Furthermore, longitudinal clinical trials included in a meta-analysis typically differ in their study duration and follow-up schedule (Jones et al., 2009; Trikalinos and Olkin, 2012). It is also not uncommon that binary covariates

may not have any variability either by design (e.g., a study of all women) or naturally in a data set (e.g., Debray et al. (2013)). Any one of these situations can pose significant estimation challenges for an IPD meta-analysis.

In the past, IPD meta-analysis applications have included a subset of studies with targeted covariates or developed a reduced model, either of which essentially deletes partially available data (i.e., list-wise deletion). Therefore, the exclusion of eligible studies or informative covariates in a meta-analysis represents an important loss of information, precision, power, and generalizability, and also diminishes the usefulness of the tested meta-analysis model. Although this discussion may be most salient for IPD, we briefly review and discuss the existing multivariate synthesis approaches to missing data for both AD and IPD because AD can be derived from IPD and also because AD and IPD can be jointly used in synthesis (e.g., Riley et al. (2008); Riley and Steyerberg (2010); Yamaguchi et al. (2014).

In the context of multivariate synthesis methods for AD, Becker and Wu (2007) showed that regression slope coefficients and their covariance estimates can be combined using a multivariate generalized least squares method. However, the slope estimates are assumed to have the same or similar measurement scales and to have come from the same or equivalent regression model in original studies and, consequently, have the same interpretation across different studies. These are strong assumptions for real applications. Furthermore, the covariance estimates associated with the slope coefficients are typically unavailable in published studies. Wu and Becker (2013) circumvented the requirement that all estimates be from the same regression model by combining bivariate correlations instead and, subsequently, analyzing a structural equation model using the pooled correlation matrix. However, this factored likelihood method by Wu and Becker uses standardized z scores to sidestep the issue of ensuring a commensurate metric across studies and requires a monotone missing data pattern to accommodate systematically missing data. Furthermore, both of these approaches (Becker and Wu, 2007; Wu and Becker, 2013) assume a common vector of coefficients across studies (i.e., fixed effects), which may not be reasonable. Recently, Wilson et al. (2016) illustrated the

use of a multilevel random-effects model to combine bivariate correlations from studies that contribute more than one correlation matrix to a pooled correlation matrix, which is then analyzed in the second stage for a meta-regression analysis. Collectively, the existing AD approaches to multivariate synthesis by combining correlations across studies and analyzing them via structural equation modeling techniques can be referred to as meta-analytic structural equation modeling (MASEM; see Cheung and Hafdahl (2016) for an introduction to the special issue in Research Synthesis Methods). The MASEM methods require the missing at random (MAR) assumption. However, if correlations were not reported in original studies because they were statistically insignificant, it would not satisfy the MAR assumption (Cheung and Cheung (2016)). Similarly, with published AD, it would be impossible to check whether assumptions for structural equation modeling are reasonable. At present, it remains an outstanding challenge to overcome between-study heterogeneity, including different measures and sample sizes across studies (see the special issue for more through discussion on MASEM).

In the context of IPD meta-analysis, a bivariate random-effects meta-analysis was proposed to combine fully and partially adjusted parameters (Collaboration et al., 2009), which requires a monotone missing data pattern. In addition, the number of monotone missing data patterns needs to be reduced to use a multivariate meta-analysis model, which limits the utilities of this approach. Other investigators have used imputation methods: a multiple imputation by chained equations (MICE) approach adapted for imputing systematically missing covariates (Resche-Rigon et al., 2013); a multilevel multiple imputation (MLMI) method, an extension of the Resche-Rigon et al. method (Jolani et al., 2015); and a multiple imputation approach (Reiter, 2008) to fill in systematically missing data at the study level, with the help of external calibration trial data (Siddique et al., 2015). These multiple imputation methods may be seen as making up data for an entire study and simultaneously for several studies. Furthermore, the MLMI method may not be feasible under certain data situations (Jolani et al., 2015). The approach by Siddique et al. (2015) appears to be encouraging in the context of their extreme data condition (i.e., no overlap in measures across studies, and only five studies). However, their imputation model does not have any terms to indicate study

membership, which may be difficult to justify. In sum, the available approaches to accommodating missing data for meta-analysis of IPD and AD suggest a need for more research in these relatively early stages of method development.

In the present study, we propose a novel three-stage approach, based on confidence distributions (CDs), to fit and combine IPD for studies with systematically missing data by design (e.g., not asked; Gelman et al. (1998)) or with non-estimable covariates (e.g., no variability) without imputing missing data. The three-stage approach consists of the following three steps: (1) the development of an underlying full (big) model for all studies included in a meta-analysis; (2) the separate analyses of IPD for each study to obtain all within-study estimates while identifying their appropriate connections to the estimates of the full model via mapping matrices; and (3) the estimation of the population-level parameters of the full model in a multivariate random-effects meta-analysis model. In short, we propose to combine data across studies to derive the full model, which we subsequently use to generate specific inferences about point estimates.

Our approach to multi-parameter synthesis (Ades and Sutton (2006)) is similar to the method demonstrated in Gasparrini and Armstrong (2011); Gasparrini et al. (2012), but faces an additional challenge. Gasparrini et al. (2012) had little to no missing data because of the nature of the data examined (complex nonlinear associations between temperature and non-accidental across multiple cities using time-series data). For a multi-parameter synthesis of clinical trial data, however, it is critical to combine partially available information across studies. To address this need, we utilize a new information combination method that combines confidence distributions (called the CD approach; Xie and Singh (2013)). A confidence distribution (CD) is a sample-dependent distribution function that contains information about confidence intervals of a parameter of interest at all levels. It can be referred to as a confidence density if presented in a density function form (Liu et al., 2015). This new method of combining information has been demonstrated as a powerful inference tool in connection with meta-analysis (Claggett et al., 2014; Liu et al., 2015; Xie et al., 2012; Yang et al., 2014, 2016). The current study extends the CD approach to a multivariate random-effects meta-analysis

model, which is based on a more reasonable assumption but computationally more challenging, compared to a multivariate fixed-effects meta-analysis model (Liu et al., 2015). The setup of the current method development, which includes heterogeneous designs and partial information, is also more general than Zeng and Lin (2015) for random-effects models.

3.2 Motivating Data Example

The current study was motivated by Project INTEGRATE (Mun et al., 2015a), which combined IPD from 24 clinical trials aimed at examining the efficacy of brief motivational interventions (BMIs) to reduce excessive alcohol use and prevent harm among college students. De-identified IPD from the 24 trials were obtained through a network of interested collaborators for the purpose of better delineating the mechanisms of behavior change. With the exception of two small studies that had only scale-level IPD, item-level IPD were obtained from all other studies. Typical BMIs are brief, and provide personalized feedback on alcohol use and alcohol-related problems, as well as educational information on alcohol.
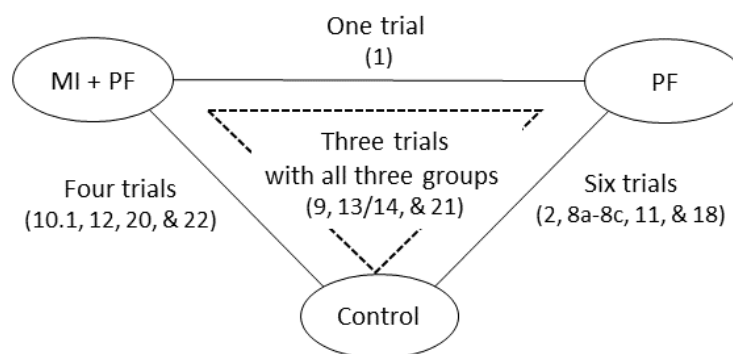
Figure ix: A diagram of the evidence network (numbers in parenthesis indicate studies).

IPD for the current analysis come from 8,920 participants from 14 trials at baseline (see Figure ix). From the original 24 studies, a subset of studies that featured a single intervention (i.e., no comparison group) or interventions that were one of its kind were excluded in the current study, resulting in a stand-alone, personalized feedback

intervention (PF); an in-person motivational intervention with personalized normative feedback profile (MI + PF); and a no-treatment control group. Forty three percent of the participants were men; 76% White; 60% first-year or incoming students who were assessed two or more times up to 12 months post intervention (see vi for different assessment schedules). Note that studies 13 and 14 were originally two independent trials but were combined in the present analysis, given their similarities in key study characteristics and relatively small samples, and the fact that no systematic differences existed across groups at baseline.

Table vi: Baseline and follow-up assessment schedule by study.

Study													

	Time in months (0 = baseline)												
Study	0	1	2	3	4	5	6	7	8	9	10	11	12
1	X				X								X
2	X		X				X						
8a	X												X
8b	X												X
8c	X												X
9	X			X			X						
10.1	X												X
11	X		X		X								
12	X	X		X			X						
13/14	X			X			X			X			
18	X	X					X						
20	X												X
21	X			X			X						X
22	X										X		

Notes: X indicates that baseline or follow-up outcome data exist at the given time.

3.2.1   Intervention and control groups

With the exception of study 1, all studies had a control group (see Figure ix). Eleven of the 14 studies had an assessment-only control group, and the remaining two studies (studies 18 and 20) had a control group who received a single page information sheet. Individuals in the latter control groups were provided with very limited, generally-written information about alcohol use. The information content was not targeted at the

participants drinking behavior, which is a critical characteristic of a BMI. Rather, it was written like a typical awareness campaign (i.e., alcohol has no nutritional value; space your drinks) on college campuses. For Project INTEGRATE, we quantitatively coded all content components of BMIs, as well as their delivery characteristics, and checked their group labels in relation to those in other studies, and appropriately renamed a few groups based on our detailed coding and analysis (Mun and Ray, 2016; Ray et al., 2014).

3.2.2 Measures

We focus on alcohol-related problems (e.g., neglecting responsibilities; friends and relatives avoiding you) as the outcome variable in the current study. Because this outcome variable was assessed differently across original studies (i.e., different items or questionnaires, referent time frames, and response options), we previously utilized a 2-parameter logistic item response theory (2-PL IRT) model to derive commensurate latent trait scores (also called theta [ ] scores) across studies and time. This IRT model was developed and implemented from a Bayesian perspective for Project INTEGRATE (see Huo et al. (2015) for technical details on the multi-unidimensional 2-PL IRT model for IDA applications). Latent trait scores from IRT models can be interpreted with direct reference to item parameters, and are independent of which items that participants were tested on or who else was tested together (Embretson, 2006). IRT models are widely used in educational test settings to estimate latent trait (ability) scores and increasingly for psychological and medical research. In the present study, alcohol-related problems trait (theta) scores at baseline ranged from -2.02 to 3.50 (mean = 0.11, standard deviation = 1).

Latent trait scores for the tendency to adopt protective behavioral strategies prior to, during, and after drinking to protect oneself from experiencing negative consequences from drinking (Martens et al., 2005), such as setting limits or alternating drinks, were derived from a generalized partial credit IRT model (Muraki, 1992) to accommodate polytomous responses for Project INTEGRATE (see Huo et al. (2015) for technical

details; see Mun et al. (2015a,b)). The estimated trait scores (PBS) for protective behavioral strategies were drawn from a standard normal distribution with an expected population mean of 0. In the present study, the trait (theta) scores at baseline ranged from 2.47 to 3.44 (mean = 0.45, standard deviation = 1).

3.2.3 Motivating examples

Example 1: Heterogeneity in the assessment of covariates. The studies included in Project INTEGRATE differed in their assessment of key covariates. For example, first-year college students typically are at risk for excessive drinking and alcohol-related problems. Therefore, whether students are in their first year in college is an important risk variable to take into account. However, several studies included in Project INTE-GRATE recruited exclusively first-year students. In these studies (studies 9, 10, 11, and 22; see Table vii), the coefficient for the binary covariate corresponding to first-year student status (1 = first-year; 0 = other) is not estimable. In addition, the intercept estimates from these four studies would indicate outcome levels of first-year students, whereas the intercepts from other studies would reflect outcome levels of the students in 2nd year and above (i.e., reference demographic group; see the full model in Section 3.3.1, assuming all other covariates are fixed. Therefore, without a methodological intervention, the intercept estimate of the full model across studies would become a mixture of the outcome levels for students in their 2nd year and above for some studies and those of the first-year students for studies 9, 10, 11, and 22, rendering the combined estimate uninterpretable and confounded with other study-level differences. If one were to merely drop this covariate from the full model, the studies consisting of exclusively first-year students would be retained. However, it would be impossible to examine or adjust for the effect of first-year student status on the outcome variable. Furthermore, resulting inferences may suffer from loss of power, and be also biased because of the omission of a well-known informative covariate in the analysis. Therefore, either optionexcluding the studies or covariatescan result in non-negligible loss of information in

a typical meta-analysis. Clearly, there is a need to properly separate the effect of interest from potential confounding effects when combining coefficients from heterogeneous studies.

Table vii: Estimable covariates for the underlying full model by study and by mapping matrix pattern.

Study	N	\multicolumn{13}{c}{Covariate position in the full model}	Mapping matrix pattern
		0	1	2	3	4	5	6	7	8	9	10	11	12	
1+	348	X	X	X	X	X	-	X	X	X	-	X	-	X	1
2	230	X	X	X	X	X	X	-	X	-	X	-	-	-	
8a	1486	X	X	X	X	X	X	-	X	-	X	-	-	-	
8b	2155	X	X	X	X	X	X	-	X	-	X	-	-	-	2
8c	600	X	X	X	X	X	X	-	X	-	X	-	-	-	
9	507	X	X	X	-	X	X	X	X	X	X	X	X	X	3
10.1	435	X	X	X	-	-	-	X	X	-	-	X	-	-	4
11	383	X	X	X	-	-	X	-	X	X	X	-	X	-	5
12	335	X	X	X	X	X	-	X	X	X	-	X	-	X	6
13/14	138	X	X	X	X	-	X	X	X	X	X	X	-	X	7
18	329	X	X	X	X	X	X	-	X	X	X	-	X	-	8
20	928	X	X	X	X	-	-	X	X	-	-	X	-	-	9
21	288	X	X	X	X	X	X	X	X	X	X	X	X	X	10
22	758	X	X	X	-	X	-	X	X	-	-	X	-	-	11

Notes: "X" indicates estimable parameters whereas "-" indicates inestimable parameters. Covariate in the full model are (0) intercept; (1) Man (=1 vs. woman = 0); (2) White (=1 vs. non-white=0); (3) First-year (=1 vs. other=0); (4) PBS (Estimated latent trait (θ) scores for utilizing protective behavioral strategies) at Baseline; (5) PF (stand-alone personalized feedback intervention) (=1 vs. control=0); (6) MI + PF (in-person motivational intervention with personalized normative feedback profile) (=1 vs. control=0); (7) LS (Linear slope of time in month); (8) QS (Quadratic slope of time in month); (9) LS × PF (vs. control); (10) LS × (MI + PF) (vs. control); (11) QS × PF (vs. control); (12) QS × (MI + PF) (vs. control). $^+$ = Study 1 did not have a control group, thus, PF served as a comparison group.

Example 2: Availability of follow-up assessments. In the Project INTEGRATE data set, some studies had a single post-intervention follow-up assessment whereas others had at least two follow-up assessments (see Table vi). Such design differences can cause study-level missing data under certain full models. For example, in a longitudinal model requiring two terms to fit data over time (e.g., linear and quadratic terms) one may be forced either to choose a simpler model that may not correctly reflect change process or to limit the analysis to a subset of studies with a sufficient number of follow-up assessments. As before, either option is not optimal. Both of the motivation examples discussed are not a missing data problem within individual studies. However, when different studies are pooled in a meta-analysis, their heterogeneous designs can create a study-level missing data problem and a significant estimation challenge. We describe

our method in Section 3.3 and illustrate it using real IPD from Project INTEGRATE in Section 3.4.

3.3   Methods

3.3.1   Underlying full model specification

Consider k independent studies each with n_i observations for $i = 1, \cdots, k$. We formulate the underlying full model as a generalized linear mixed model to highlight the flexibility of the proposed approach in terms of modeling choice. Specifically, we assume a random-intercept model with a total of $p - 1$ covariates across k studies:

$$g(E(y_{ijt})) = \beta_{0i} + \beta_{1i}x_{1ijt} + \beta_{2i}x_{2ijt} + \beta_{3i}x_{3ijt} + \cdots + \beta_{p-1,i}x_{p-1,ijt} + u_{ij0},$$

where $g(\cdot)$ is the link function and $E(\cdot)$ denotes expectation. y_{ijt} indicates the outcome for participant j in study i at time t, and x_{dijt} is value of the dth covariate for participant j in study i and time t. β_{di} indicates the coefficient associated with the dth covariate for study i with $d = 0, \cdots, p - 1$. The term u_{ij0} indicates participant-specific random intercept effects. The link function can be an identity link for a linear model with a continuous normal outcome; a logit link for a logit model with a binary outcome variable (see Yang et al. (2014) for a binary data example); and a log link for a loglinear model with a count variable. In the present study, we use an identity link function for a normally distributed outcome. For clarity, the full model in the current study is:

$$
\begin{aligned}
y_{ijt} = {} & \beta_{0i} + \beta_{1i}\text{male}_{ij} + \beta_{2i}\text{white}_{ij} + \beta_{3i}\text{freshman}_{ij} + \beta_{4i}\text{m0pbs}_{ij} + \beta_{5i}\text{PF}_{ij} \\
& + \beta_{6i}(\text{MI} + \text{PF})_{ij} + \beta_{7i}\text{month}_{ij} + \beta_{8i}\text{month}^2_{ij} + \beta_{9i}(\text{PF}\times\text{month})_{ij} \\
& + \beta_{10i}((\text{MI} + \text{PF})\times\text{month})_{ij} + \beta_{11i}(\text{PF}\times\text{month}^2)_{ij} + \beta_{12i}((\text{MI} + \text{PF})\times\text{month}^2)_{ij} \\
& + u_{ij0} + \varepsilon_{ijt},
\end{aligned}
$$

where male, white, freshman, PF and (MI+PF) are binary indicator variables; m0pbs and month are continuous variables; and the rest of the terms are either interaction terms between the aforementioned ones and/or a quadratic form. The last term ε_{ijt}

represents a residual error term for participant j in study i at time t, and u_{ij0} and ε_{ijt} are independent.

3.3.2 Mapping matrix approach to partial information

At the second stage of the analysis, separate analyses are conducted sequentially for each and every study. Upon obtaining regression parameter estimates from all studies and their covariance matrices, they can be connected to the full model via mapping matrices. For a special case where all original studies have the $p - 1$ covariates, the estimates for the assumed full model shown in Section 3.3.1 can be combined directly using the standard multivariate random-effects meta-analysis model discussed by, for example, Jackson et al. (2010, 2011) and others. However, for more typical situations where some trials provide partial information, one can obtain the p_i length parameter vector $\boldsymbol{\beta}_i$ for study i with $p_i \leq p$. Our task is to appropriately link the estimates from studies with partial information to the estimates from other studies, and to incorporate all available evidence for efficient and valid inference for the entire parameter vector of interest $\boldsymbol{\beta}$. Let $\boldsymbol{M}_i$ be the mapping function that links $\boldsymbol{\beta}_i$ to $\boldsymbol{\beta}$, i.e., $\boldsymbol{\beta}_i \equiv \boldsymbol{M}_i(\boldsymbol{\beta})$. In a linear model, in which we often have linear mapping, the relationship $\boldsymbol{\beta}_i \equiv \boldsymbol{M}_i(\boldsymbol{\beta})$ can typically be simplified to the following linear equation: $\boldsymbol{\beta}_i \equiv \boldsymbol{M}_i\boldsymbol{\beta}$, where $\boldsymbol{M}_i$ is a $p_i \times p$ matrix.

As a simple illustration of this method, let us assume that we fit a response variable y with two continuous covariates and x_1 and x_2 in a fixed-effect full model:

$$E(y_{ij}) = \beta_{0i} + \beta_{1i}x_{1ij} + \beta_{2i}x_{2ij}$$

where subscripts i and j index studies and participants, respectively. If study i has all the covariates available, then

$$\boldsymbol{M}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let us consider a situation where β_{2i} cold not be estimated for study i because x_{2i} was

not assessed. If we can assume that x_2 has an expected zero mean, then

$$\boldsymbol{M}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Consequently, the resulting vector of estimable parameters for study i would be reduced to $\boldsymbol{\beta}_i = \boldsymbol{M}_i\boldsymbol{\beta} = (\beta_0, \beta_1)^T$.

If x_2 is a binary variable with a constant value $x_{2i} = 1$ for study i, then β_{2i} could not be estimated because there is no variability. In this data situation, an appropriate mapping matrix is

$$\boldsymbol{M}_i = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

with $\boldsymbol{\beta}_i = \boldsymbol{M}_i\boldsymbol{\beta} = (\beta_0 + \beta_2, \beta_1)^T$, where an estimable intercept term in study i in the context of the full model is a sum of β_0 and β_2. The identification of two different mapping matrices for the seemingly similar situations shows that it is necessary to take into account between-study differences in designs and measures from the perspective of estimating all parameters of the full model. For more technical details and nonlinear mapping functions, see Liu et al. (2015) and Yang et al. (2014).

Let us consider a few specific data examples in our project. Study 1 (White et al., 2007) from Project INTEGRATE (Mun et al., 2015a) tested the efficacy of two BMIs (i.e., MI + PF and PF) without a no-treatment control group, whereas typical trials included in Project INTEGRATE were two-arm trials with a single intervention arm and a control arm (see Figure ix). In the context of a network meta-analysis (Cipriani et al., 2013; Jansen et al., 2011), the relative intervention benefit between two BMIs (MI + PF and PF) can be seen as the difference of the differences between MI + PF and control and between PF and control. In other words, MI + PF vs. PF = (MI + PF vs. control) (PF vs. control). Therefore, the study-specific (within-study) parameter of the relative intervention effect from study 1 can provide valuable information in the context of a network analysis as long as it can properly be linked to other parameters of the intervention effects from other studies. Clearly, we can estimate study-specific parameters $\boldsymbol{\beta}_i$ for study 1 that correspond to the true effects of demographic and other

model covariates, which include sex (1 = man; 0 = woman), race (1 = White; 0 = nonwhite), first-year status (1 = first-year student; 0 = other), protective behavioral strategies (PBS) at baseline, the effect of the MI + PF intervention in comparison with PF, the effects of linear and quadratic slopes, and the interaction effects between growth slopes and intervention groups using IPD, but not the population parameter vector, $\boldsymbol{\beta}$, which additionally includes three parameters involving PF in comparison to control. See Section 3.4.1 for justification of the full model for our motivation example.

To include data from study 1 in our meta-analysis and draw valid inferences from the combined estimates, we need to identify an appropriate mapping matrix to link $\boldsymbol{\beta}_i$ to $\boldsymbol{\beta}$. This mapping matrix can be identified as follows (see also Table 3). First, remove the 6th, 10th, and 12th rows from a 13×13 identity matrix by changing the diagonal 1s into 0s. These three rows correspond to the parameters comparing PF with control (i.e., PF vs. control; [PF vs. control] linear growth slope; and [PF vs. control] quadratic growth slope, respectively). Second, change the first row to indicate that the intercept estimate from study 1 describes the study-specific average outcome response of PF, rather than control, in the full model. Third, contrast MI + PF against PF when applicable. The resulting vector of parameters for study 1 is

$$\boldsymbol{\beta}_1 = (\beta_0 + \beta_5, \beta_1, \beta_2, \beta_3, \beta_4, -\beta_5 + \beta_6, \beta_7, \beta_8, -\beta_9 + \beta_{10}, -\beta_{11} + \beta_{12})^T.$$

Table viii: Mapping matrix pattern. The pattern indices here correspond to the numbers shown in the last column in Table vii. Bold and underlined elements indicate how the identity matrix for the full parameter vector was modified to link partial data between studies.

Covariate	0	1	2	3	4	5	6	7	8	9	10	11	12
Pattern 1 (Study 1)													
0. Intercept	**1**	0	0	0	0	**1**	0	0	0	0	0	0	0
1. Man (=1 vs. woman=0)	0	**1**	0	0	0	0	0	0	0	0	0	0	0
2. White (=1 vs. nonwhite=0)	0	0	**1**	0	0	0	0	0	0	0	0	0	0
3. First-year (=1 vs. other=0)	0	0	0	**1**	0	0	0	0	0	0	0	0	0

Table viii Continued from previous page

Covariate	0	1	2	3	4	5	6	7	8	9	10	11	12
4. PBS at Baseline	0	0	0	0	**1**	0	0	0	0	0	0	0	0
5. PF (=1 vs. control=0)	0	0	0	0	0	0	0	0	0	0	0	0	0
6. MI + PF (=1 vs. control=0)	0	0	0	0	0	**-1**	**1**	0	0	0	0	0	0
7. LS	0	0	0	0	0	0	0	**1**	0	0	0	0	0
8. QS	0	0	0	0	0	0	0	0	**1**	0	0	0	0
9. LS * PF (vs. control)	0	0	0	0	0	0	0	0	0	0	0	0	0
10. LS * MI + PF (vs. control)	0	0	0	0	0	0	0	0	0	**-1**	**1**	0	0
11. QS * PF (vs. control)	0	0	0	0	0	0	0	0	0	0	0	0	0
12. QS * MI + PF (vs. control)	0	0	0	0	0	0	0	0	0	0	0	**-1**	**1**
Pattern 3 (Study 9)													
0. Intercept	**1**	0	0	**1**	0	0	0	0	0	0	0	0	0
1. Man (=1 vs. woman=0)	0	**1**	0	0	0	0	0	0	0	0	0	0	0
2. White (=1 vs. nonwhite=0)	0	0	**1**	0	0	0	0	0	0	0	0	0	0
3. First-year (=1 vs. other=0)	0	0	0	0	0	0	0	0	0	0	0	0	0
4. PBS at Baseline	0	0	0	0	**1**	0	0	0	0	0	0	0	0
5. PF (=1 vs. control=0)	0	0	0	0	0	**1**	0	0	0	0	0	0	0
6. MI + PF (=1 vs. control=0)	0	0	0	0	0	0	**1**	0	0	0	0	0	0
7. LS	0	0	0	0	0	0	0	**1**	0	0	0	0	0
8. QS	0	0	0	0	0	0	0	0	**1**	0	0	0	0
9. LS * PF (vs. control)	0	0	0	0	0	0	0	0	0	**1**	0	0	0
10. LS * MI + PF (vs. control)	0	0	0	0	0	0	0	0	0	0	**1**	0	0
11. QS * PF (vs. control)	0	0	0	0	0	0	0	0	0	0	0	**1**	0
12. QS * MI + PF (vs. control)	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Pattern 4 (Study 11)													
0. Intercept	**1**	0	0	**1**	0	0	0	0	0	0	0	0	0
1. Man (=1 vs. woman=0)	0	**1**	0	0	0	0	0	0	0	0	0	0	0
2. White (=1 vs. nonwhite=0)	0	0	**1**	0	0	0	0	0	0	0	0	0	0
3. First-year (=1 vs. other=0)	0	0	0	0	0	0	0	0	0	0	0	0	0

Table viii Continued from previous page

Covariate	0	1	2	3	4	5	6	7	8	9	10	11	12
4. PBS at Baseline	0	0	0	0	0	0	0	0	0	0	0	0	0
5. PF (=1 vs. control=0)	0	0	0	0	0	**1**	0	0	0	0	0	0	0
6. MI + PF (=1 vs. control=0)	0	0	0	0	0	0	0	0	0	0	0	0	0
7. LS	0	0	0	0	0	0	0	**1**	0	0	0	0	0
8. QS	0	0	0	0	0	0	0	0	**1**	0	0	0	0
9. LS * PF (vs. control)	0	0	0	0	0	0	0	0	0	**1**	0	0	0
10. LS * MI + PF (vs. control)	0	0	0	0	0	0	0	0	0	0	0	0	0
11. QS * PF (vs. control)	0	0	0	0	0	0	0	0	0	0	0	**1**	0
12. QS * MI + PF (vs. control)	0	0	0	0	0	0	0	0	0	0	0	0	0

Note that all study-specific parameters $\boldsymbol{\beta}_i$ are within-study estimates. When study-specific, within-study estimates are combined to obtain the population-level parameter vector of the full model (see Section 3.3.3), the intercept term from study 1 correctly contributes to the estimation of two different parameters β_0 and β_5 of the full parameter vector. Similarly, the study-specific parameters involved in the contrasts between MI + PF and PF contribute to the estimation of the β_5, β_6, β_9, β_{10}, β_{11} and β_{12} of the full model.

To provide another example, study 11 exclusively recruited first-year students, did not assess protective behavioral strategies at baseline, and tested the efficacy of PF against a control in a two-arm trial. Therefore, to map estimable parameters from study 11 into the full parameter vector, the rows of an identify matrix corresponding to these variables were removed by changing the diagonal 1s into 0s. In addition, the first row of the mapping matrix was modified to indicate that the intercept parameter from study 11 is essentially the average outcome of first-year students in the context of the full model. The row for PBS could be removed because the estimated PBS trait scores were drawn from a standard normal distribution with an expected population mean of

0 in a previous IRT analysis. The resulting, reduced vector of parameters for study 11 is

$$\boldsymbol{\beta}_i = (\beta_0 + \beta_3, \beta_1, \beta_2, \beta_5, \beta_7, \beta_8, \beta_9, \beta_{11})^T .$$

3.3.3 Estimation of a multivariate random-effects meta-analysis model using the CD method

Once we obtain all study-specific parameters and identify their connections to the full model, we now obtain estimates of the full-length, population-level parameter vector $\boldsymbol{\beta} \equiv (\beta_0, \beta_1, \beta_2, \cdots, \beta_{p-1})^T$ by assuming a multivariate random-effects meta-analysis model, which is an extension of its univariate counterpart (Normand, 1999). Specifically, following the notations from Section 3.3.1, we denote $\boldsymbol{b}_i$ as the study-specific estimate of the corresponding population-level sub-vector $\boldsymbol{\beta}_i$ for study i. Accordingly, we have

$$\text{Level 1} : \boldsymbol{b}_i|\boldsymbol{\beta}_i, \boldsymbol{S}_i \sim \text{MVN}_{p_i}\left(\boldsymbol{\beta_i}, \boldsymbol{S}_i\right)$$

$$\text{Level 2} : \boldsymbol{\beta}_i|\boldsymbol{\beta}, \boldsymbol{\Sigma} \sim \text{MVN}_{p_i}\left(\boldsymbol{M}_i\boldsymbol{\beta}, \boldsymbol{M}_i\boldsymbol{\Sigma}\boldsymbol{M}_i^T\right),$$

for study $i = 1, 2, \cdots, k$ and covariate $d = 1, 2, \cdots, p-1$. MVN_{p_i} stands for the multivariate normal distribution with dimension p_i; $\boldsymbol{S}_i$ is the observed covariance matrix for study i; $\boldsymbol{M}_i$ denotes the mapping matrix for study i; and $\boldsymbol{\Sigma}$ is the unknown between-study covariance matrix that needs to be estimated.

To estimate $\boldsymbol{\beta}$, first, one needs to estimate the between-study covariance matrix $\boldsymbol{\Sigma}$. In our implementation, we use the restricted maximum likelihood (REML) method while using the estimates from the method of moments (Chen et al. (2012)) as starting values to achieve faster convergence. Estimation of $\boldsymbol{\Sigma}$ follows the formula given by Jennrich and Schluchter (1986):

$$\hat{\boldsymbol{\Sigma}}_{\text{REML}} = \arg\max_{\boldsymbol{\Sigma}} \left\{ -\frac{1}{2}\sum_{i=1}^k \log|\boldsymbol{S}_i + \boldsymbol{M}_i\boldsymbol{\Sigma}\boldsymbol{M}_i^T| - \frac{1}{2}\log\left|\sum_{i=1}^k \boldsymbol{M}_i^+ \left(\boldsymbol{S}_i + \boldsymbol{M}_i\boldsymbol{\Sigma}\boldsymbol{M}_i^T\right)^{-1} \boldsymbol{M}_i\right| \right.$$
$$\left. -\frac{1}{2}\sum_{i=1}^k \left(\boldsymbol{b}_i - \boldsymbol{M}_i\hat{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{S}_i + \boldsymbol{M}_i\boldsymbol{\Sigma}\boldsymbol{M}_i^T\right)^{-1} \left(\boldsymbol{b}_i - \boldsymbol{M}_i\hat{\boldsymbol{\beta}}\right) \right\},$$

Where

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{k} \boldsymbol{M}_i^+ \left(\boldsymbol{S}_i + \boldsymbol{M}_i \boldsymbol{\Sigma} \boldsymbol{M}_i^T \right)^{-1} \right)^{-1} \left(\sum_{i=1}^{k} \boldsymbol{M}_i^+ \left(\boldsymbol{S}_i + \boldsymbol{M}_i \boldsymbol{\Sigma} \boldsymbol{M}_i^T \right)^{-1} \boldsymbol{M}_i \boldsymbol{M}_i^+ \boldsymbol{b}_i \right).$$

Once $\hat{\boldsymbol{\Sigma}}_{\mathrm{REML}}$ is estimated, $\boldsymbol{\beta}$ can be estimated from a combined multivariate normal CD as follows. First, to accommodate the multidimensional nature of $\boldsymbol{\beta}$ (see Xie and Singh (2013) for the CD approach for univariate applications), we construct a multivariate normal CD function for $\boldsymbol{\beta}$ (Singh et al., 2007). By definition, $H(\cdot)$ is a multivariate normal CD function for a $p \times 1$ vector $\boldsymbol{\beta}$ if and only if the projected distribution of $H_\lambda(\cdot)$ on a $p \times 1$ vector λ, for any given $\lambda \in \mathfrak{R}^p$, is an univariate normal CD for $\lambda^T \boldsymbol{\beta}$. At the study level, $H_i(\boldsymbol{\beta}_i)$ is a corresponding multivariate CD function for study i, where $\boldsymbol{\beta}_i = \boldsymbol{M}_i \boldsymbol{\beta}$. On the conditions that $\boldsymbol{M}_i$ is positive or semi-positive definite and that all parameters can be linked across studies, the combined multivariate CD function for the population-level parameter vector $\boldsymbol{\beta}$ has been shown as

$$H^{(c)}(\boldsymbol{\beta}) = \Phi_p \left(\boldsymbol{\Sigma}^{-1/2} \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(c)} \right) \right),$$

where $H^{(c)}(\cdot)$ is the combined CD and $\Phi_p(\cdot)$ is the cumulative distribution function for the standard multivariate normal distribution with p dimensions (see Yang et al. (2014) for a formal definition). $\boldsymbol{\beta}$ can then be directly estimated from the combined multivariate normal CD, $H^{(c)}(\boldsymbol{\beta})$, by using the following formulas,

$$\hat{\boldsymbol{\beta}}^{(c)} = \left(\sum_{i=1}^{k} \boldsymbol{W}_i \right)^{-1} \left(\sum_{i=1}^{k} \boldsymbol{W}_i \boldsymbol{M}_i^+ \boldsymbol{b}_i \right) \quad \text{for the estimated mean vector and ,}$$

$$\hat{\boldsymbol{\Sigma}}^{(c)} = \left(\sum_{i=1}^{k} \boldsymbol{W}_i \right)^{-1} \quad \text{for its covariance matrix,}$$

where $\boldsymbol{W}_i$ is defined as $\boldsymbol{W}_i = \boldsymbol{M}_i^+ \left(\boldsymbol{S}_i + \boldsymbol{M}_i \hat{\boldsymbol{\Sigma}}_{\mathrm{REML}} \boldsymbol{M}_i^T \right)^{-1} \boldsymbol{M}_i$, and $\boldsymbol{M}_i^+$ is the Moore-Penrose generalized inverse of M_i. Therefore, $\hat{\boldsymbol{\beta}}^{(c)}$ is the vector of CD point estimator for $\boldsymbol{\beta}$ with the CD covariance matrix $\boldsymbol{\Sigma}^{(c)}$. Note that we use the sample covariance estimators $\boldsymbol{S}_i$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{REML}}$ because the combined CD function $H^{(c)}(\boldsymbol{\beta})$ would be an asymptotic multivariate normal CD as long as these estimators are consistent (see Yang et al. (2014)). The CD-based approach yields estimates with several desirable properties (e.g., asymptotically as efficient as the MLE; robust against model

misspecification). More detailed technical descriptions and proofs are provided in Liu et al. (2015) and Yang et al. (2014).

3.3.4 Inference

Upon obtaining all estimates of the full model across studies, inferences can be made flexibly using the combined full model. For example, to interpret time-related intervention effects, one can estimate and compare outcomes at a given time across intervention groups. To estimate outcomes at specific values of the covariates, the estimated parameters ($\hat{\boldsymbol{\beta}}$) in the present study) can be used to construct the estimated full model. We can then use the full model to obtain model-based mean $\hat{y}_0$ and its variance by plugging in a set of in-sample covariate values x_0 using the following formula

$$\hat{y}_0 = \boldsymbol{x}_0^T \hat{\boldsymbol{\beta}} \text{ and } \text{var}(\hat{y}_0) = \boldsymbol{x}_0^T \text{Cov}(\boldsymbol{\beta}) \boldsymbol{x}_0,$$

We will illustrate this using real data in the next section.

3.4 Data Example

3.4.1 Underlying full model specification

Alcohol use trajectories among college students after interventions typically show a sharp decline, followed by a rebound over time (Huh et al., 2015; White et al., 2007). Therefore, we chose a quadratic growth model and tested it using IPD from several individual studies separately, which supported the use of the model. To test intervention effects over time, we included interaction terms between time and intervention groups. We included gender, first-year student status, and race (white or otherwise) as demographic covariates. In addition, we conducted a separate analysis within individual studies to see if attrition at follow-ups can be explained by participant-level covariates. Based on this attrition analysis, we discovered that the tendency to use protective behavioral strategies prior to and while drinking, such as setting drinking limits, was related to greater chances for participants to drop out at follow-ups in some

of the studies. We subsequently added this covariate to the full model. All analyses were performed using R (version 3.2.3). The nlme R package (Pinheiro et al., 2014) was used to fit a random-intercept growth model. We developed R codes to identify patterns of estimable covariates and to construct mapping matrices and used the optimx package (Nash et al., 2011) to obtain the REML estimates of the between-study covariance matrix.

3.4.2   Partial information and mapping matrices

Table 1 shows all 13 coefficients included in the current analysis and their availability by study. Coefficients could not be estimated because (1) variables were unassessed by study (e.g., PBS; studies 10.1, 11, 13/14, and 20); (2) the entire sample consisted of only first-year students (studies 9, 10.1, 11, and 22); (3) not all intervention groups were included (studies 1, 2, 8a, 8b, 8c, 10.1, 11, 12, 18, 20, and 22); and (4) only one follow-up assessment was available (i.e., only a linear slope term could be estimated; studies 2, 8a, 8b, 8c, 10.1, 20, and 22). A total of just three covariate coefficients were estimable across all studies (i.e., man vs. woman, white vs. non-white, and a linear slope of time), and only one study (study 21) had the necessary data to estimate all coefficients. Thus, it is clear that, without a methodological solution to combine data from studies providing partial information, the underlying full model could not be estimated. There were a total of 11 different partial data patterns across 14 studies, requiring 11 different mapping matrices. With the exception of study 21, all other studies required mapping matrices with reduced dimensions.

3.4.3   Estimation and interpretation

The combined estimate $\hat{\boldsymbol{\beta}}$ (see Table ix) and its covariance matrix $\text{Cov}(\hat{\boldsymbol{\beta}})$ were obtained by applying the estimation method described in Section 3.3.3. Table x shows that all correlation estimates of the coefficients were not boundary estimates (i.e., away from ± 1), which can be sometimes observed in multivariate meta-analysis models, suggesting estimation difficulties (Riley et al., 2007). Substantively, results indicated that there

was a significant interaction between MI + PF and the linear slope of time. To interpret this interaction effect, we calculated model-implied outcome values for all groups based on the estimated full model. Namely, we used in-sample covariate values (i.e., first-year, male, white students, a mean PBS score) and obtained model-implied means for alcohol-related problems and their estimated variances at 6 months and 12 months post intervention. Figure x shows the expected mean levels for all three groups, which showed a reduction in alcohol-related problems at 6 months, followed by a rebound at 12 months. When we further probed this by comparing PF and MI + PF with control at three time points (Figure xi), a statistically significant group difference was found for MI + PF (vs. control) at 12 months (p-value = 0.023).

Table ix: Combined parameter estimates from the multivariate random-effects meta-analysis.

Covariate	Estimate	p value
0. Intercept	0.4276	0.0046
1. Man (=1 vs. woman=0)	0.0191	0.6857
2. White (=1 vs. nonwhite=0)	0.0511	0.1308
3. First-year (=1 vs. other=0)	0.0530	0.0541
4. PBS at Baseline	-0.2796	0.0000
5. PF (=1 vs. control=0)	-0.0045	0.8781
6. MI + PF (=1 vs. control=0)	0.0946	0.1180
7. LS	-0.0469	0.0082
8. QS	0.0046	0.0091
9. LS * PF	0.0035	0.3514
10. LS * MI + PF	-0.0293	0.0213
11. QS * PF	-0.0007	0.3880
12. QS * MI + PF	-0.0003	0.8026

Notes: PBS = Estimated latent trait (θ) scores for utilizing protective behavioral strategies; PF = stand-alone personalized feedback intervention; MI + PF = in-person motivational intervention with personalized normative feedback profile; LS = Linear slope (time in month); and QS = Quadratic slope (months squared).

3.4.4 Sensitivity analysis

To examine if the reported results were overly influenced by outlying studies, we further conducted a sensitivity analysis by excluding one study at a time and repeating the
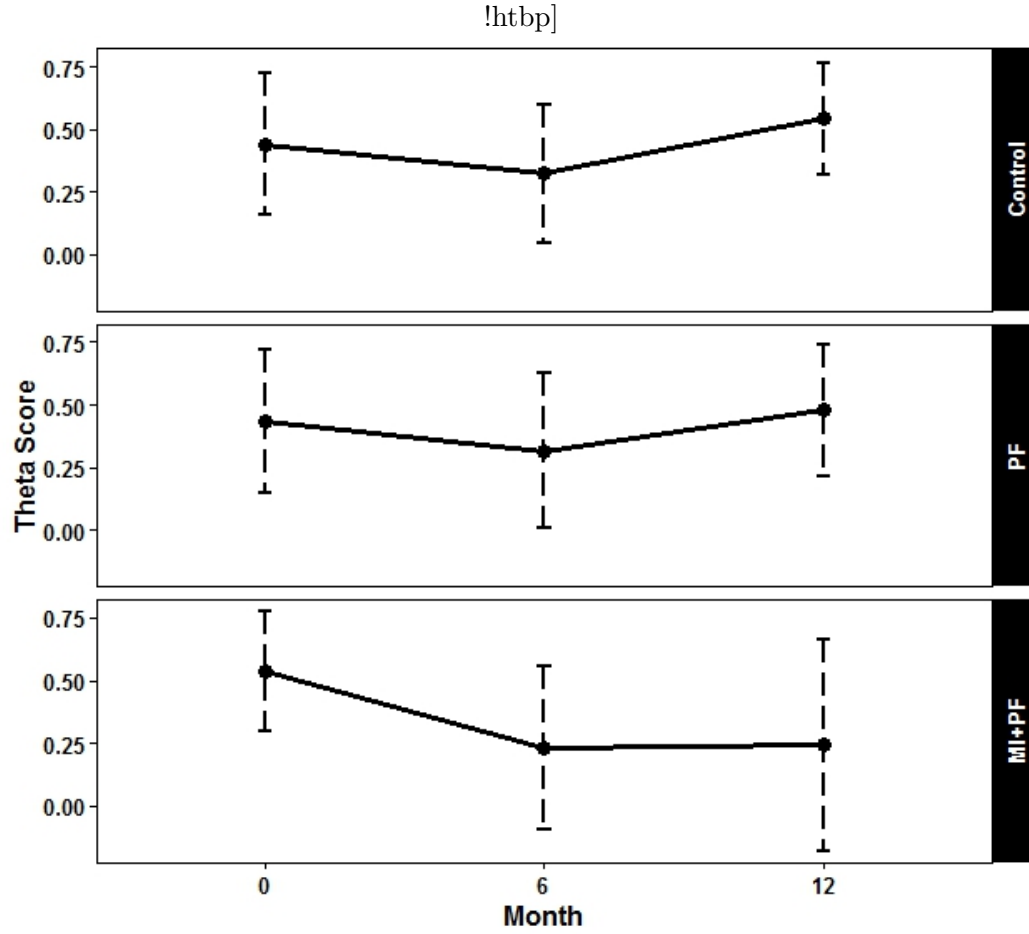
Figure x: Model-based mean estimates for three different groups using the estimated full model shown in Table ix. Theta score = latent trait severity score for alcohol-related problems. PF = stand-alone personalized feedback intervention; MI + PF = in-person motivational intervention with personalized normative feedback profile. Values for covariates were set for White, first-year, male students with a mean PBS score at baseline. Vertical dotted lines indicate 95% confidence intervals.

Table x: Synthesized between-study correlation matrix of the combined parameter estimates.

	0	1	2	3	4	5	6	7	8	9	10	11	12
0. Intercept	1	0.31	-0.68	-0.07	-0.4	-0.03	-0.61	0.22	-0.61	0.63	0.53	0.56	0.21
1. Man (=1 vs. woman=0)		1	-0.52	-0.66	0.01	-0.09	-0.39	0.4	-0.34	0.66	0.55	-0.21	-0.02
2. White (=1 vs. nonwhite=0)			1	0.12	0.36	0.24	0.64	-0.56	0.75	-0.47	-0.77	-0.52	0.34
3. First-year (=1 vs. other=0)				1	0.18	0.23	0.18	-0.68	0.44	-0.3	-0.64	0.2	0.34
4. PBS at Baseline					1	-0.49	-0.27	-0.62	0.81	0.36	-0.36	-0.88	0.49
5. PF (=1 vs. control=0)						1	0.77	-0.22	0.01	-0.45	-0.52	0.34	0.22
6. MI + PF (=1 vs. control=0)							1	-0.2	0.28	-0.84	-0.67	0.04	-0.06
7. LS								1	-0.87	-0.02	0.86	0.43	-0.84
8. QS									1	-0.1	-0.83	-0.75	0.61
9. LS * PF										1	0.46	-0.27	0.44
10. LS * MI + PF											1	0.31	-0.56
11. QS * PF												1	-0.38
12. QS * MI + PF													1

Notes: PBS = Estimated latent trait (θ) scores for utilizing protective behavioral strategies; PF = stand-alone personalized feedback intervention; MI + PF = in-person motivational intervention with personalized normative feedback profile; LS = Linear slope (time in month); and QS = Quadratic slope (months squared).
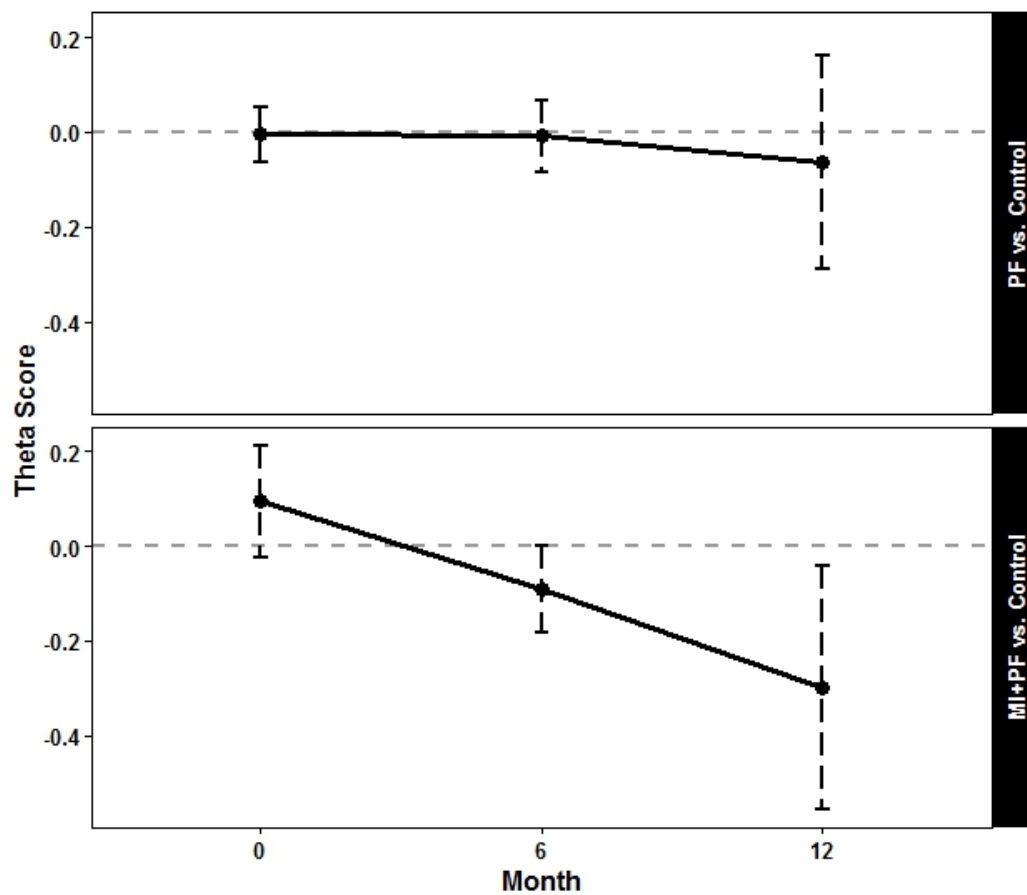
Figure xi: Model-based mean difference estimates of the two intervention groups, compared to control, in alcohol-related problems. Theta score = latent trait severity score for alcohol-related problems. PF = stand-alone personalized feedback intervention; MI + PF = in-person motivational intervention with personalized normative feedback profile. Vertical dotted lines indicate 95% confidence intervals. A horizontal dashed line at zero indicates no group difference.

analysis. Results indicated that while individual regression parameters changed in magnitude to some extent, the overall findings remained largely the same (results not reported in the current study but available upon request). The most influential study was study 9. Without the contribution of study 9, the average outcomes of the MI + PF group across time were similar to those of other groups, showing a more pronounced rebound at 12 months. This was perhaps due to the fact that study 9 was one of the few studies that had almost all estimable parameters and, consequently, had a greater influence on the final estimates, which may not be a surprise given that studies with more information are weighed more in estimation. In addition, we sequentially removed two different covariates from the analysis at two different stages of the analysis and examined their impact on two key coefficients (PF x Month and [PF + MI] x Month). Figure xii shows the results when we removed PBS (top) and first-year student status (bottom) in the final stage of the analysis, whereas Figure xiii shows the results when we removed them throughout the entire analysis. Both of the sensitivity results suggest that the estimated full model (shown in filled diamond symbols in bottom) was fairly robust and that an omitted covariate made little impact on the final estimates, regardless of whether it was a continuous or binary covariate. We also inspected other coefficients in the full model. The reported robust findings from the sensitivity analyses were also observed for other coefficients.

3.5   Discussion

The current study adopted a new estimation approach to IPD for multivariate random-effects meta-analysis applications, which incorporates partially available information by utilizing the CD concept (Xie and Singh, 2013; Xie et al., 2012). The CD concept has been studied in connection with meta-analysis in recent literature (Claggett et al., 2014; Liu et al., 2015; Yang et al., 2014, 2016). We extended the CD-based approach to a multivariate random-effects meta-analysis model in the current study from the multi-parameter synthesis perspective (Ades and Sutton (2006)). The three-stage CD-based approach differs from the existing complex synthesis approaches in the sense that the

Study excluded	Coefficient estimate [95% CI]		Coefficient estimate [95% CI]
PBS removed from			
study 1	0.003 [-0.004 , 0.011]		-0.029 [-0.053 , -0.005]
study 2	0.004 [-0.003 , 0.011]		-0.029 [-0.054 , -0.004]
study 8a	0.002 [-0.005 , 0.008]		-0.026 [-0.051 , -0.001]
study 8b	0.003 [-0.004 , 0.010]		-0.028 [-0.053 , -0.004]
study 8c	0.003 [-0.004 , 0.010]		-0.027 [-0.051 , -0.002]
study 9	0.004 [-0.004 , 0.011]		-0.029 [-0.054 , -0.005]
study 12	0.004 [-0.003 , 0.010]		-0.024 [-0.045 , -0.003]
study 18	0.003 [-0.005 , 0.011]		-0.028 [-0.051 , -0.005]
study 21	0.004 [-0.004 , 0.011]		-0.029 [-0.053 , -0.004]
study 22	0.003 [-0.004 , 0.011]		-0.025 [-0.050 , -0.001]
1st-year removed from			
study 1	0.003 [-0.005 , 0.011]		-0.029 [-0.053 , -0.005]
study 2	0.003 [-0.004 , 0.011]		-0.029 [-0.053 , -0.005]
study 8a	0.003 [-0.004 , 0.011]		-0.030 [-0.055 , -0.004]
study 8b	0.003 [-0.004 , 0.011]		-0.027 [-0.050 , -0.004]
study 8c	0.003 [-0.004 , 0.011]		-0.030 [-0.056 , -0.005]
study 12	0.003 [-0.004 , 0.011]		-0.029 [-0.053 , -0.004]
study 13/14	0.004 [-0.004 , 0.011]		-0.030 [-0.055 , -0.004]
study 18	0.003 [-0.004 , 0.011]		-0.030 [-0.054 , -0.005]
study 20	0.004 [-0.004 , 0.011]		-0.030 [-0.055 , -0.004]
study 21	0.003 [-0.004 , 0.011]		-0.029 [-0.053 , -0.005]
Full model	**0.004 [-0.004 , 0.011]**		**-0.029 [-0.054 , -0.004]**

-0.080   0.000   0.020   -0.080   0.000   0.020

Coefficient estimate for PFxMonth   **Coefficient estimate for (PF+MI)xMonth**

Figure xii: Results from sensitivity analyses where each covariate from each study was treated as systematically missing in the final-stage of the analysis. The effects of the exclusion of a continuous covariate (PBS at baseline; top) and a binary covariate (first-year student status; bottom) on the combined estimates of PF x Month (left) and (MI+PF) x Month (right) are shown, respectively. Filled diamond symbols indicate the combined estimates from all 14 studies as reported in Table ix. The estimates from sensitivity analyses are shown in filled squares.

Study excluded		Coefficient estimate [95% CI]		Coefficient estimate [95% CI]
PBS removed from				
study 1		0.004 [-0.004 , 0.011]		-0.024 [-0.043 , -0.005]
study 2		0.005 [-0.002 , 0.012]		-0.029 [-0.053 , -0.005]
study 8a		0.002 [-0.005 , 0.008]		-0.029 [-0.056 , -0.002]
study 8b		0.003 [-0.003 , 0.009]		-0.030 [-0.055 , -0.006]
study 8c		0.004 [-0.003 , 0.011]		-0.028 [-0.053 , -0.004]
study 9		0.004 [-0.003 , 0.012]		-0.029 [-0.054 , -0.004]
study 12		0.004 [-0.003 , 0.010]		-0.024 [-0.045 , -0.003]
study 18		0.006 [-0.003 , 0.014]		-0.033 [-0.061 , -0.006]
study 21		0.003 [-0.005 , 0.011]		-0.030 [-0.055 , -0.005]
study 22		0.002 [-0.006 , 0.010]		-0.024 [-0.046 , -0.001]
1st-year removed from				
study 1		0.003 [-0.004 , 0.011]		-0.029 [-0.052 , -0.005]
study 2		0.003 [-0.005 , 0.010]		-0.029 [-0.053 , -0.005]
study 8a		0.004 [-0.004 , 0.011]		-0.029 [-0.054 , -0.004]
study 8b		0.003 [-0.004 , 0.010]		-0.027 [-0.049 , -0.004]
study 8c		0.003 [-0.004 , 0.010]		-0.030 [-0.055 , -0.005]
study 12		0.003 [-0.004 , 0.011]		-0.029 [-0.054 , -0.004]
study 13/14		0.004 [-0.004 , 0.011]		-0.030 [-0.055 , -0.004]
study 18		0.003 [-0.004 , 0.011]		-0.030 [-0.054 , -0.005]
study 20		0.004 [-0.004 , 0.011]		-0.030 [-0.055 , -0.004]
study 21		0.003 [-0.005 , 0.011]		-0.029 [-0.053 , -0.005]
Full model		**0.004 [-0.004 , 0.011]**		**-0.029 [-0.054 , -0.004]**

Coefficient estimate for PFxMonth   Coefficient estimate for (PF+MI)xMonth

Figure xiii: Results from sensitivity analyses where a covariate from each study was sequentially removed throughout the entire analysis. The effects of the exclusion of a continuous covariate (PBS at baseline; top) and a binary covariate (first-year student status; bottom) on the combined estimates of PF x Month (left) and (MI+PF) x Month (right) are shown, respectively. Filled diamond symbols indicate the combined estimates from all 14 studies as reported in Table 4. The estimates from sensitivity analyses are shown in filled squares.
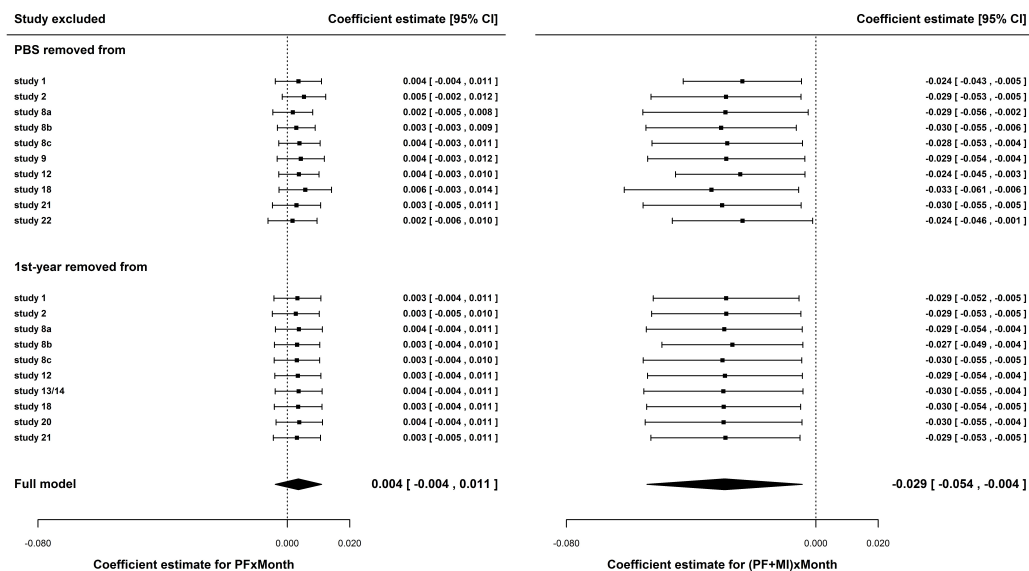
former is aimed at combining the entire full model, which is subsequently used to derive model-based estimates for flexible inference, whereas the latter is focused on deriving point estimates. In the present study, the combined full model had 13 point estimates and their 13 x 13 covariance estimates across 11 different estimable patterns of coefficients for 14 different clinical trials. We used a mapping matrix approach to identify appropriate connections of the study-specific estimates to the full model vector and, subsequently, combined all estimates using a multivariate CD-based approach. The approach we illustrated may provide the field with a valuable methodological approach to consider, in connection with methods of aggregating published prediction models (Debray et al., 2012, 2014), dealing with systematic missing data (Collaboration et al., 2009; Jolani et al., 2015; Resche-Rigon et al., 2013; Siddique et al., 2015), exploring subgroups that may respond differently to an intervention (Fisher et al., 2011; Riley et al., 2015b), and combining multiple parameter estimates from AD or IPD (Cheung and Hafdahl, 2016; Gasparrini et al., 2012).

The promise of the current method may be helpful for a situation described by Riley et al. (2015b). Riley and colleagues explored an extension application in which different treatment effects are examined separately for each subgroup using a multivariate meta-analysis formulation. However, this approach, as Riley et al. discussed, can result in a confounded estimate when some of the studies included in a meta-analysis do not have all subgroups. For example, when there are studies of only men or women, the resulting treatment difference for men vs. women from the bivariate meta-analysis approach as formulated in Riley et al. would include not only the within-study, relative treatment difference estimate but also the estimate difference between a set of studies with all men and another set of studies with all women. This approach may be defensible if within-study and between-study covariate interactions are the same, which may be unrealistic for clinical trials. Riley et al. discussed the advantages (e.g., power) and disadvantages (e.g., ecological bias and study-level confounding) of combining within-study estimates with study-level, between-study estimates of the approach. The method we illustrated in the current study may offer a more favorable solution to this challenge. Instead of quantifying a few isolated treatment effect sizes for different subgroups in

a synthesis, we estimate a full model that includes treatment by covariate interaction terms and derive model-based estimates of subgroup responses to a treatment. Via mapping matrices, all study-specific coefficients combined for the full model would be within-study interaction terms that are made comparable to the estimates from other trials by taking into account between-study differences. Therefore, there would be no study-level confounding. Although we illustrated this method using IPD, it can also apply to AD.

One of the most important advantages of the CD-based method may be that it helps to expand the dimension of evidence from which inference can be drawn by combining the entire full model rather than a few point estimates. Multivariate meta-analysis, despite its well established rationale and promise, has resulted in a rather small improvement in the statistical properties of the individual estimates (Jackson et al., 2011; Trikalinos et al., 2014). In typical multivariate meta-analysis applications, the dimension of coefficients combined has been rather limited, which may have been a contributing factor for the marginal gain reported in multivariate applications. In contrast, we had more to draw on when we borrowed information from within-study correlations (up to 13 coefficients). In addition, through the use of a multivariate random-effects model, we borrowed from the between-study correlations across 14 studies. As expected, trials with large intervention effects for one treatment arm tended to have large effects for the other treatment arm in the current study (see Table 5). In sum, compared to a typical multivariate meta-analysis of just a few point estimates, the proposed method can make more specific inferences, which may be helpful for the development of personalized treatment approaches using clinical trial data (i.e., the Precision Medicine Initiative; Collins and Varmus (2015)).

It is also important to note that the proposed CD-based method is flexible with respect to different types of multiple related coefficients or different outcome distributions. In the current study, we simultaneously combined three different types of related coefficients: the relative intervention benefits (network), informative covariates (regression), and longitudinal associations (longitudinal). The modeling flexibility of the current

method also extends to outcome data distributions because there are no restrictions on the distributions of outcome variables for the underlying full model.

Having discussed the promise of the proposed CD-based approach, it may be helpful to draw attention to the assumptions utilized in the current study and potential caveats. First, we made an assumption that the underlying full model is the correct model for all studies. We developed the full model by drawing on expertise in alcohol interventions for college student populations and also by closely checking individual data sets. We fully leveraged modeling flexibilities of having IPD, such as the ability to check data quality; check any imbalance in RCTs; check model assumptions; and perform attrition/missing data analysis and any other additional analyses. However, this process took considerable time. Furthermore, the identification of appropriate mapping matrices is needed for each missing data pattern. In sum, the proposed synthesis method can take considerable time and efforts and require a wide range of complementary expertise and skills. If the full model is incorrect or if mapping matrices are incorrect, how robust the method is to model misspecification is unknown and remains to be fully studied. Although the sensitivity analyses suggested that our approach was generally robust, future simulation studies of this new approachs empirical performance under various data situations, relative to other approaches, would be informative.

Second, the included studies should be sufficiently similar in terms of their methodological and clinical characteristics to justify combining data. In the present study, we used new measures that had been harmonized and analyzed in previous studies via advanced IRT models so that covariates and outcomes could be meaningfully compared and interpreted across different studies (Huo et al., 2015; Mun et al., 2015a). Moreover, the original 24 studies included in Project INTEGRATE were selected for their similarity in interventions and target populations, and the specific 14 studies included in the current synthesis met the additional inclusion criteria. Furthermore, because we combined data from RCTs, participants did not self-select into a trial or an intervention group in the current synthesis. Consequently, there was little evidence of covariation among the covariates included in the full model. In essence, we made an assumption that the

underlying full model is a true model for all studies except that some of the covariates may not be estimable for the reasons that do not affect the derived estimates, which directly leads to the next assumption.

Third, our approach assumes that the pattern of omitted covariates at the study level meets the MAR assumption, which may be quite reasonable for RCTs (Riley et al., 2015a). All of the study-level missing data occurred because the original studies did not assess the covariates included in the analysis or did not have all intervention groups by design. This is a reasonable assumption in our case because we had access to all IPD in all studies and the combined coefficients did not come from published studies. Therefore, the results from the current study may be less prone to selective reporting bias, under which the MAR may not be reasonable.

Some of the limitations of the present study are that we did not accommodate any uncertainty surrounding the covariance estimators $\boldsymbol{S}_i$ and $\hat{\boldsymbol{\Sigma}}_{\text{REML}}$. However, individual studies in our motivated data example had moderate to large sample sizes and the number of studies analyzed was not small. In future meta-analysis studies, this uncertainty may be reflected, for example, by inflating confidence intervals for the REML method (Jackson and Riley, 2014). Second, to make IPD comparable across studies for the proposed method, an additional set of complex analyses using item-level IPD may be required, which can take considerable efforts and time and may not always be feasible especially in a large synthesis study (Mun et al., 2015a). Third, when a data set for synthesis differs from the larger data set used to calibrate item parameters and derive commensurate latent trait scores across studies because of an additional inclusion criteria needed for synthesis, this difference, as well as the assumptions and model specifications used in the IRT analysis, may need to be carefully evaluated. Fourth, the CD-based approach to combine information may not work well for other studies if covariates are expected to be highly correlated within studies. In such situations, the interpretation of each coefficient reflects the list of all other predictors in the model. Consequently, submodels may not be comparable across different studies, and the resulting study-specific coefficients may not validly be linked to the full model. In the

current synthesis of RCTs under the specific full model with mostly binary covariates, the CD-based approach worked quite well. However, how this method works under different conditions may require more carefully planned simulation studies.

Substantively, we found the positive effect of the MI + PF intervention on alcohol-related problems, which is consistent with the findings reported by Huh et al. (2015), despite using a different methodological approach to IPD meta-analysis. Huh et al. (2015) estimated a Bayesian three-level model in a one-stage analysis of IPD using the MCMCglmm R package (Hadfield et al., 2010). In addition, Huh et al. (2015) used a different full model, which specified alcohol-related problem scores at baseline as a covariate. In terms of taking advantage of within- and between-study correlations, the highest level was set at the level of the randomized groups within studies (or study group), instead of at the level of studies. This model specification was a practical compromise to simultaneously analyze IPD in a one-stage analysis using data from studies with heterogeneous designs. Consequently, the correct data structure that multiple intervention groups were nested within studies was overlooked in Huh et al. (2015). Moreover, studies without a control group were excluded in the analysis because the intervention effect size was estimated by comparing the estimates of intervention groups with their corresponding control group within study. The covariates examined were also limited in Huh et al. (2015). because only a common set of covariates available across all studies could be considered.

From the comparison of the one-stage approach (Huh et al., 2015) and the current three-stage CD-based approach, albeit indirectly, we reach two conclusions. First, the convergent findings from the two studies increase our confidence in the substantive conclusion that MI + PF has an advantage over PF in terms of reducing alcohol-related problems for college students. Second, the three-stage CD-based approach to IPD multivariate meta-analysis may be better suited to build a scalable evidence base, compared to the one-stage approach utilized in Huh et al. (2015). The term scalable applies not only to the number of studies but also to the number of informative covariates that can be examined in a meta-analysis. Given that the resources and time required for IPD

meta-analysis are considerable (Berlin et al., 2013; Hussong et al., 2013; Mun et al., 2015a; Mun and Ray, 2016; Steinberg et al., 1997), especially in the context that both IPD meta-analysis and AD meta-analysis produce similar conclusions about main treatment effects (e.g., (Olkin and Sampson, 1998)), the best use of IPD may be to examine informative covariates and subgroups (Tian et al., 2012; Zhang et al., 2008). Toward this end, the CD-based method may be an important new tool. In conclusion, to provide answers to complex questions from the available large-scale data, it is critical to account for between-study differences, which has been discussed as the most significant challenge for the current meta-analysis field (Hedges, 2016). The proposed method is aimed at promoting large-scale, complex evidence synthesis of IPD to shed light on complex phenomena by offering one methodical route to overcome this critical barrier for multivariate meta-analysis models and meta-analysis of IPD.

Chapter 4

Concluding Remarks

In this dissertation, we briefly review the existing CD-based inference methods and extend them in several important ways. We also apply these novel approaches to real world problems. With a confidence distribution as the synthesizing instrument, CD approaches showed its potential and promise, compared to conventional meta-analysis methods based on point or interval estimates. The CD approach is more efficient, can be more flexible to incorporate prior information, and can mitigate the influence of outlying studies. Furthermore, it is generally robust to model mis-specifications. More importantly, it provides a unified meta-analysis framework that subsumes most conventional methods.

In Chapter 2, we develop the CD framework to analyze 2×2 tables. The proposed method can takes into account of the data sampling scheme used to obtain data and can be applied to develop confidence distributions for most commonly used metrics of a 2×2 table. By incorporating prior information, the approach can also be applied to rare events data for 2×2 tables with zero observed events without any artificial corrections. In Chapter 3, we extend the current method of combining CD random vectors to a multivariate random-effects meta-analysis model and tackle the challenge of applying the methodology to real word complex data sets with a higher dimension and incomplete data. Conclusion from the analysis of Project INTEGRATE are consistent with current fundings in the field. The sensitivity analysis shows that the results is robust. Future studies should investigate the empirical performance of the proposed method under various data situations and examine robustness of the method under different types of model mis-specifications, such as incorrectly specified mapping matrices.

To close this dissertation, we remark that the idea of combining confidence distributions to synthesize information is very powerful and can be potentially used to deal with other open and challenging problems in the field of meta-analysis.

Bibliography

Ades, A. and Sutton, A. (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(1):5–35.

Agresti, A. (2003). *Categorical Data Analysis*. John Wiley & Sons, Inc.

Agresti, A. (2007). *An introduction to categorical data analysis*, volume 135. Wiley New York.

Bauer, D. J. and Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychological methods*, 14(2):101.

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

Berlin, J. A., Crowe, B. J., Whalen, E., Xia, H. A., Koro, C. E., and Kuebler, J. (2013). Meta-analysis of clinical trial safety data in a drug development program: answers to frequently asked questions. *Clinical Trials*, 10(1):20–31.

Chen, H., Manning, A. K., and Dupuis, J. (2012). A method of moments estimator for random effect multivariate meta-analysis. *Biometrics*, 68(4):1278–1284.

Cheung, M. W.-L. (2015). *Meta-analysis: A structural equation modeling approach*. John Wiley & Sons.

Cheung, M. W.-L. and Cheung, S. F. (2016). Random-effects models for meta-analytic structural equation modeling: review, issues, and illustrations. *Research Synthesis Methods*, 7(2):140–155.

Cheung, M. W.-L. and Hafdahl, A. R. (2016). Special issue on meta-analytic structural equation modeling: introduction from the guest editors. *Research Synthesis Methods*, 7(2):112–120.

Cipriani, A., Higgins, J. P., Geddes, J. R., and Salanti, G. (2013). Conceptual and technical challenges in network meta-analysis. *Annals of Internal Medicine*, 159(2):130–137.

Claggett, B., Xie, M., and Tian, L. (2014). Meta-analysis with fixed, unknown, study-specific parameters. *Journal of the American Statistical Association*, 109(508):1660–1671.

Collaboration, F. S. et al. (2009). Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Statistics in medicine*, 28(8):1218.

Collins, F. S. and Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, 372(9):793–795.

Cox, D. R. (1958). Some problems connected with statistical inference. *The Annals of Mathematical Statistics*, 29(2):357–372.

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., and Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: the role of item response theory in integrative data analysis. *Developmental Psychology*, 44(2):365.

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., Sher, K., and Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate behavioral research*, 49(3):214–231.

Debray, T., Koffijberg, H., Nieboer, D., Vergouwe, Y., Steyerberg, E. W., and Moons, K. G. (2014). Meta-analysis and aggregation of multiple published prediction models. *Statistics in medicine*, 33(14):2341–2362.

Debray, T., Koffijberg, H., Vergouwe, Y., Moons, K. G., and Steyerberg, E. W. (2012). Aggregating published prediction models with individual participant data: a comparison of different approaches. *Statistics in Medicine*, 31(23):2697–2712.

Debray, T. P., Moons, K. G., Abo-Zaid, G. M. A., Koffijberg, H., and Riley, R. D. (2013). Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PLoS One*, 8(4):e60650.

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

Efron, B. (1998). Ra fisher in the 21st century. *Statistical Science*, pages 95–114.

Embretson, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist*.

Fisher, D., Copas, A., Tierney, J., and Parmar, M. (2011). A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. *Journal of clinical epidemiology*, 64(9):949–967.

Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of eugenics*, 6(4):391–398.

Fisher, S. R. A. (1956). 6 mathematics of a lady tasting tea. *The World of Mathematics*.

Fraser, D. A. S. (1991). Statistical inference: Likelihood to significance. *Journal of the American Statistical Association*, 86(414):258–265.

Gasparrini, A. and Armstrong, B. (2011). Multivariate meta-analysis: A method to summarize non-linear associations. *Statistics in medicine*, 30(20):2504–2506.

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

Gelman, A., King, G., and Liu, C. (1998). Not asked and not answered: Multiple imputation for multiple surveys. *Journal of the American Statistical Association*, 93(443):846–857.

Hadfield, J. D. et al. (2010). Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22.

Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica*, pages 491–544.

Hannig, J. (2013). Generalized fiducial inference via discretization. *Statistica Sinica*, 19:489–514.

Hannig, J., Iyer, H., Lai, R. C., and Lee, T. C. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111:1346–1361.

Hedges, L. V. (2016). Applying meta-analysis to structural equation modeling. *Research Synthesis Methods*, 7(2):209–214.

Hofer, S. M. and Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological methods*, 14(2):150.

Huh, D., Mun, E.-Y., Larimer, M. E., White, H. R., Ray, A. E., Rhew, I. C., Kim, S.-Y., Jiao, Y., and Atkins, D. C. (2015). Brief motivational interventions for college student drinking may not be as powerful as we think: An individual participant-level data meta-analysis. *Alcoholism: Clinical and Experimental Research*, 39(5):919–931.

Huo, Y., de la Torre, J., Mun, E.-Y., Kim, S.-Y., Ray, A. E., Jiao, Y., and White, H. R. (2015). A hierarchical multi-unidimensional irt approach for analyzing sparse, multi-group data for integrative data analysis. *Psychometrika*, 80(3):834–855.

Hussong, A. M., Curran, P. J., and Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual review of clinical psychology*, 9:61.

Ioannidis, J. (2010). Meta-research: The art of getting it wrong. *Research Synthesis Methods*, 1(3-4):169–184.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, 2(8):e124.

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

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

Jackson, D. and Riley, R. D. (2014). A refined method for multivariate meta-analysis and meta-regression. *Statistics in medicine*, 33(4):541–554.

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

Jansen, J. P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., Lee, K., Boersma, C., Annemans, L., and Cappelleri, J. C. (2011). Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ispor task force on indirect treatment comparisons good research practices: part 1. *Value in Health*, 14(4):417–428.

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

Jolani, S., Debray, T., Koffijberg, H., Buuren, S., and Moons, K. G. (2015). Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using mice. *Statistics in medicine*, 34(11):1841–1863.

Jones, A. P., Riley, R. D., Williamson, P. R., and Whitehead, A. (2009). Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clinical Trials*, 6(1):16–27.

Liu, D., Liu, R. Y., and Xie, M. (2015). Multivariate meta-analysis of heterogeneous

studies using only summary statistics: efficiency and robustness. *Journal of the American Statistical Association*, 110(509):326–340.

Liu, D., Liu, R. Y., and Xie, M.-g. (2014). Exact meta-analysis approach for discrete data and its application to 2× 2 tables with rare events. *Journal of the American Statistical Association*, 109(508):1450–1465.

Martens, M. P., Ferrier, A. G., Sheehy, M. J., Corbett, K., Anderson, D. A., and Simmons, A. (2005). Development of the protective behavioral strategies survey. *Journal of studies on alcohol*, 66(5):698–705.

Mun, E.-Y., De La Torre, J., Atkins, D. C., White, H. R., Ray, A. E., Kim, S.-Y., Jiao, Y., Clarke, N., Huo, Y., Larimer, M. E., et al. (2015a). Project integrate: An integrative study of brief alcohol interventions for college students. *Psychology of Addictive Behaviors*, 29(1):34.

Mun, E.-Y., Jiao, Y., and Xie, M. (2015b). Integrative data analysis for research in developmental psychopathology. *Developmental Psychopathology*.

Mun, E.-Y. and Ray, A. E. (2016). Integrative data analysis from a unifying research synthesis perspective. *In Developmental perspectives of alcohol and other addictions over the life span Eds. HE Fitzgerald and LI Puttler (in press)*.

Muraki, E. (1992). A generalized partial credit model: Application of an em algorithm. *ETS Research Report Series*, 1992(1):i–30.

Nash, J. C., Varadhan, R., et al. (2011). Unifying optimization algorithms to aid software system users: optimx for r. *Journal of Statistical Software*, 43(9):1–14.

Nissen, S. E. and Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine*, 356(24):2457–2471.

Normand, S. (1999). Meta-analysis: formulating, evaluating, combining and reporting stat med 18: 321–359. *Find this article online.*

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

O'Rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*, 100(12):579–582.

Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2014). R core team (2014) nlme: linear and nonlinear mixed effects models. r package version 3.1-117. *See http://CRAN. R-project. org/package= nlme.*

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, volume 1. Sage.

Ray, A. E., Kim, S.-Y., White, H. R., Larimer, M. E., Mun, E.-Y., Clarke, N., Jiao, Y., Atkins, D. C., and Huh, D. (2014). When less is more and more is less in brief motivational interventions: Characteristics of intervention content and their associations with drinking outcomes. *Psychology of Addictive Behaviors*, 28(4):1026.

Reiter, J. P. (2008). Multiple imputation when records used for imputation are not used or disseminated for analysis. *Biometrika*, 95(4):933–946.

Resche-Rigon, M., White, I. R., Bartlett, J. W., Peters, S. A., and Thompson, S. G. (2013). Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Statistics in Medicine*, 32(28):4890–4905.

Riley, R., Price, M., Jackson, D., Wardle, M., Gueyffier, F., Wang, J., Staessen, J. A., and White, I. (2015a). Multivariate meta-analysis using individual participant data. *Research synthesis methods*, 6(2):157–174.

Riley, R. D., Abrams, K. R., Sutton, A. J., Lambert, P. C., and Thompson, J. R. (2007). Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology*, 7(1):1.

Riley, R. D., Elia, E. G., Malin, G., Hemming, K., and Price, M. P. (2015b). Multivariate meta-analysis of prognostic factor studies with multiple cut-points and/or methods of measurement. *Statistics in medicine*, 34(17):2481–2496.

Riley, R. D., Lambert, P. C., and Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *Bmj*, 340:c221.

Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L., and Boutitie, F. (2008). Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in medicine*, 27(11):1870–1893.

Riley, R. D. and Steyerberg, E. W. (2010). Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods*, 1(1):2–19.

Siddique, J., Reiter, J. P., Brincks, A., Gibbons, R. D., Crespi, C. M., and Brown, C. H. (2015). Multiple imputation for harmonizing longitudinal non-commensurate measures in individual participant data meta-analysis. *Statistics in medicine*, 34(26):3399–3414.

Simmonds, M. and Higgins, J. (2007). Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. *Statistics in medicine*, 26(15):2982–2999.

Simmonds, M., Stewart, G., and Stewart, L. (2015). A decade of individual participant data meta-analyses: a review of current practice. *Contemporary clinical trials*, 45:76–83.

Simmonds, M. C., Higginsa, J. P., Stewartb, L. A., Tierneyb, J. F., Clarke, M. J., and Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials*, 2(3):209–217.

Singh, K., Xie, M., and Strawderman, W. E. (2007). Confidence distribution (cd): Distribution estimator of a parameter. *Lecture Notes-Monograph Series*, pages 132–150.

Singh, K., Xie, M., Strawderman, W. E., et al. (2005). Combining information from independent sources through confidence distributions. *The Annals of Statistics*, 33(1):159–183.

Steinberg, K., Smith, S., Stroup, D., Olkin, I., Lee, N., Williamson, G., and Thacker, S. (1997). Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. *American Journal of Epidemiology*, 145(10):917–925.

Sutton, A. J. and Higgins, J. (2008). Recent developments in meta-analysis. *Statistics in medicine*, 27(5):625–650.

Tian, L., Cai, T., Zhao, L., and Wei, L.-J. (2012). On the covariate-adjusted estimation for an overall treatment difference with data from a randomized comparative clinical trial. *Biostatistics*, 13(2):256–273.

Trikalinos, T. A., Hoaglin, D. C., Small, K. M., Terrin, N., and Schmid, C. H. (2014). Methods for the joint meta-analysis of multiple tests. *Research synthesis methods*, 5(4):294–312.

Trikalinos, T. A. and Olkin, I. (2012). Meta-analysis of effect sizes reported at multiple time points: a multivariate approach. *Clinical Trials*, 9(5):610–620.

White, H. R., Mun, E. Y., Pugh, L., and Morgan, T. J. (2007). Long-term effects of brief substance use interventions for mandated college students: Sleeper effects of an in-person personal feedback intervention. *Alcoholism: Clinical and Experimental Research*, 31(8):1380–1391.

Wilson, S. J., Polanin, J. R., and Lipsey, M. W. (2016). Fitting meta-analytic structural equation models with complex datasets. *Research Synthesis Methods*, 7(2):121–139.

Wu, M.-J. and Becker, B. J. (2013). Synthesizing regression results: a factored likelihood method. *Research synthesis methods*, 4(2):127–143.

Xie, M., Singh, K., and Strawderman, W. E. (2012). Confidence distributions and a

unifying framework for meta-analysis. *Journal of the American Statistical Association.*

Xie, M.-g. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: a review. *International Statistical Review*, 81(1):3–39.

Yamaguchi, Y., Sakamoto, W., Goto, M., Staessen, J. A., Wang, J., Gueyffier, F., and Riley, R. D. (2014). Meta-analysis of a continuous outcome combining individual patient data and aggregate data: a method based on simulated individual patient data. *Research synthesis methods*, 5(4):322–351.

Yang, G., Liu, D., Liu, R. Y., Xie, M., and Hoaglin, D. C. (2014). Efficient network meta-analysis: A confidence distribution approach. *Statistical methodology*, 20:105–125.

Yang, G., Liu, D., Wang, J., and Xie, M.-g. (2016). Meta-analysis framework for exact inferences with application to the analysis of rare events. *Biometrics.*

Zeng, D. Z. and Lin, D. (2015). On random-effects meta-analysis. *Biometrika.*

Zhang, M., Tsiatis, A. A., and Davidian, M. (2008). Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics*, 64(3):707–715.