Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness

Dungang Liu[1], Regina Liu[2], and Minge Xie[2]

[1]Department of Biostatistics, Yale University School of Public Health, New Haven, CT 06511, USA

[2]Department of Statistics and Biostatistics, Rutgers University, New Brunswick, NJ 08854, USA

January 22, 2014

[12]Dungang Liu is Postdoctoral Associate, Department of Biostatistics, Yale University School of Public Health, New Haven, CT 06511. Email: dungang.liu@yale.edu. Regina Liu and Minge Xie are Professors, Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854. Emails: rliu; mxie@stat.rutgers.edu. The authors thank Dr. Andrew Cheng from the Federal Aviation Administration for providing expertise on aviation domain. They also thank the Associate Editor and two anonymous referees for their helpful comments. This work is based on part of the first author's Ph.D. dissertation. The first author gratefully acknowledges Fellowship from Rutgers University, Research Assistantship from NSF grant DMS 0707053 and support from NIH-R01-DA016750-09. The second and third authors acknowledge support from grants: NSF-DMS 1107012; 1007683; 0915139; 0707053, NSF-SES 0851521, and NSA-H11-1-0157.

Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness

Abstract

Meta-analysis has been widely used to synthesize evidence from multiple studies for common hypotheses or parameters of interest. However, it has not yet been fully developed for incorporating heterogeneous studies, which arise often in applications due to different study designs, populations or outcomes. For heterogeneous studies, the parameter of interest may not be estimable for certain studies, and in such a case, these studies are typically excluded from conventional meta-analysis. The exclusion of part of the studies can lead to a non-negligible loss of information. This paper introduces a meta-analysis for heterogeneous studies by combining the *confidence density functions* derived from the summary statistics of individual studies, hence referred to as the CD approach. It includes all the studies in the analysis and makes use of all information, direct as well as indirect. Under a general likelihood inference framework, this new approach is shown to have several desirable properties, including: i) it is asymptotically as efficient as the maximum likelihood approach using individual participant data (IPD) from all studies; ii) unlike the IPD analysis, it suffices to use summary statistics to carry out the CD approach. Individual-level data are not required; and iii) it is robust against misspecification of the working covariance structure of the parameter estimates. Besides its own theoretical significance, the last property also substantially broadens the applicability of the CD approach. All the properties of the CD approach are further confirmed by data simulated from a randomized clinical trials setting as well as by real data on aircraft landing performance. Overall, one obtains an unifying approach for combining summary statistics, subsuming many of the existing meta-analysis methods as special cases.

Key words: combining information, complex evidence synthesis, confidence distribution, efficiency, generalized estimating equations, heterogeneous studies, indirect evidence, individual participant data, multivariate meta-analysis.

1 Introduction

Meta-analysis is one of the most commonly used approaches for combining findings from a series of independent studies. It has been used extensively in many fields, such as medicine, epidemiology, education, psychology, among others. The growing emphasis on evidence-based decision-making in those fields has brought to the fore the need for developing meta-analysis for more complex settings. The goal of this paper is to introduce a meta-analysis for the complex setting of heterogeneous studies, using the approach of combining the *confidence density functions* derived from the summary statistics of individual studies. This approach, referred to as the CD approach henceforth, unlike the traditional meta analysis, can include all the studies in the analysis and make use of both direct and indirect evidence. In addition, under a general likelihood inference framework, we can show that: i) The CD approach is asymptotically as efficient as the maximum likelihood approach using individual participant data (IPD) from all the studies; ii) It suffices to use summary statistics to carry out the CD approach and individual-level IPD data are not required; and iii) The CD approach is robust against misspecification of the working covariance structure of the parameter estimates.

In many meta-analysis investigations, the studies under consideration are often heterogeneous. For instance, Sutton and Higgins (2008) showed that the heterogeneity can arise from the differences in study populations, designs or outcomes. This naturally leads to *parameter heterogeneity* among the studies, in the sense that the estimable parameters are different from one study to another. At times, the parameter of interest may not even be estimable in some of the studies. In such situations, since these studies do not provide any direct information for inference for the parameter of interest, such as point estimates, they are generally excluded from the conventional meta-analysis. However, this exclusion can lead to a non-negligible or even substantial loss of information. To overcome this problem, we propose a meta-analysis approach that can incorporate all the studies in an efficient manner.

We use a basic fixed-effects model to illustrate a broad range of heterogeneity settings. Consider a meta-analysis of K independent clinical trials with the following fixed-effects linear model:

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

where Y_{ij} is the response for the j-th subject in the i-th study, X_{ij} the treatment status ($1/0$ for treatment/control), Z_{ij} the covariate of interest (e.g., drug dosage), and ε_{ij} the noise variable following $N(0, \sigma_i^2)$. Here, α_i's are the study-specific effects and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ is the common effect among all studies.

This model is often used to examine, in addition to the treatment effect, the covariate effect as well as the treatment-covariate interaction effect in randomized clinical trials, as shown in, for example, Simmonds and Higgins (2007). Using this model, Simmonds and Higgins (2007) investigated the power of different meta-analysis methods in detecting the interaction effect β_3. They showed that the conventional meta-analysis method of simply weighting the point estimates of β_3 from each of the studies suffers loss of power in testing, or equivalently, of efficiency in estimation. Although Lin and Zeng (2010) showed that this loss of efficiency can be avoided if the point estimates of the vector parameter $\boldsymbol{\beta}$ are combined using the inverse of the corresponding covariance matrix as the weight, both methods break down when heterogeneity is present among the studies, as illustrated in Examples 1-3 below. In each of these examples, at least one of the parameters is not estimable in certain studies, due to the heterogeneity in populations, designs, or/and outcomes. Consequently, some studies can not be utilized in conventional analysis, resulting in a loss of efficiency. This point will be elaborated further both theoretically and numerically.

Example 1 (Heterogeneity in populations)**.** The studies collected for meta-analysis may be from different populations defined by distinct gender, race or disease status of study subjects. The population heterogeneity may affect the effect size and thus require the specification of study-specific effects in statistical modeling, such as α_i's in Model (1). The study-specific effects are of interest when the study-specific inference is part of the research goal. For example, in clinical trials, it is also crucial to evaluate the treatment effect for certain subpopulations (see, e.g., Entsuah, Huang, and Thase, 2001; Shekelle et al., 2003; Zarrouf et al., 2009). However, it is clear that the study-specific effect Study 1, namely α_1, is not estimable in the other studies. Hence, none of Study 2, $\cdots$, Study K can be utilized in the conventional meta-analysis for making inference on α_1.

Example 2 (Heterogeneity in covariate designs)**.** The covariate designs may be different from one study to another since each study may have its own specific considerations and constraints. In particular, Simmonds and Higgins (2007) pointed out a situation with *missing covariate designs*; that is, certain studies do not have the design covariate that is of current research interest. For example, if Study 1 does not aim to examine the effect of the covariate Z_{1j} (e.g., drug dosage), the Z_{1j} of all the subjects in Study 1 are typically controlled at a fixed value (i.e., same dosage), say $Z_{1j} \equiv z$. Therefore, for Study 1, Model (1) becomes

$$Y_{1j} = (\alpha_1 + z\beta_2) + (\beta_1 + z\beta_3)X_{1j} + \varepsilon_{1j}, \quad j = 1, \ldots, n_1. \tag{2}$$

Here the interaction effect β_3 is not estimable, and thus Study 1 can not be utilized in the conventional

meta-analysis for making inference on β_3.

Example 3 (Heterogeneity in outcomes)**.** Another challenging problem in meta-analysis arises in the difficulty of synthesizing studies with different types of outcome report (Dominici and Parmigiani, 2000), often due to different data report policies. Whitehead, Bailey, and Elbourne (1999) discussed an example on reporting blood loss in women labor. In that example, even though the outcome of interest (namely, blood loss) is continuous, some studies may choose to report only binary outcomes (e.g., "severe" or "not severe") indicating whether or not the underlying continuous outcome exceeds a prefixed threshold. Under the working Model (1), suppose that Study 1 reports only the binary responses d_{1j}, where $d_{1j} = 1$ if $y_{1j} \geq \tau_1$ and $d_{1j} = 0$ otherwise, with the prefixed threshold τ_1. It is easy to see that, for Study 1, Model (1) now reduces to the following probit model:

$$\Pr(d_{1j} = 1) = \Phi\left(\frac{\alpha_1 - \tau_1}{\sigma_1} + \frac{\beta_1}{\sigma_1}X_{1j} + \frac{\beta_2}{\sigma_1}Z_{1j} + \frac{\beta_3}{\sigma_1}Z_{1j}X_{1j}\right), \quad j = 1, \ldots, n_1. \tag{3}$$

Here again, the interaction effect β_3 is not estimable (only β_3/σ_1 is estimable since σ_1 is unknown), and Study 1 can not be utilized in the conventional meta-analysis for making inference on β_3.

Note that all the examples above involve multiple parameters, and the information for one parameter may potentially impact the inference on the other parameters. In other words, the studies excluded from the conventional meta-analysis may contain *indirect evidence* for the common parameters. Clearly, such indirect evidence is useful and can be utilized if IPD data are available for all the relevant studies, since, in this case, an effective inference can be made from multiplying individual-level likelihood functions from all the studies. It is worth pointing out that although the IPD method is regarded as the "gold standard" for combining information in the literature, its implementation can be difficult, costly, and often impractical for various reasons, such as concerns over data confidentiality issues and reluctance of original researchers to release the full data.

This last observation naturally raises the question: *Can we retain full efficiency (achieved by using IPD data) by using only summary statistics to perform meta-analysis of heterogeneous studies?* This question has been investigated in the literature of meta-analysis in various settings but all within the scope of homogeneous studies, see, e.g., Olkin and Sampson (1998), Mathew and Nordstrom (1999), Simmonds and Higgins (2007), and Lin and Zeng (2010), among others. In particular, Lin and Zeng (2010) showed that asymptotically there is no loss of efficiency by analyzing only summary statistics in the framework of likelihood inference for homogeneous studies. However, the question on relative efficiency of analyzing

summary statistics versus IPD data in meta analysis for heterogeneous studies is more challenging and has remained unresolved thus far.

This paper provides an affirmative answer to this question. Our estimator derived from summary statistics is asymptotically as efficient as the IPD estimator and suffers no loss of efficiency, even when the studies are heterogeneous. This theoretical conclusion covers the results established in Olkin and Sampson (1998), Mathew and Nordstrom (1999), and Lin and Zeng (2010) as special cases. Those papers focused on the situation where the parameter of interest is estimable across all the studies. Ours extends the results to complex situations where the parameter of interest is not estimable in some studies due to the presence of heterogeneous studies.

Our method is to combine *confidence density functions* from relevant studies. The concept of confidence density and its cumulative counterpart, confidence distribution, has been developed extensively in recent years; see Xie and Singh (2013) and the references therein for a comprehensive review. Roughly speaking, the confidence distribution (density) is a sample-dependent distribution (density) function that can be used to estimate and provide all aspects of statistical inference for a parameter of interest. A confidence density function can carry a wealth of information (e.g., correlation) for multiple parameters, and is particularly well suited for the problem setting of this paper. This useful feature is also seen in Tian et al. (2010), where the confidence density is used to make joint inference about a set of constrained parameters in survival analysis.

The idea of combining confidence density functions lends itself to an unifying approach for combining summary statistics, which subsumes many existing meta-analysis methods. More specifically, in some special cases of the setting considered in this paper, the estimators derived from our approach reduce exactly to the ones from those existing methods. In particular, our approach covers the commonly used method of weighting point estimates (Lin and Zeng, 2010) and the multivariate generalized least square method (Becker and Wu, 2007), both of which perform linear combination of summary statistics. Our approach is broader, allowing incorporating indirect evidence in the analysis, and adapts to more complicated problem settings, by performing non-linear combination of summary statistics.

We also establish a robustness property of our approach. This property is of considerable importance, because it broadens substantially the applicability of our approach. Specifically, it shows that the approach applies to the situation when only estimates of the individual variances, rather than the full covariance matrices, of the parameter estimates can be obtained from the studies. In other words, the approach is

valid even if the covariance structure of parameter estimates is misspecified. As a matter of fact, our estimator remains consistent and asymptotically normal. This asymptotic result in essence follows the theory for *generalized estimating equations* described in Liang and Zeger (1986). We further show that our approach can achieve considerable gain in efficiency for the overall inference by using suitably chosen "working" correlation matrices.

The rest of this paper is organized as follows. In Section 2, we describe our approach and show that it is asymptotically as efficient as the IPD approach under a general likelihood inference framework. In Section 3, we establish the robustness property of our approach and illustrate its applications to the situation where only the marginal variance estimates are available. The theoretical results established in Section 2 and 3 are then numerically demonstrated in Section 4 using the data simulated from a setting of randomized clinical trials. In Section 5, we apply our approach to analyzing a data set on aircraft landing performance collected at some typical commercial airports by the Federal Aviation Administration. In fact, the project of analyzing aircraft landing performance is the original motivation for us to investigate meta-analysis for heterogenous studies and thus the research in this paper. Finally, we provide in Section 6 a discussion of some practical implications of the methodological development in this paper. All proofs for theoretical results are deferred to Appendix.

2 Methodology

2.1 The problem setting in its general likelihood inference framework

Consider K independent studies with n_i participants in the i-th study, $i = 1, \cdots, K$. Assume that we are interested in making inference for a $p \times 1$ vector parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$, where $\theta_1, \ldots, \theta_p$ are unknown fixed parameters associated with the K studies. We denote by $\{(X_{ij}, Y_{ij}), j = 1, \ldots, n_i\}$ the full data of the i-th study and assume that they are a random sample drawn from the density $f_i^*(x, y; \boldsymbol{\gamma}_i)$, where the $p_i \times 1$ parameter vector $\boldsymbol{\gamma}_i$ of the i-th study is linked to the parameter of interest $\boldsymbol{\theta}$ through a known smooth $\mathbb{R}^p \to \mathbb{R}^{p_i}$ mapping function $\boldsymbol{M}_i$ with $\boldsymbol{\gamma}_i = \boldsymbol{M}_i(\boldsymbol{\theta})$. We write the density function $f_i(x, y; \boldsymbol{\theta}) = f_i^*(x, y; \boldsymbol{\gamma}_i)$ and its likelihood function $L_i(\boldsymbol{\theta}) \equiv \prod_{j=1}^{n_i} f_i(x_{ij}, y_{ij}; \boldsymbol{\theta}) = \prod_{j=1}^{n_i} f_i(x_{ij}, y_{ij}; \boldsymbol{\gamma}_i) \equiv L_i^*(\boldsymbol{\gamma}_i)$. In this paper, we assume that $L_i^*(\boldsymbol{\gamma}_i)$ is identifiable with respect to the parameter $\boldsymbol{\gamma}_i$ in the sense that it can not be written in terms of a smaller subset of $\boldsymbol{\gamma}_i$. Also, for simplicity, we assume that the mapping function $\boldsymbol{M}_i$ is three times differentiable. Note that the function $\boldsymbol{M}_i$ is often determined by study designs. For instance, in Model (1),

if we let $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_K, \boldsymbol{\beta})^T$ and $\boldsymbol{\gamma}_i = (\alpha_i, \boldsymbol{\beta})^T$, then $\boldsymbol{M}_i$ is the mapping from $\boldsymbol{\theta}$ to $\boldsymbol{\gamma}_i$.

Consider likelihood inference. For the i-th study, we denote by $\hat{\boldsymbol{\gamma}}_i$ the maximum likelihood estimate of $\boldsymbol{\gamma}_i$, namely $\hat{\boldsymbol{\gamma}}_i = \arg\max_{\boldsymbol{\gamma}_i} L_i^*(\boldsymbol{\gamma}_i)$; and express the observed information matrix as $\Gamma_i(\boldsymbol{\gamma}_i) = -\partial^2 \log L_i^*(\boldsymbol{\gamma}_i)/\partial\boldsymbol{\gamma}_i \partial\boldsymbol{\gamma}_i^T$. Then, $\hat{\Sigma}_i = \Gamma_i^{-1}(\hat{\boldsymbol{\gamma}}_i)$ is an estimate of the covariance matrix for $\hat{\boldsymbol{\gamma}}_i$. Unless stated otherwise, we shall treat $\hat{\boldsymbol{\gamma}}_i$ and $\hat{\Sigma}_i$ as summary statistics of the i-th study throughout this paper. In our development, we focus on an asymptotic setting that n_i goes to infinity while K remains fixed. Let $n = \sum_{i=1}^K n_i$ and assume that $n_i/n \to c_i \in (0,1)$ as $n \to \infty$. We also assume throughout the paper the regularity conditions stated in Appendix B. Then, it follows that $\Gamma_i(\boldsymbol{\gamma}_i)/n_i \to I_i$ in probability and $n_i^{-1/2}\{\partial \log L_i^*(\boldsymbol{\gamma}_i)/\partial\boldsymbol{\gamma}_i\} \to \mathrm{MN}(\mathbf{0}, I_i)$ in distribution, as $n_i \to \infty$. Here, "MN" stands for multivariate normal distribution, and I_i is the $p_i \times p_i$ Fisher information matrix.

The setup above allows us to investigate the relative efficiency of analyzing summary statistics versus IPD data. It subsumes many commonly used parametric models, including, but not limited to, generalized linear models for continuous and categorical data, survival models for censored data, and mixed models for longitudinal data. The setup also allows us to assume different likelihood functions for the individual studies, such as linear regression model for continuous data in one study and logistic regression model for binary data in another study. Furthermore, the likelihood functions are not necessarily all from regression models, as seen in the case of meta-analysis of diagnostic accuracy data where sensitivity and specificity are of interest. Most important, since $\boldsymbol{M}_i$ can be any complex function satisfying some mild smoothness conditions, the framework here can be easily adapted to encompass a vast range of heterogeneous studies, including those in Examples 1-3 illustrated in Introduction. To sum up, *the key idea of the proposed approach is to bridge the estimable parameters $\boldsymbol{\gamma}_i$'s in different studies by using their associated functions $\boldsymbol{M}_i$'s to establish links to the common parameter $\boldsymbol{\theta}$*. Obviously, our setup reduces to the special case considered in Lin and Zeng (2010) if $\boldsymbol{M}_i$ is the identity transformation and $\boldsymbol{\gamma}_i \equiv \boldsymbol{\theta}$ for every i.

2.2 Combining confidence density functions

A confidence distribution is often viewed as *a sample-dependent distribution function that can represent confidence intervals of all levels for a parameter of interest*, (see, e.g., Cox, 1958; Efron, 1993; and the review in Xie and Singh (2013)). Cox (2013) stated that the CD approach provides "simple and interpretable summaries of what can reasonably be learned from data (and an assumed model)." A confidence distribution, if presented in a density function form, is referred to as *a confidence density*. More details

can be found in Xie and Singh (2013) and the references therein. Under the general setup in Section 2.1, the multivariate normal distribution $\text{MN}(\hat{\boldsymbol{\gamma}}_i, \hat{\Sigma}_i)$ can serve as a confidence distribution for the parameter $\boldsymbol{\gamma}_i$, as discussed in Singh et al. (2007). Accordingly, the density function of $\text{MN}(\hat{\boldsymbol{\gamma}}_i, \hat{\Sigma}_i)$ is a *confidence density* for the parameter $\boldsymbol{\gamma}_i$. Denote this density function by $h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i)$, where $\boldsymbol{S}_i$ represents the sample in the i-th study. More specifically,

$$h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i) = \frac{1}{(2\pi)^{p_i/2}|\hat{\Sigma}_i|^{1/2}} \exp\left\{\frac{1}{2}(\boldsymbol{\gamma}_i - \hat{\boldsymbol{\gamma}}_i)^T \hat{\Sigma}_i^{-1}(\boldsymbol{\gamma}_i - \hat{\boldsymbol{\gamma}}_i)\right\}, \quad i = 1, \ldots, K. \tag{4}$$

This sample-dependent density function contains rich information for frequentist inference. For instance, it provides confidence regions of all confidence levels for $\boldsymbol{\gamma}_i$; moreover, it provides confidence intervals of all confidence levels for any linear combination of the components of $\boldsymbol{\gamma}_i$. These points were elaborated in Singh et al. (2007) and Xie and Singh (2013).

We propose to combine the confidence density functions $h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i),\ i = 1, \ldots, K$, the same way as we combine likelihood functions for inference. Specifically, let

$$h(\boldsymbol{\theta}; \boldsymbol{S}_1, \ldots, \boldsymbol{S}_K) = \prod_{i=1}^{K} h_i(\boldsymbol{\gamma}_i; \boldsymbol{S}_i) = \prod_{i=1}^{K} h_i(\boldsymbol{M}_i(\boldsymbol{\theta}); \boldsymbol{S}_i). \tag{5}$$

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

$$\hat{\boldsymbol{\theta}}_{CD} = \arg\max_{\boldsymbol{\theta}} h(\boldsymbol{\theta}). \tag{6}$$

We establish the asymptotic properties of $\hat{\boldsymbol{\theta}}_{CD}$ in Theorem 1 below, which immediately imply the key result Theorem 2 claiming that the CD estimator and the IPD estimator are equally efficient asymptotically.

Theorem 1. *Under the setting specified in Section 2.1 and the regularity conditions stated in Appendix B, the CD estimator $\hat{\boldsymbol{\theta}}_{CD}$ obtained from (6) satisfies the following: as $n \to \infty$,*

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

$$n^{1/2}(\hat{\boldsymbol{\theta}}_{CD} - \boldsymbol{\theta}) \xrightarrow{d} \text{MN}\left(\boldsymbol{0}, \left\{\sum_{i=1}^{K} c_i J_i(\boldsymbol{\theta})^T I_i J_i(\boldsymbol{\theta})\right\}^{-1}\right), \tag{7}$$

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

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

$$\hat{\Sigma}_{CD}=\left\{-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\log h(\hat{\boldsymbol{\theta}}_{CD})\right\}^{-1}. \tag{8}$$

When individual-level data are available, the IPD estimator (namely, the MLE in this case) can be obtained by maximizing the multiplied likelihood function $L(\boldsymbol{\theta})=\prod_{i=1}^{K}L_i(\boldsymbol{\theta})$, namely

$$\hat{\boldsymbol{\theta}}_{IPD}=\arg\max_{\boldsymbol{\theta}}L(\boldsymbol{\theta}). \tag{9}$$

The IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$ is consistent and asymptotically normally distributed. In our framework, these can be expressed explicitly as follows (see Appendix for more details):

$$n^{1/2}(\hat{\boldsymbol{\theta}}_{IPD}-\boldsymbol{\theta})\xrightarrow{d}\text{MN}\left(\mathbf{0},\left\{\sum_{i=1}^{K}c_iJ_i(\boldsymbol{\theta})^TI_iJ_i(\boldsymbol{\theta})\right\}^{-1}\right). \tag{10}$$

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

$$\hat{\Sigma}_{IPD}=\left\{-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\log L(\hat{\boldsymbol{\theta}}_{IPD})\right\}^{-1}. \tag{11}$$

The statements (7) and (10) show that the estimators $\hat{\boldsymbol{\theta}}_{CD}$ and $\hat{\boldsymbol{\theta}}_{IPD}$ have the same limiting covariance matrix. Hence, we have established the claim that analyzing summary statistics using our CD approach incurs no loss of efficiency in comparison to analyzing individual-level data using the IPD approach.

Theorem 2. *Under the assumptions of Theorem 1, the CD estimator* $\hat{\boldsymbol{\theta}}_{CD}$ *is asymptotically as efficient as the IPD estimator* $\hat{\boldsymbol{\theta}}_{IPD}$*.*

Conceptually, the fact that our approach achieves full efficiency can be attributed to the following two features which are built in: First, our approach takes advantage of the reparameterization of the problem setting. This reparameterization connects each study-specific parameter $\boldsymbol{\gamma}_i$ to the common full parameter $\boldsymbol{\theta}$ using the transformation function $\boldsymbol{M}_i$. It essentially pools the information collected from all individual studies including those in which only some lower-dimensional functions (e.g., linear combination or ratio) of the full parameter are estimable (such as the cases in Example 2-3). Thus the reparameterization provides a mechanism for incorporating indirect evidence in $\boldsymbol{\gamma}_i$, in order to improve the overall inference for $\boldsymbol{\theta}$. Second, each individual confidence density function in (4) carries information on the correlation between the estimates of component parameters. Such within-study correlation enables borrowing of

strength across the studies for the synthesis of evidence. This point will be elaborated further in the next subsection. It is important to note that this kind of borrowing of strength is possible only if at least some individual studies share some common parameters. In this case, by multiplying the study-specific confidence densities, the overall maximization in (6) "shrinks" the study-specific estimates towards the parameters that are common to all the studies. Such shrinkage can also be viewed as another form of borrowing of strength.

The confidence density function in (4) may be treated as an approximation of the likelihood function $L_i^*(\boldsymbol{\gamma}_i)$, after being normalized by the constant $\int L_i^*(\boldsymbol{\gamma}_i)d\boldsymbol{\gamma}_i$ (which only depends on samples but not on the parameters). In this sense, our approach by combining confidence density functions in (5) can be viewed as an approximate likelihood approach, in which we construct and make inference based on some approximate functions of the true likelihood function. However, the CD framework has been shown to be very broad (Xie and Singh, 2013). In particular, the CD formulation can encompass a broad range of functions, including p-value functions, bootstrap distribution functions, normalized likelihood functions and even some Bayesian posteriors and priors. Thus, the applicability of the CD combining approach is far greater than that of the likelihood approach.

The advantage of combining confidence density functions not only gains full efficiency without having to resort to individual-level data, but it also gives rise to a unifying approach for combining summary statistics. It covers many existing meta-analysis approaches as special cases. For example, when $\boldsymbol{M}_i$'s are linear functions of $\boldsymbol{\theta}$, say, $\boldsymbol{M}_i(\boldsymbol{\theta}) = B_i^T\boldsymbol{\theta}$, our approach yields the same estimator as the one from the multivariate generalized least squares approach of Becker and Wu (2007). Here, B_i is $p_i \times p$ deterministic matrix, not depending on $\boldsymbol{\theta}$, $i = 1,\ldots,K$. In this particular case, our estimator $\hat{\boldsymbol{\theta}}_{CD}$ in (6) has the following explicit solution,

$$\hat{\boldsymbol{\theta}}_{CD} = \left(\sum_{i=1}^{K} B_i^T\hat{\Sigma}_i^{-1}B_i\right)^{-1}\left(\sum_{i=1}^{K} B_i^T\hat{\Sigma}_i^{-1}\hat{\boldsymbol{\gamma}}_i\right). \tag{12}$$

Furthermore, if $\boldsymbol{M}_i$ is the identity transformation (i.e., B_i is the $p \times p$ identity matrix) for every i, (12) reduces to $\hat{\boldsymbol{\theta}}_{CD} = \left(\sum_{i=1}^{K}\hat{\Sigma}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{K}\hat{\Sigma}_i^{-1}\hat{\boldsymbol{\gamma}}_i\right)$ which is exactly the same estimator derived from the usual approach of weighting point estimates; see, e.g., Lin and Zeng (2010). Note that our approach is applicable even when $\boldsymbol{M}_i$'s assume more complex functions, and incorporate a broader scope of indirect evidence to achieve efficiency gain, as shown in the next subsection.

2.3 Gain of efficiency from utilizing indirect evidence

In this section, we assess the relative efficiency of our approach versus the conventional approach and show that the utilization of indirect evidence in our approach can gain asymptotic efficiency via correlation. Without loss of generality, we consider a setting where $\boldsymbol{\gamma}_i = (\alpha_i, \boldsymbol{\beta})$, for $i = 1, \ldots, K$. Here, α_i's are the study-specific parameters and $\boldsymbol{\beta}$ is the common vector parameter. This setting is used widely in parametric and semiparametric models (see the examples in Simmonds and Higgins (2007) and Lin and Zeng (2010)). It also covers Model (1) in Introduction. Therefore, the results in this section apply readily to Examples 1-3.

In what follows we assume that the study-specific parameters α_i's are of interest, such as the case in Example 1. Using the IPD estimator as the benchmark for comparisons, the corollary below shows that the CD estimator $\hat{\alpha}_{i,CD}$ is asymptotically more efficient than the study-specific estimator $\hat{\alpha}_i$.

Corollary 1. *Consider the following three estimators for α_i's: $\hat{\alpha}_{i,CD}$ be the CD estimator, $\hat{\alpha}_{i,IPD}$ the IPD estimator, and $\hat{\alpha}_i$ the study-specific estimator from Study i. Let "aVar" stand for asymptotic variance. Then, we have, for each i, $i = 1, \ldots, K$,*

$$aVar(\hat{\alpha}_{i,IPD}) = aVar(\hat{\alpha}_{i,CD}) \leq aVar(\hat{\alpha}_i). \tag{13}$$

Corollary 1 shows that there is efficiency gain in the CD approach estimator $\hat{\alpha}_{i,CD}$ over the study-specific estimator $\hat{\alpha}_i$. It implies that other studies (Study j, for all $j \neq i$) can contribute to the estimation of the study-specific parameter α_i. At first, this may seem counterintuitive considering that the j-th study does not involve the parameter α_i and it is completely independent of the i-th study. However, it is important to realize that the i-th and j-th studies share a common parameter $\boldsymbol{\beta}$, and that the information for α_i and $\boldsymbol{\beta}$ is often correlated within the i-th study. When the estimation of $\boldsymbol{\beta}$ is improved from combining multiple studies, the estimation of α_i in turn is also improved through this correlation. Thus, the common parameter $\boldsymbol{\beta}$ serves as a catalyst that enables borrowing information from other studies for the estimation of the study-specific parameter α_i in Study i. This phenomenon of borrowing strength from indirect evidence is not yet well appreciated in conventional meta-analysis, and the study-specific estimators $\hat{\alpha}_i$'s are generally reported as the final estimators. Corollary 1 also shows that the CD estimator $\hat{\alpha}_{i,CD}$ is asymptotically as efficient as the IPD estimator $\hat{\alpha}_{i,IPD}$, which clearly implies that the CD approach utilizes fully the correlation information in the summary statistics for α_i's and $\boldsymbol{\beta}$. Both simulation and real data analysis in Sections 4 and 5 show that numerically $\hat{\alpha}_{i,CD}$ and $\hat{\alpha}_{i,IPD}$ are very close.

Now, we consider the estimation of a common scalar parameter $\eta = g(\boldsymbol{\beta})$, where g is a scalar function of the common parameter vector $\boldsymbol{\beta}$. For instance, in Model (1), the interaction effect $\eta = \beta_3$ is often of primary interest. The conventional approach combines the study-specific estimators $\hat{\eta}_i = g(\hat{\boldsymbol{\beta}}_i)$ using $w_i = 1/\widehat{\text{aVar}}(\hat{\eta}_i)$ as the weights, provided that the estimate $\hat{\eta}_i$ is available from the i-th study. More precisely, the conventional estimator $\hat{\eta}_{cvt} \equiv \sum w_i \hat{\eta}_i / \sum w_i$. The corollary below compares $\hat{\eta}_{cvt}$ with our estimator $\hat{\eta}_{CD} (\equiv g(\hat{\boldsymbol{\beta}}_{CD}))$, again using the IPD estimator $\hat{\eta}_{IPD} (\equiv g(\hat{\boldsymbol{\beta}}_{IPD}))$ as the benchmark.

Corollary 2.

$$aVar(\hat{\eta}_{IPD}) = aVar(\hat{\eta}_{CD}) \leq aVar(\hat{\eta}_{cvt}). \tag{14}$$

Corollary 2 shows the gain of efficiency of $\hat{\eta}_{CD}$ over $\hat{\eta}_{cvt}$. The efficiency loss by using $\hat{\eta}_{cvt}$ can occur in the situation where $\hat{\eta}_i$'s are available from all the studies, as also reported in Simmonds and Higgins (2007) and Lin and Zeng (2010). Obviously, the efficiency loss of the conventional approach can be exacerbated when $\hat{\eta}_i$'s are not available from some of the studies, such as the cases in Examples 2 and 3. Once again, Corollary 2 shows that our proposed estimator $\hat{\eta}_{CD}$ is asymptotically as efficient as the IPD estimator $\hat{\eta}_{IPD}$, and implies that our approach utilizes fully the correlation information. The result in Corollary 2 will also be demonstrated in the simulation and real data analyses.

3 Robustness against misspecification of covariance structure

In this section, we show that the CD approach is robust against misspecification of the covariance structure of the parameter estimates. Specifically, the CD estimator of $\boldsymbol{\theta}$ remains consistent and asymptotically normally distributed with appropriately adjusted limiting covariance matrix. This robustness property greatly enhances the applicability of the approach, even in the scenario that only the estimates of the variances, rather than the full covariance matrices, of $\hat{\boldsymbol{\gamma}}_i$ are reported.

Let $\Sigma_{i,W}$ denote a "working" covariance matrix of $\hat{\boldsymbol{\gamma}}_i$ in the i-th study, in the same sense of Liang and Zeger (1986). In this section, we use $(\hat{\boldsymbol{\gamma}}_i, \Sigma_{i,W})$ in place of $(\hat{\boldsymbol{\gamma}}_i, \hat{\Sigma}_i)$ as summary statistics and then show that the CD approach remains valid. Specifically, we denote by $\hat{\boldsymbol{\theta}}_W$ the new estimator obtained from (6) after replacing $\hat{\Sigma}_i$ in (5) with $\Sigma_{i,W}$. The next theorem shows that $\hat{\boldsymbol{\theta}}_W$ is consistent and asymptotically normally distributed with a "sandwich" covariance matrix.

Theorem 3. *Under the assumptions of Theorem 1 and the assumption that* $(n_i \Sigma_{i,W})^{-1} \to A_i$ *in probability as* $n_i \to \infty$*, where* A_i *is a positive definite matrix, the estimator* $\hat{\boldsymbol{\theta}}_W$ *is consistent and asymptotically*

normally distributed:

$$n^{1/2}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{MN}(\mathbf{0}, \Delta), \tag{15}$$

where

$$\Delta = \left\{\sum_{i=1}^{K} c_i J_i(\boldsymbol{\theta})^T A_i J_i(\boldsymbol{\theta})\right\}^{-1} \left\{\sum_{i=1}^{K} c_i J_i(\boldsymbol{\theta})^T A_i I_i^{-1} A_i J_i(\boldsymbol{\theta})\right\} \left\{\sum_{i=1}^{K} c_i J_i(\boldsymbol{\theta})^T A_i J_i(\boldsymbol{\theta})\right\}^{-1}. \tag{16}$$

The covariance of $\hat{\boldsymbol{\theta}}_W$ given in Theorem 3 can be estimated by

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

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

The proof of Theorem 3 is straightforward by noting that the estimating equation as in (21) in Appendix A.1 now changes to

$$\sum_{i=1}^{K} \frac{\partial \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{K} J_i(\boldsymbol{\theta})^T \Sigma_{i,W}^{-1} (\hat{\boldsymbol{\gamma}}_i - \boldsymbol{M}_i(\boldsymbol{\theta})) = \mathbf{0}. \tag{17}$$

Since $\hat{\boldsymbol{\gamma}}_i$ is an asymptotically unbiased estimator of $\boldsymbol{M}_i(\boldsymbol{\theta})$, the solution to the above estimating equation holds consistency and asymptotic normality even if $\Sigma_{i,W}$ is not the true covariance structure of $\hat{\boldsymbol{\gamma}}_i$. This robustness in principle follows the asymptotic theory for generalized estimating equations (GEE); that is, the solution to a GEE remains consistent and asymptotically normal even if the second moment (covariance structure) of the response is misspecified, as long as the first moment (mean) of the response is specified correctly (see, e.g., Liang and Zeger, 1986; Fahrmeir and Tutz, 2001, pp 119-129; Xie and Yang, 2003). But, different from Liang and Zeger (1986) where $K \to \infty$ and n_i are fixed, the result in Theorem 3 is established when $n_i \to \infty$ and K is fixed, which is also a setting considered in Xie and Yang (2003). Similar to what is observed in the GEE approach, although the specification of $\Sigma_{i,W}$ does not alter the consistency, it does affect the efficiency of the estimator $\hat{\boldsymbol{\theta}}_W$. Not surprisingly, $\hat{\boldsymbol{\theta}}_W$ achieves higher efficiency when $\Sigma_{i,W}$ is closer to $\hat{\Sigma}_i$. If $\Sigma_{i,W} = \hat{\Sigma}_i$, then $\hat{\boldsymbol{\theta}}_W = \hat{\boldsymbol{\theta}}_{CD}$, and it is again asymptotically as efficient as the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$.

The robustness of our approach has important ramifications. For instance, scientific or media publications typically report only the estimates of the variances, but not the full covariance matrix, for the parameter estimates. Theorem 3 justifies the application of our approach in this situation, and moreover, provides directions on how to choose suitable working correlation structure to achieve greater efficiency. To elaborate this point, we let the working covariance matrix of $\hat{\boldsymbol{\gamma}}_i$ be

$$\Sigma_{i,W} = \hat{V}_i^{1/2} R_{i,W}(\phi) \hat{V}_i^{1/2}, \tag{18}$$

where $\hat{V}_i = \text{diag}\{\hat{\Sigma}_i\}$ is a $p_i \times p_i$ diagonal matrix whose diagonal entries are the estimates of the variances of each component of $\hat{\boldsymbol{\gamma}}_i$, and $R_{i,W}(\phi)$ is a working correlation matrix fully characterized by a (possibly vector-valued) parameter ϕ. This specification of working covariance structures is similar to that proposed for longitudinal data analysis in Liang and Zeger (1986). Theorem 3 guarantees that the estimator $\hat{\boldsymbol{\theta}}_W$ derived using $\Sigma_{i,W}$ in (18) is consistent with *any* choice of $R_{i,W}(\hat{\phi})$, provided that $\hat{\phi}$ is a consistent estimator of ϕ given $\boldsymbol{\theta}$. In practice, the working correlation matrix $R_{i,W}(\phi)$ in (18) should be specified weighing the compromise between simplicity and loss of efficiency due to incorrect specification. For example, we can set $R_{i,W}(\phi) = R_{i,0}$, for any given correlation matrix $R_{i,0}$. The simplest choice of $R_{i,0}$ is the identity matrix (i.e., simply using the working independence assumption), and in this case $\Sigma_{i,W} = \hat{V}_i$. We elaborate later in our simulation analysis that the study designs, generally are reported, can provide useful guidelines for choosing $R_{i,0}$.

We emphasize that the robustness property enables our approach to use all studies, with both direct and indirect evidence, even if the correlation information is not reported with the summary statistics. Remarkably, *even naïvely using the working independence correlation structure, our approach still gains efficiency via the incorporation of indirect evidence*. To illustrate this point, we consider a toy example of combining two studies: Study 1 reports summary statistics for $\boldsymbol{\gamma}_1 = (\alpha, \beta)$, i.e., the point estimates $\hat{\alpha}$ and $\hat{\beta}$ and the corresponding variance estimates $\hat{\sigma}^2_\alpha$ and $\hat{\sigma}^2_\beta$, but without correlation estimate. Study 2 reports summary statistics for $\gamma_2 = \alpha + \beta$, i.e., the point estimate $\hat{\gamma}_2$ and the corresponding variance estimate $\hat{\sigma}^2_{\gamma_2}$. If β is of primary interest, the conventional method simply uses $\hat{\beta}$ from Study 1 for inference, ignoring Study 2 since it does not provide an estimate for β and thus plays the role of indirect evidence. Our proposed method, under the working independence correlation structure, makes full use of both studies and yields the combined estimate $\hat{\beta}_{CD} = \left\{(\hat{\sigma}^2_\alpha + \hat{\sigma}^2_{\gamma_2})\hat{\beta} + \hat{\sigma}^2_\beta(\hat{\gamma}_2 - \hat{\alpha})\right\}/(\hat{\sigma}^2_\alpha + \hat{\sigma}^2_\beta + \hat{\sigma}^2_{\gamma_2})$. This shows clearly that the indirect evidence (Study 2) is integrated into our inference, unlike the case for conventional meta-analysis which simply uses $\hat{\beta}$. In fact, even if $\hat{\alpha}$ and $\hat{\beta}$ are indeed independent, $Var(\hat{\beta}_{CD}) \leq Var(\hat{\beta})$ still holds. This toy example shows that in the absence of knowledge of correlation among the parameter estimates, the CD approach is still capable of extracting useful information from studies of indirect evidence for overall inference. Similar but more complex examples can be found in the analysis of the aircraft landing data set in Section 5. Intuitively speaking, borrowing strength from indirect evidence is achieved by reparameterization and shrinkage effect, as elaborated in Section 2.2.

Finally, we remark that, if $\Sigma_{i,W}$ is not equivalent to $\hat{\Sigma}_i$ asymptotically, the density function $h_i(\boldsymbol{\gamma}_i)$ in

(4) with $\Sigma_{i,W}$ replacing $\hat{\Sigma}_i$ is no longer a confidence density for the parameter $\boldsymbol{\gamma}_i$ asymptotically. This observation does not invalidate our approach. To the contrary, the development in this section highlights the great flexibility and broad applicability of our approach.

4 Simulation studies

We conduct simulation studies to numerically examine the theoretical results established in Sections 2 and 3. We mimic meta-analysis of randomized clinical trials, and simulate $K = 3$ independent studies using Model (1) in Introduction. For the i-th study, the treatment indicator X_{ij} is 1 or 0 with 0.5 probability, the covariate Z_{ij} (e.g., drug dosage) has three levels of 1, 2 and 5, and each level is assigned to $n_i/3$ subjects. In this case, let D_i denote the the i-th study's design matrix formed by stacking the four regression terms $(1,\ X_{ij},\ Z_{ij},\ Z_{ij}X_{ij})$, then $n_i(D_i'D_i)^{-1}$ converges to (after normalization)

$$\begin{pmatrix} 1 & -0.71 & -0.84 & 0.60 \\ & 1 & 0.60 & -0.84 \\ & & 1 & -0.71 \\ & & & 1 \end{pmatrix},$$

which indicates moderate to mildly strong correlation between the regression parameter estimates. To examine the impact of heterogenous studies, we modify the design of Study 1 to make it different from the other two studies. Then, we analyze the three studies using the conventional meta-analysis method, the IPD method and the CD method. The results are based on 1000 replicates when $n_1 = n_2 = n_3 = 150$ and the parameters $\alpha_1 = -1, \alpha_2 = 0, \alpha_3 = 1, \beta_1 = 1, \beta_2 = 2, \beta_3 = -1$, $\sigma_1 = 3, \sigma_2 = 4$ and $\sigma_3 = 3$.

For the first part of the simulation study, Study 1 follows a missing covariate design as described in Example 2. Specifically, the covariate Z_{1j} is set at a fixed level, say, $Z_{1j} \equiv 1$ for all $j = 1, \ldots, n_1$. This mimics the situation where a clinical trial is designed to examine solely the treatment effect by controlling the variable Z_{1j} to eliminate the covariate effect. In this case, as shown in Example 1, the model for Study 1 reduces to Model (2). The estimable parameter $\boldsymbol{\gamma}_1 = (\alpha_1 + \beta_1, \beta_1 + \beta_3)^T$, and $\boldsymbol{\gamma}_1$ can be linked to $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3)^T$ through a self-explanatory function $\boldsymbol{M}_1$. The setting clearly fits the general likelihood inference setting in Section 2.1, and Study 1 can be included in the analysis using the CD approach. The analysis results are presented in Table 1.

Table 1 shows that the CD approach enables the estimation of α_1 which is not estimable in Study 1

alone. For all the regression parameters, our estimators are nearly unbiased, and the estimated standard errors are almost identical to the standard errors calculated from the estimates from 1000 replications. In comparison with the IPD method, our method yields virtually identical results in terms of point estimation and standard error estimation. This shows that the CD method and the IPD method have very similar numerical performance even for moderate sample sizes. Table 1 also shows clearly that the conventional method is deficient in the following two regards. First, it can not analyze Study 1 due to the design heterogeneity, and thus fails to estimate the associated study-specific parameter α_1. Second, its standard errors of the estimates for the study-specific parameters α_2 and α_3 and the common parameter β_1 are considerably larger than those obtained from the CD or IPD method. This is a great loss of efficiency resulting from the failure of fully utilizing Study 1 (as indirect evidence) and the correlation information.

For the second part of the simulation study, we set the responses in Study 1 to be dichotomized as described in Example 3. Specifically, we create binary responses d_{1j} in such a way that $d_{1j} = 1$ if the observation $y_{1j} \geq 4$ and $d_{1j} = 0$ otherwise, for $j = 1, \ldots, n_1$. Then, we discard all original continuous responses y_{1j} and only keep the binary responses d_{1j} for analysis. This mimics the situation where a clinical center routinely reports "censored" outcomes instead of the original outcomes. In this case, as shown in Example 3, the model for Study 1 reduces to Model (3). The estimable parameter $\boldsymbol{\gamma}_1 = ((\alpha_1 - 4)/\sigma_1, \beta_1/\sigma_1, \beta_2/\sigma_1, \beta_3/\sigma_1)^T$, and $\boldsymbol{\gamma}_1$ can be linked to $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \sigma_1)^T$ through the function $\boldsymbol{M}_1$. Note that $\boldsymbol{M}_1$ is a non-linear function in this case. Again, this setting fits the general likelihood inference setting in Section 2.1, and Study 1 can be included in the analysis using our CD approach. The analysis results are presented in Table 2.

Table 2 shows that the CD approach yields estimates for all the regression parameters, including α_1, and their numerical values are quite close to those from the IPD method. The conventional method can not utilize Study 1 due to the heterogeneity in outcome, and is thus unable to estimate the study-specific parameter α_1. Moreover, it suffers substantial loss of efficiency in estimating some parameters, such as α_2, α_3 and β_1. Detailed comparison among the three methods can be summarized in conclusions similar to those of the first simulation study. We also include in Table 2 the analysis results obtained from dichotomizing the continuous responses in Study 2 and Study 3. As discussed in Dominici and Parmigiani (2000), when combined inference is desired in the mix of continuous and dichotomous responses, a common practice is to dichotomize all continuous responses and then proceed as if all responses were binary. However, Table 2 shows that this dichotomization method yields worse estimates than the CD or

IPD method. It shows that dichotomization can lead to substantial loss of information in inference.

The last part of the simulation study concerns the setting where we have only the variances estimates, but not the full covariance matrices estimates, of the estimates of parameters. For convenience, we use the same setting as the first part of the simulation study but assuming the availability of only $\hat{V}_i = \text{diag}\{\hat{\Sigma}_i\}$ instead of $\hat{\Sigma}_i$. To implement our approach, we consider three approaches for specifying working correlation matrices, with a varying degree of implementation complexity: 1) the independence method, i.e., using identity matrices as the working correlation matrices; 2) the design-driven method, i.e., using our "best guess" to set working correlation matrices based on the knowledge of the study designs; 3) the data-driven method, i.e., using correlation matrices estimated from the observed data. More concretely, the design-driven approach is implemented as follows. For Study 1 with Model (2), we set two $n_1 \times 1$ vectors $\mathbf{1}_0 = (1,\ldots,1)^T$ and $\boldsymbol{x}_{1,0} = (0,1,\ldots,0,1)^T$. Then we let the off-diagonal entries of the 2×2 working correlation matrix $R_{1,0}$ be $r_{12} = r_{21} = \langle \mathbf{1}_0^{\perp}, \boldsymbol{x}_{1,0}^{\perp}\rangle / (\|\mathbf{1}_0^{\perp}\|\|\boldsymbol{x}_{1,0}^{\perp}\|)$. Here, $\langle\cdot,\cdot\rangle$ denotes the inner product, $\|\cdot\|$ denotes the Euclidean norm, and $\mathbf{1}_0^{\perp} = \mathbf{1}_0 - P(\mathbf{1}_0|\mathscr{L}(\boldsymbol{x}_{1,0}))$ is the residual after projecting $\mathbf{1}_0$ to the linear space $\mathscr{L}(\boldsymbol{x}_{1,0})$ spanned by $\boldsymbol{x}_{1,0}$. Note that $R_{1,0}$ is the true correlation matrix for $\hat{\boldsymbol{\gamma}}_1$ if $(\mathbf{1}_0, \boldsymbol{x}_{1,0})$ is the actual design matrix for Model (2). For Study 2 and Study 3 with Model (1), the working correlation matrices $R_{2,0}$ and $R_{3,0}$ can be specified in a similar way. It is worth noting that this specification approach is based on the prefixed design of randomized clinical trials, not on the data observed after the experiment. Thus, $R_{i,0}$'s are considered as fixed. The data-driven approach is implemented only when there are multiple studies that have the same correlation structure. For such an approach, we generate 10 independent copies for each of Study 1, 2 and 3, which yields $K = 3\times 10 = 30$ studies in total. Although it is rarely the case in practice to have 10 independent studies with the same correlation structure, we present the result here for two purposes: 1) to provide an understanding of how much efficiency our approach can retain assuming such covariance estimation were possible; and 2) to provide a numerical comparison with the other two cases. The analysis results for all three approaches are presented in Table 3.

Table 3 shows the bias, standard error (SE) and efficiency for $\hat{\boldsymbol{\theta}}_W$ obtained using working correlation matrices based on the independence approach ($K = 3$), design-driven approach ($K = 3$) and data-driven approach ($K = 3\times 10$), respectively. Note that all approaches yield estimates for *all* the regression parameters, including α_1 that is not estimable in Study 1. Moreover, $\hat{\boldsymbol{\theta}}_W$ is unbiased regardless of the approach used. This numerically confirms the theoretical finding in Section 3 that the specification of working correlation matrices does not alter the consistency of the CD estimator. Note that, the SEs reported in the

data-driven approach are much smaller, due to the 10 folder increase of sample sizes in this particular simulation setting. A meaningful measure for effectiveness is the relative efficiency of $\hat{\boldsymbol{\theta}}_W$ relative to the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$. Obviously, the independence method can lead to noticeable loss of efficiency in estimating certain parameters, such as α_2, α_3 and β_3, because the working zero correlation prevents the full use of the information in Study 1. On the other hand, both the design-driven and data-driven approaches are highly efficient. In particular, the design-driven approach essentially suffers no efficiency loss. This implies that, for the self-designed studies, such as randomized clinical trials, the working correlation matrices derived from the design-driven approach are remarkably efficient, and thus the design-driven approach should be recommended.

5 Real data example: aircraft landing data

The Federal Aviation Administration (FAA) is the oversight agency responsible for regulating air traffic and safety. The rapid growth in air traffic density has led the FAA to initiate many new research efforts in aviation safety. In particular, the trend of increasing runway incidents has prompted FAA to expand its research on aircraft landing performance and possibly set new advisory directive on aircraft landing operations. It is reported (see, e.g., Van Es (2005)) that the most frequently reported aircraft landing incidents are *runway overruns*, meaning that landing aircraft are unable to stop before the end of the runway. It is also concluded in Van Es (2005) that there is a significant increase in overrun risk when an aircraft has long *landing distance*. The landing distance here refers to the distance from the beginning of the runway to the aircraft touchdown point. Hence, it is of vital importance to examine how landing distance is affected by aircraft air-borne performance.

Our project, though with a broader objective, has a specific task to model and analyze the impact of aircraft air-borne performance measures on landing distance on a standard runway for all passenger flights. From the observed sample flights, summary statistics are reported by studies according to aircraft make-and-model. For simplicity, we illustrate the application of our CD approach to this project using only two studies. The first is of $n_1 = 6565$ flights carried out by Airbus 321, and the second of $n_2 = 15809$ flights by Boeing 737. The observation of each flight is supposed to contain various air-borne performance measures and landing distance.

Intuitively, one would want to model landing distance based on air-borne performance measures for

each individual study to obtain, presumably, the best model for each aircraft model-and-make. However, the purpose of our project is to provide the FAA a model that is broad enough to encompass all passenger flights, so that the FAA can, in accordance with policy issuances, issue a single advisory directive with general landing performance guidelines for all flights. This is the reason why we need to combine studies from different aircraft. More discussion on whether or when heterogeneous studies should be combined is given in Section 6.

To combine the studies from Airbus 321 and Boeing 737, we consider the following linear model:

$$Y_{ij} = \alpha_i + \beta_1 X_{1ij} + \cdots + \beta_5 X_{5ij} + \beta_6 X_{6ij} + \beta_7 X_{5ij} X_{6ij} + \varepsilon_{ij}, \quad i = 1,2,\ j = 1,\ldots,n_i. \tag{19}$$

Here Y_{ij} is the logarithm of landing distance for the j-th observation in the i-th data set ($i = 1$ for Airbus and $i = 2$ for Boeing), $X_{1ij},\ldots,X_{6ij}$ are different air-borne performance measures (e.g., height, air-speed, flaps, ...etc.) when the aircraft passes the runway threshold, and ε_{ij} the noise variable with $\varepsilon_{ij} \sim N(0,\sigma_i^2)$. The six air-borne performance measures included in Model (19) are selected using statistical variable selection procedures and also independently confirmed by subject matter experts.

For our project, we were initially given only the summary statistics from the two studies for Airbus 321 and Boeing 737, as presented in Part 1 of Table 4. Generally, individual flight data are not released due to confidentiality pertaining to airlines, aircraft manufacturers or any specific flights, among other concerns. Only after we had reported the findings from our meta-analysis based on the summary statistics, our request for individual flight data was granted. Even then, these individual data were permitted for the sole purpose of comparing the findings from summary statistics and IPD method. Our success in obtaining such individual-level data should be viewed more of an exception rather than the norm in practice when involving proprietary studies. This again highlights the usefulness of our CD approach in retaining the efficiency without requiring individual data.

Part 1 of Landing Data Analysis – Reflecting the difference in aircraft design, the measure "flaps" (the regressor X_6) for the Airbus data is set at a constant level by design, namely $X_{61j} \equiv 24$ for all j's in Model (19). Thus, this study has a *missing covariate design*, with the estimable parameter $\boldsymbol{\gamma}_1 = (\alpha_1 + 24\beta_6, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 + 24\beta_7)^T$. The situation is similar to that of Example 2 in Introduction.

Under the setting above, Part 1 of Table 4 shows the individual analysis results for all the regression coefficients in Model (19). Clearly, Airbus Study by itself does not provide estimates for the study-specific parameter α_1 and the common slope parameters β_5, β_6 and β_7, due to the missing covariate design. Therefore, the conventional method can not utilize Airbus Study to improve the inference for those parameters,

as seen in Part 1 of Table 5. On the other hand, the column labeled "CD" in Part 1 of Table 5 shows the analysis results obtained from our CD method, given the point estimates of the estimable parameters and corresponding covariance estimates. Clearly, our CD method provides estimates for all the parameters, including those that are not estimable in Airbus Study. Moreover, in comparison with the IPD method, our method yields almost identical point estimates and standard error estimates. This confirms once again that our CD method and the IPD method behave similarly in large sample settings. Note that the column labeled "Robust" in Table 5 are the analysis results obtained from our CD method using only variances estimates. Here we use the independence approach to generate working correlation matrices. Even with this simple approach, Table 5 shows that our CD method is still able to utilize both studies and provide meaningful estimates for all the parameters. We note that for the common parameters $\beta_1, \ldots, \beta_4$, although our robust method uses working independence assumption, there are negligible differences between the results from our CD method and the IPD method. The underlying reason is that the corresponding regressors $X_1, \ldots, X_4$ are only mildly correlated to others. But for $\alpha_1, \alpha_2, \beta_5, \ldots, \beta_7$, the differences are more noticeable because these regressors have stronger correlation with others.

Part 2 of Landing Data Analysis – As an illustrative example, we also dichotomize the response variable in the Boeing data set to mimic a possible scenario that only binary evaluation outcome of the landing distance is reported. In other words, instead of the actual landing distance, the landing distance is only coded as long landing-distance (indicating that the landing distance is longer than a prefixed safety standard) or not. Specifically, we define binary responses d_{2j} by letting $d_{2j} = 1$ if $y_{2j} \geq \tau$ and $d_{2j} = 0$ otherwise, where τ is a standard threshold used in aviation analysis to assess the achievable performance of most flights. One example is the 90% quantile of all acceptable landing distances from commercial flights. The observation of $d_{2j} = 1$ indicates *long landing distance* of a flight and is indicative of higher risk. Assume that only the binary responses d_{2j}'s are available in Boeing Study (in place of Y_{2j}'s). This setting is similar to Example 3 in Introduction. The estimable parameter in this case is $\boldsymbol{\gamma}_2 = ((\alpha_2 - \tau_2)/\sigma_2, \beta_1/\sigma_2, \ldots, \beta_7/\sigma_2)^T$. The analysis results under this particular setting are reported in Part 2 of Table 4 and Table 5.

For individual studies, Part 2 of Table 4 shows that most of the slope parameters in Model (19) are not estimable, due to the fact that the Airbus Study has missing covariate design and the Boeing Study has dichotomous responses. In this situation, the conventional meta-analysis method is not able to combine any information at all, as seen in Part 2 of Table 5. On the other hand, Part 2 of Table 5 shows that our

CD method still provides estimates for all the intercept and slope parameters in Model (19), including those that are not estimable in either study. For comparison purpose, we also provide the IPD estimates in Table 5. Clearly, both the point estimates and the corresponding standard error estimates from our method are very close to those obtained from the IPD method. Moveover, given only variances estimates, our method is robust enough to still be able to combine the two studies and provide inference for all the parameters. Table 5 shows that the results obtained using working independence correlation matrices remain quite close to the IPD estimates. This observation shows further that, even in the absence of correlation information, our CD approach still benefits from incorporating indirect evidence in the analysis in gaining efficiency in inference. This is a clear improvement over the conventional method which discards indirect evidence.

It is worth noting that the data analysis in this section shows that when the "information" in each study (e.g., measured by $n_i J_i(\boldsymbol{\theta})^T I_i J_i(\boldsymbol{\theta})$) is large, our CD estimates are quite close to the IPD estimates. When such "information" is reduced, as seen in Part 2 of our analysis, the difference in numerical results based on IPD and summary statistics may become more appreciable.

6 Discussions

This paper proposes a general multivariate meta-analysis approach through combining confidence density functions. Although this proposed CD approach only requires summary statistics from relevant studies, it is shown to be asymptotically as efficient as the IPD approach which requires individual-level data from all studies. The CD approach is shown to be applicable to a broad range of heterogeneous studies. It also enables us to incorporate indirect evidence in the analysis and borrow strength (c.f. the discussion in Section 2.3) to achieve gains in efficiency in the overall inference. Furthermore, the CD approach is shown to have a robustness property which ensures that the approach remains valid even when the covariance estimates of the parameter estimates are misspecified. This robustness property substantially broadens the applicability of the CD approach. All those desirable properties of the CD approach are also confirmed in the stimulation studies and real data analysis in this paper.

The development in this paper has far-reaching practical implications on the issue of analyzing individual level IPD data versus summary statistics in meta-analysis. It is well known that analyzing IPD data from the original studies has many benefits. For example, by analyzing individual-level data, one can

achieve efficient inference, provided that the model is specified correctly. Moreover, by accessing the original data, one can enhance comparability among the studies with respect to inclusion/exclusion criteria, creation of subgroups, and adjustments of covariates, as discussed in Lin and Zeng (2010). Despite these obvious advantages, the majority of meta-analysis is not performed using the IPD approach, as observed in Sutton and Higgins (2008). One practical limitation in carrying out an IPD analysis is that, in most occasions, individual-level data from all the studies are not available. In fact, there are ongoing debates on whether the benefit of using the IPD method can outweigh the tremendous cost of retrieving IPD from all relevant studies (Sutton and Higgins, 2008). The development in this paper clearly shows that there is no need to retrieve IPD, because the aforementioned benefit of the IPD approach can all be retained simply by using the CD approach to analyze summary statistics. First of all, the CD approach of analyzing summary statistics has no asymptotic efficiency loss compared to the IPD approach, under the general likelihood inference framework considered in this paper. Note that this framework covers most of the commonly used fixed-effects models. Second, the CD approach adapts to a broad scope of heterogeneous studies, which implies that, given only summary statistics, we can still incorporate indirect evidence in the analysis. As a result, our meta-analysis can improve inference for subpopulations, adjustment for important covariates, and joint analysis of mixed outcomes. These useful properties are clearly demonstrated in our simulation and real data analysis.

There has been much work in meta-analysis on examining the relative efficiency of using summary statistics versus IPD. For some special settings, Olkin and Sampson (1998) and Mathew and Nordstrom (1999) showed that there is no efficiency loss by analyzing summary statistics. Lin and Zeng (2010) reached the same conclusion under a more general likelihood inference setting but the focus there is only on the parameter that is estimable across all studies. The development in our paper is more general. It does not require that the parameter $\boldsymbol{\theta}$ be estimable in all studies. This is yet another reason why our analysis can include a broad class of heterogenous studies. Also, the approach of combining summary statistics investigated in Lin and Zeng (2010) is essentially a linear weighting of point estimators. Our approach is fundamentally different, as it is to combine confidence density functions. It can process efficiently the summary statistics for $\boldsymbol{\gamma}_i = \boldsymbol{M}_i(\boldsymbol{\theta})$ (cf. Section 2) for any complex function $\boldsymbol{M}_i$ satisfying mild smoothness conditions, including the special case of $\boldsymbol{M}_i$ being the identity transformation of $\boldsymbol{\theta}$ for all studies which is exactly the setting considered in Section 2.1 of Lin and Zeng (2010).

This paper shows that the idea of combining confidence density functions not only yields a unified

treatment for combining summary statistics, but it also provides a new alternative to performing *complex evidence synthesis* by allowing the mapping functions $\boldsymbol{M}_i$'s to be any complex functions satisfying some mild smoothness conditions. Here, the phrase *complex evidence synthesis* refers to combining information from the models that "incorporate evidence on multiple parameters and/or that specifically model data from different study designs" (Sutton and Higgins, 2008). In recent years, complex evidence synthesis has received increasing attention, and one important development in this area has been on mixed treatment comparisons (also referred to as indirect treatment comparisons) in randomized clinical trials (see, e.g., Jansen et al., 2011; Hoaglin et al., 2011). Most of the approaches developed for complex evidence synthesis so far are within the Bayesian framework (see, e.g., Ades and Sutton, 2006). The proposed CD approach shares a similar spirit with the Bayesian approaches, in the sense that the CD approach also combines density functions but it achieves the same goal under the frequentist framework. On the other hand, unlike Bayesian procedures, our CD approach does not require specification of priors and/or Markov chain Monte Carlo procedures.

Although the confidence distributions used in this paper are all in the form of multivariate normal distributions, which suffice for the investigation in the asymptotic setting, the framework of combining confidence distributions can be much more general. For example, one may consider using multivariate t distributions (to account for small sample properties) or other general forms of confidence distributions.

The research on confidence distributions is part of the recent emerging developments on distributional inference. In general, the goal of distributional inference is to define a sample-dependent distribution on the parameter space that can provide meaningful answers to questions related to statistical inference. Beside confidence distributions, the developments also include generalized fiducial inference (e.g., Hannig, 2009; Hannig and Lee, 2009; Hannig, 2013), belief function and inferential models (e.g., Dempster, 2008; Martin et al., 2010; Martin and Liu, 2013) and objective Bayes methods (e.g., Berger, 2006; Berger and Sun, 2009). All these developments help shed light on the common perspective of the Bayesian and frequentist ideals.

Our CD approach has been demonstrated so far for fixed-effects where study-specific effects, representing between-study variation, are assumed fixed but unknown. One natural extension is to assume that study-specific effects are realizations of a random variable and to apply random-effects models to draw inference. Such a random effects meta-analysis (e.g., by hierarchical modeling) can potentially enable another layer of "borrowing strength" through between-study correlation. The CD approach also applies

to random-effects models as long as the between-study covariance Ω can be estimated. In this case, we replace the within-study covariance estimate $\hat{\Sigma}_i$ by the total covariance estimate $\hat{\Sigma}_i + \hat{\Omega}$ and then follow the same procedure to make inference.

We should point out that the term "heterogeneity" considered in this paper should not be confused with the same term often used in the literature on random-effects meta-analysis. In the latter, heterogeneity means that the effect of interest is different across studies due to heterogeneous populations, and the goal is to assess, test or model such heterogeneity, mostly in univariate cases. Our paper investigates heterogeneity with respect to a wide scope of models, covering cases of heterogeneous study designs, outcome types, and more generally, complex evidences.

It is important that the desire for combining evidence from a cohort of studies be justified by the assumption that at least part of the studies share a common effect, say β (although β is not required to be estimable in each study). Thus, an important step in meta-analysis is to assess the assumption that $\beta_i = \beta$ for all i. Such an assessment may use subject matter knowledge and statistical techniques. From the perspective of observational studies in epidemiology and/or randomized clinical trials, respectively, Stroup et al. (2000) and Berlin et al. (2013) suggested that subject matter knowledge should be taken as the first consideration in evaluating the "similarity" of the studies and determining which study should not be combined with others. Such knowledge includes population characteristics, experiment duration and other crucial information associated with the specific research goals. On the other hand, some statistical techniques have been developed for assessing the assumption that $\beta_i = \beta$ for all i when β_i is estimable (see Sutton and Higgins, 2008 pp. 628). But further research is needed when some of β_i's are not estimable.

Finally, we use our aircraft landing study project as an illustrative example to provide a few brief comments on the question of "to combine of not to combine" when facing heterogeneous studies.

We stress that the focus of this paper is on developing an efficient approach for combining heterogeneous studies after the combination is already deemed suitable and needed. It is difficult to formulate precise rules on how to decide whether a given set of heterogeneous studies should be combined. But the decision can be made more meaningful by carefully examining the technical assessments in each case and weighing them against the needs/suitability justified by domain experts. For instance, among other considerations, the decision of combining the studies in our aircraft landing project can be justified by the need of a common model for all aircraft to enable the FAA to issue a *single* advisory directive (AC) on aircraft landing safety for *all commercial passenger flights, regardless of aircraft types, makes or models.*

More specifically, the FAA, in its capacity as a regulatory and oversight agency for aviation safety, is required to issue a single AC that provides general guidelines for safe landing operations on airport runways for all commercial passenger jets (referred to as FAR Part 121). (More details about this and other FAA regulations can be found in http://www.faa.gov/regulations_policies/.) The combination of studies from Airbus and Boeing illustrated in Section 5 can be used to support such regulatory efforts.

When facing heterogeneous studies in meta-analysis, before undertaking the task of combining the studies, it is also important to evaluate the nature and degree of heterogeneity among those studies. This evaluation has crucial impact on the downstream analysis. For instance, one should avoid combining studies whose degree of heterogeneity is so great as to render useless the analysis outcome or incoherent its interpretation. A case in point would be NOT to include helicopter landing performance in our aircraft landing studies. Even though helicopters are aircraft, just like the planes Airbus 321 and Boeing 737, and their landing performance on landing pads (or helipads) is subject to a similar requirement of landing within certain preset boundary, just like that of the landing performance of the Airbus 321 and Boeing 737 on airport runways, the vertical landing of helicopters constitutes too great a degree of heterogeneity for a meaningful combined analysis. On the other hand, the two types of aircraft Airbus 321 and Boeing 737 share sufficiently many similar features, such as weight, function, designs, and flight purpose (i.e., scheduled commercial air transportation). In fact, the FAA AC requires that they operate within the same safe guidelines.

APPENDIX A: PROOFS

A.1 Proofs for the results in Section 2.2

This Appendix contains technical details for the theoretical results in Section 2.2. First, we prove the asymptotic property (10) for the IPD estimator. Second, we prove below Lemma 1 which implies the asymptotic equivalence between the IPD estimator and our CD estimator. The asymptotic properties of our CD estimator in Theorem 1 then follow.

Assume that the regularity conditions specified in Appendix B hold for the density function $f_i^*(x,y;\boldsymbol{\gamma}_i) \equiv f_i(x,y;\boldsymbol{\theta})$ in each study and that $\boldsymbol{\theta}$ is identifiable in the multiplied density function $\prod_{i=1}^K f_i(x_i,y_i;\boldsymbol{\theta})$. Then, the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$ is consistent. To proceed to prove (10) for $\hat{\boldsymbol{\theta}}_{IPD}$, we apply the Taylor expansion

and obtain

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\hat{\boldsymbol{\theta}}_{IPD}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) + \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log L(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{IPD} - \boldsymbol{\theta}) + O_p(1). \tag{20}$$

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

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

and

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

Plug the above results into Equation (20), and after some algebraic manipulations, we derive

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

Under the conditions specified in Section 2.1, we can conclude the asymptotic normality (10) for $\hat{\boldsymbol{\theta}}_{IPD}$.

Lemma 1. *The gradient of the log-confidence density function* $\log h_i(\boldsymbol{\gamma}_i) \equiv \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))$*, with respect to* $\boldsymbol{\theta}$*, is asymptotically equivalent to the score function* $\boldsymbol{s}_i(\boldsymbol{\theta}) = \partial \log L_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$*. More precisely,*

$$\frac{\partial \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = J_i(\boldsymbol{\theta})^T \hat{\Sigma}_i^{-1}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{M}_i(\boldsymbol{\theta})) = \boldsymbol{s}_i(\boldsymbol{\theta}) + O_p(1), \quad i = 1, \ldots, K. \tag{21}$$

Here, the $p_i \times p$ *matrix* $J_i(\boldsymbol{\theta}) = \partial \boldsymbol{M}_i(\boldsymbol{\theta}))/\partial \boldsymbol{\theta}$*.*

Proof. It is straightforward to obtain the first equation in (21) by differentiation. We only need to show the second equation in (21). Since $\boldsymbol{s}_i(\boldsymbol{\theta}) = \partial \log L_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = J_i(\boldsymbol{\theta})^T \{\partial \log L_i^*(\boldsymbol{\gamma}_i)/\partial \boldsymbol{\gamma}_i\}$, we have the Taylor series expansion of $\partial \log L_i^*(\boldsymbol{\gamma}_i)/\partial \boldsymbol{\gamma}_i$ around the consistent estimate $\hat{\boldsymbol{\gamma}}_i$ as follows

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

The second equation in (21) immediately follows. This completes the proof of Lemma 1. □

In light of Equation (21), our estimator $\hat{\boldsymbol{\theta}}_{CD}$ has the same asymptotic properties as the IPD estimator $\hat{\boldsymbol{\theta}}_{IPD}$. The statements in Theorem 1 are therefore implied by the established results.

A.2 Proofs for the results in Section 2.3

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

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

From (7), the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{CD}$ is

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

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

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

where D_1 and D_2 are $l \times l$ matrices. Under this setup, we prove Corollary 1 and Corollary 2 as follows.

First, we show $\text{aVar}(\hat{\alpha}_{i,CD}) \le \text{aVar}(\hat{\alpha}_i)$ as in Corollary 1 for $i = 1$. It is easy to see that $\text{aVar}(\hat{\alpha}_i) = a_1$, and

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

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

$$\left\{\left(c_1 J_1^T I_1 J_1 + c_2 J_2^T I_2 J_2\right)^{-1}\right\}_{[1,1]} \le a_1. \tag{22}$$

By Lemma 2 in Appendix A.4,

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

Using the results in Lemma 3 in Appendix A.4, we obtain

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

which leads to the establishment of (22). This completes the proof of Corollary 1.

Next, we show $\text{aVar}(\hat{\eta}_{CD}) \le \text{aVar}(\hat{\eta}_{cvt})$ as in Corollary 2 for $\eta = g(\boldsymbol{\beta})$. For simplicity, we assume $\eta = g(\boldsymbol{\beta}) = \boldsymbol{\lambda}^T \boldsymbol{\beta}$, where $\boldsymbol{\lambda}$ is a l-dimensional vector. By Lemma 2 in Appendix A.4,

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

Thus, $\text{aVar}(\hat{\eta}_{CD}) = \boldsymbol{\lambda}^T (D_1^{-1} + D_2^{-1})^{-1} \boldsymbol{\lambda}$. On the other hand, $\text{aVar}(\hat{\eta}_{cvt}) = \left\{(\boldsymbol{\lambda}^T D_1 \boldsymbol{\lambda})^{-1} + (\boldsymbol{\lambda}^T D_2 \boldsymbol{\lambda})^{-1}\right\}^{-1}$. It follows from Lemma 3 in Appendix A.4 that $\text{aVar}(\hat{\eta}_{CD}) \le \text{aVar}(\hat{\eta}_{cvt})$. This completes the proof of Corollary 2.

A.3 Proof for the result in Section 3

We provide a sketch of the proof for Theorem 3 and refer readers to Xie and Yang (2003) for technical details for a formal treatment.

Assuming the regularity conditions as stated in Appendix B, the consistency of $\hat{\boldsymbol{\theta}}_W$ is a direct result of the observation that $E\left\{\sum_{i=1}^K \partial \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))/\partial \boldsymbol{\theta}\right\}/n \to 0$ as $n \to \infty$. To show the asymptotic normality of $\hat{\boldsymbol{\theta}}_W$, we notice that

$$\sum_{i=1}^K \frac{\partial \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \sim \mathrm{MN}\left(\mathbf{0}, \left\{\sum_{i=1}^K J_i(\boldsymbol{\theta})^T \Sigma_{i,w}^{-1} \mathrm{Cov}(\hat{\boldsymbol{\gamma}}_i) \Sigma_{i,w}^{-1} J_i(\boldsymbol{\theta})\right\}\right),$$

based on (17). On the other hand, using Taylor expansion, we obtain

$$\sum_{i=1}^K \frac{\partial \log h_i(\boldsymbol{M}_i(\hat{\boldsymbol{\theta}}_w))}{\partial \boldsymbol{\theta}} = \sum_{i=1}^K \frac{\partial \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} + \sum_{i=1}^K \frac{\partial \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(\hat{\boldsymbol{\theta}}_w - \boldsymbol{\theta}) + O_p(1).$$

Therefore,

$$(\hat{\boldsymbol{\theta}}_w - \boldsymbol{\theta}) = \left\{\sum_{i=1}^K \frac{\partial \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right\}^{-1} \left\{-\sum_{i=1}^K \frac{\partial \log h_i(\boldsymbol{M}_i(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\right\} + O_p(1/n),$$

which leads to the establishment of (15).

A.4 Some useful matrix results

Lemma 2. *Under the setup in Appendix A.2, we have the following results:*

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

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

Proof. Using the blockwise matrix inversion formula, we have

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

and similarly,

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

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

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

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

Lemma 3. *Suppose W_1 and W_2 are $q \times q$ positive definite matrices. Then, for any q-dimensional vector $\boldsymbol{v}$,*

$$\left(\boldsymbol{v}^T W_1^{-1} \boldsymbol{v}\right)^{-1} + \left(\boldsymbol{v}^T W_2^{-1} \boldsymbol{v}\right)^{-1} \leq \left\{\boldsymbol{v}^T (W_1 + W_2)^{-1} \boldsymbol{v}\right\}^{-1}. \tag{23}$$

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

Proof. Since $W_1 > 0$ and $W_2 > 0$, we can find a nonsingular matrix P such that $W_1 = P\text{diag}\{r_1, \ldots, r_q\} P^T$ and $W_2 = P\text{diag}\{u_1, \ldots, u_q\} P^T$, where $r_i > 0$ and $u_i > 0$ for all $i = 1, \ldots, q$. By redefining $\boldsymbol{v}$ as $P^{-1}\boldsymbol{v}$, it suffices to prove the lemma when $W_1 = \text{diag}\{r_1, \ldots, r_q\}$ and $W_2 = \text{diag}\{u_1, \ldots, u_q\}$. Denoting $\boldsymbol{v} = (v_1, \ldots, v_q)^T$, the inequality (23) becomes

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

After rearrangement, the above inequality can be equivalently written as

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

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

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

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

APPENDIX B: REGULARITY CONDITIONS

Here we state the regularity conditions used throughout the paper. For each study, we assume that the density function $f_i^*(x, y; \boldsymbol{\gamma}_i)$ satisfies the following conditions. For notational convenience, we suppress

the study index i in $f_i^*(x,y;\boldsymbol{\gamma}_i)$ and write $f^*(x,y;\boldsymbol{\gamma})$ instead.

(a) The density function $f^*(\boldsymbol{\gamma})$ is identifiable, i.e., $f^*(\boldsymbol{\gamma}_1) \equiv f^*(\boldsymbol{\gamma}_2) \Rightarrow \gamma_1 = \gamma_2$.

(b) The density function $f^*(\boldsymbol{\gamma})$ has a common support, i.e., the set $S = \{(x,y) \mid f^*(x,y;\boldsymbol{\gamma}) > 0\}$ is independent of $\boldsymbol{\gamma}$.

(c) The parameter space Γ contains an open set Γ^O of which the true parameter value $\boldsymbol{\gamma}^0$ is an interior point, and $f^*(\boldsymbol{\gamma})$ admits all third derivatives for all $\boldsymbol{\gamma} \in \Gamma^O$.

(d) The first and second derivatives of the logarithm of f^* satisfy the equations $E_{\boldsymbol{\gamma}}\left\{\partial \log f^*(X,Y;\boldsymbol{\gamma})/\partial\gamma_j\right\} = 0$ and $E_{\boldsymbol{\gamma}}\left\{\partial \log f^*(X,Y;\boldsymbol{\gamma})/\partial\gamma_j \times \partial \log f^*(X,Y;\boldsymbol{\gamma})/\partial\gamma_k\right\} = E_{\boldsymbol{\gamma}}\left\{-\partial^2 \log f^*(X,Y;\boldsymbol{\gamma})/\partial\gamma_j\gamma_k\right\}$ for $j,k = 1,\ldots,p$.

(e) The information matrix $I(\boldsymbol{\gamma})$ is positive definite for all $\boldsymbol{\gamma} \in \Gamma^O$, where the (jk)-th element of $I(\boldsymbol{\gamma})$ is defined by $I_{jk}(\boldsymbol{\gamma}) = \mathrm{cov}\left\{\partial \log f^*(X,Y;\boldsymbol{\gamma})/\partial\gamma_j, \partial \log f^*(X,Y;\boldsymbol{\gamma})/\partial\gamma_k\right\}$.

(f) There exists functions G_{jkl} such that $\mid \partial^3 \log f^*(x,y;\boldsymbol{\gamma})/\partial\gamma_j\gamma_k\gamma_l \mid \leq G_{jkl}(x,y)$ for all $\boldsymbol{\gamma} \in \Gamma^O$ where $E_{\boldsymbol{\gamma}^0}\left\{G_{jkl}(X,Y)\right\} < \infty$ for all j,k,l.

References

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

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

Berger, J. O.and Bernardo, J. M. and Sun, D. (2009), "The case for objective Bayesian analysis," *Ann. Statist.*, 37, 905–938.

Berger, J. O. (2006), "The case for objective Bayesian analysis," *Bayesian Anal.*, 1, 385–402.

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, 20–31.

Cox, D. (2013), "Discussion of 'Confidence distribution, the frequentist distribution estimator of a parameter — a Review' by Xie and Singh," *Int. Statist. Rev.*, 81, 40–41.

Cox, D. R. (1958), "Some problems connected with statistical inference," *Ann. Math. Statist*, 29, 357–372.

Dempster, A. P. (2008), "DempsterShafer calculus for statisticians," *Internat. J. of Approx. Reason*, 48, 265–277.

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

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

Entsuah, A. R., Huang, H., and Thase, M. E. (2001), "Response and remission rates in different subpopulations with major depressive disorder administered venlafaxine, selective serotonin reuptake inhibitors, or placebo." *Journal of Clinical Psychiatry*, 62, 869–877.

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

Hannig, J. (2009), "On Generalized Fiducial Inference," *Statistica Sinica*, 19, 491–544.

— (2013), "Generalized Fiducial Inference via Discretizion," *Statistica Sinica*, 23, 489–514.

Hannig, J. and Lee, T. (2009), "Generalized fiducial inference for wavelet regression," *Biometrika*, 96, 491–544.

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

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

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

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

Martin, R. and Liu, C. (2013), "Inferential models: A framework for prior-free posterior probabilistic inference," *Journal of the American Statistical Association*, 108, 301–313.

Martin, R., Zhang, J., and Liu, C. (2010), "DempsterShafer theory and statistical inference with weak beliefs," *Statist. Sci.*, 25, 905–938.

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

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

Shekelle, P. G., Rich, M. W., Morton, S. C., Atkinson, C. S. W., Tu, W., Maglione, M., Rhodes, S., Barrett, M., Fonarow, G. C., Greenberg, B., Heidenreich, P. A., Knabel, T., Konstam, M. A., Steimle, A., and Stevenson, L. W. (2003), "Efficacy of angiotensin-converting enzyme inhibitors and beta-blockers in the management of left ventricular systolic dysfunction according to race, gender, and diabetic status: A meta-analysis of major clinical trials," *Journal of the American College of Cardiology*, 41, 1529–1538.

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

Singh, K., Xie, M., and Strawderman, W. E. (2007), "Confidence distribution (CD): distribution estimator of a parameter," in *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond. IMS Lecture Notes-Monograph Series*, eds. Liu, R., Strawderman, W., and Zhang, C.-H., Beachwood, Ohio: Institute of Mathematical Statistics, vol. 54, pp. 132–150.

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A., and Thacker, S. B. (2000), "Meta-analysis of observational studies in epidemiology," *Journal of the American Medical Association*, 283, 2008–2012.

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

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

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

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

Xie, M. and Singh, K. (2013), "Confidence distribution, the frequentist distribution estimator of a parameter — a Review (with discussion)," *Int. Statist. Rev.*, 81, 3–39.

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

Zarrouf, F. A., Artz, S., Griffith, J., Sirbu, C., and Kommor, M. (2009), "Testosterone and depression: systematic review and meta-analysis," *Journal of Psychiatric Practice*, 15, 289–305.

Table 1: Meta-analysis in the presence of a study with missing covariate design (# of replicates = 1000)

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

Mean: mean of parameter estimates; SE: standard error of parameter estimates; SEE: mean of standard error estimates.

Table 2: Meta-analysis in the presence of a study with dichotomized responses (# of replicates = 1000)

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

Mean–mean of parameter estimates; SE–standard error of parameter estimates; SEE– mean of standard error estimates.

Table 3: Robust meta-analysis with a variety of working correlation matrices (# of replicates = 1000)

	Independence			Design Driven			Data Driven		
Parameters	Bias	SE	RE	Bias	SE	RE	Bias	SE	RE
α_1	0.00	0.38	0.97	0.00	0.36	1.00	0.00	0.12	0.97
α_2	-0.01	0.84	0.57	0.00	0.51	1.00	0.00	0.18	0.89
α_3	-0.02	0.66	0.72	-0.02	0.47	0.98	0.00	0.16	0.94
β_1	0.02	0.52	1.00	0.01	0.52	1.00	-0.01	0.16	0.98
β_2	0.01	0.16	0.94	0.00	0.15	1.00	0.00	0.05	0.95
β_3	-0.01	0.26	0.77	-0.01	0.20	1.00	0.00	0.06	0.95

Bias–mean of parameter estimates minus the true value of the parameter; SE–standard error of parameter estimates; RE –relative efficiency to the IPD estimator.

Table 4: Summary statistics for regression coefficients from individual studies of the landing data

	Part 1: continuous outcomes		Part 2: dichotomized Boeing outcomes	
Parameters	Airbus 321	Boeing 737	Airbus 321	Boeing 737
α_1 (Airbus)	NA	NA	NA	NA
α_2 (Boeing)	NA	3.40 (0.24)	NA	NA
β_1 (x_1)	-0.65 (1.02)	-2.87 (0.89)	-0.65 (1.02)	NA
β_2 (x_2)	7.83 (0.13)	8.70 (0.10)	7.83 (0.13)	NA
β_3 (x_3)	2.27 (0.10)	2.21 (0.06)	2.27 (0.10)	NA
β_4 (x_4)	-0.50 (0.75)	-1.51 (0.42)	-0.50 (0.75)	NA
β_5 (x_5)	NA	2.35 (0.36)	NA	NA
β_6 (x_6)	NA	4.08 (0.76)	NA	NA
β_7 $(x_5 : x_6)$	NA	-4.69 (1.13)	NA	NA

Within the parentheses is the estimated standard error of the corresponding parameter estimate.

Table 5: Meta-analysis of the landing data

	Part 1: continuous outcomes				Part 2: dichotomized Boeing outcomes			
Parameters	Conventional	IPD	CD	Robust	Conventional	IPD	CD	Robust
α_1 (Airbus)	NA	4.10 (0.15)	4.10 (0.15)	3.83 (0.20)	NA	4.62 (0.27)	4.63 (0.27)	4.35 (0.50)
α_2 (Boeing)	NA	3.95 (0.15)	3.95 (0.15)	3.40 (0.24)	NA	4.47 (0.27)	4.48 (0.27)	3.57 (0.65)
β_1 (x_1)	-1.91 (0.67)	-1.98 (0.67)	-1.98 (0.67)	-1.91 (0.67)	NA	-0.97 (0.89)	-0.93 (0.89)	-1.37 (0.91)
β_2 (x_2)	8.38 (0.08)	8.37 (0.08)	8.37 (0.08)	8.36 (0.08)	NA	7.67 (0.12)	7.65 (0.12)	7.83 (0.13)
β_3 (x_3)	2.23 (0.05)	2.20 (0.05)	2.19 (0.05)	2.22 (0.05)	NA	2.22 (0.08)	2.22 (0.08)	2.26 (0.09)
β_4 (x_4)	-1.27 (0.37)	-1.34 (0.37)	-1.34 (0.37)	-1.27 (0.37)	NA	-0.25 (0.58)	-0.18 (0.58)	-0.22 (0.59)
β_5 (x_5)	NA	1.61 (0.23)	1.61 (0.23)	2.19 (0.20)	NA	1.12 (0.38)	1.09 (0.38)	1.81 (0.49)
β_6 (x_6)	NA	2.50 (0.52)	2.50 (0.52)	4.08 (0.76)	NA	0.16 (1.00)	0.14 (1.00)	1.93 (2.05)
β_7 $(x_5 : x_6)$	NA	-2.40 (0.76)	-2.40 (0.76)	-5.08 (0.83)	NA	0.28 (1.42)	0.37 (1.43)	-3.52 (2.06)

Within the parentheses is the estimated standard error of the corresponding parameter estimate.