Relative efficiency of using summary versus individual data in random-effects meta-analysis

Ding-Geng Chen[1,2], Dungang Liu [*3], Xiaoyi Min[4,5] and Heping Zhang[6]

[1] School of Social Work & Department of Biostatistics,
University of North Carolina, Chapel Hill, NC 27599, USA

[2] Department of Statistics, University of Pretoria, South Africa

[3] Department of Operations, Business Analytics, and Information Systems
Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221, USA

[4]Department of Mathematics and Statistics, Georgia State University
Atlanta, GA 30303, USA

[5] Hudson Data, New York, NY 10004, USA

[6]Department of Biostatistics, Yale University, New Haven, CT 06510, USA

February 1, 2020

[*]The author to whom all correspondence should be addressed (Email: dungang.liu@uc.edu)

Relative efficiency of using summary versus individual data in random-effects meta-analysis

Abstract

Meta-analysis is a statistical methodology for combining information from diverse sources so that a more reliable and efficient conclusion can be reached. It can be conducted by either synthesizing study-level summary statistics or drawing inference from an overarching model for individual participant data (IPD) if available. The latter is often viewed as the "gold standard". For random-effects models, however, it remains not fully understood whether the use of IPD indeed gains efficiency over summary statistics. In this paper, we examine the relative efficiency of the two methods under a general likelihood inference setting. We show theoretically and numerically that summary-statistics-based analysis is *at most* as efficient as IPD analysis, provided that the random effects follow the Gaussian distribution and maximum likelihood estimation is used to obtain summary statistics. More specifically, (i) the two methods are equivalent in an asymptotic sense; and (ii) summary-statistics-based inference can incur an appreciable loss of efficiency if the sample sizes are not sufficiently large. Our results are established under the assumption that the between-study heterogeneity parameter remains constant regardless of the sample sizes, which is different from a previous study. Our findings are confirmed by the analyses of simulated data sets and a real world study of alcohol interventions.

Key Words: Divide and conquer; evidence synthesis; individual participant data; literature review; one-stage IPD; two-stage IPD.

1 Introduction

In the big data era, it has become the norm, rather than the exception, that the data collected to address the same/similar scientific question come from diverse sources. The art and science of synthesizing information from diverse sources to draw a more effective inference is generally referred to as meta-analysis. In the past thirty years, meta-analysis has played an important role in health and medical sciences, and its applications have led to numerous scientific discoveries. For example, meta-analysis results are reported in more than six hundred articles in the *New England Journal of Medicine* in the past decade. Although there is a rich literature on meta-analysis (e.g., Borenstein et al., 2009; Pigott, 2012; Chen and Peace, 2013), many important issues remain unsettled. One of them is: *Does analyzing individual participant data (IPD) from all studies indeed gain efficiency over combining summary statistics from each study?*

Traditional meta-analysis was confined to synthesis of research findings, such as reported effect sizes, from publications. Given the estimates of a common effect size from several studies, a meta-analyst combines these (summary) statistics, with the goal of achieving a more efficient estimate. Nowadays, as data sharing has been increasingly encouraged, original data at the individual level may be accessible on certain database platforms such as dbGaP (2020). When individual data are available, it is generally believed that the so-called IPD method may reduce bias and gain efficiency in inference. The IPD method refers to building an overarching model for all individuals and drawing inference from the overall likelihood using the maximum likelihood method. Despite several alleged advantages of analyzing IPD (Sutton and Higgins, 2008), summary-statistics-based meta-analysis is still prevalent in practice for several reasons. First, the retrieval of original data could be unforeseeably time-consuming. Whether the benefits of using the IPD method can outweigh the tremendous cost of obtaining original data is still under debate (Sutton and Higgins,

2008). In certain scenarios, it is not unusual that IPD are inaccessible due to privacy issues (Lee et al., 2017). In fact, the lack of data sharing regulations in many fields discourages research institutes from making individual-level data available to the public. Second, having access to IPD does not necessarily mean that we are able to analyze the entire data all at once. When the data volume is too large to be analyzed on a single computer, as often seen in computer science or machine learning problems, the only feasible solution is to conduct statistical learning from each study and combine learning results (e.g., summary statistics) at the end (Jordan et al., 2013; Chen and Xie, 2014; Cheung and Jak, 2016; Lee et al., 2017). In health care and medical research, there is also an on-going discussion on when the two-stage IPD method, which is easy to implement and interpret for practitioners, may produce similar result to the one-stage IPD method, which nevertheless requires more advanced statistical and computational support (see, e.g., Burke et al., 2017; Kontopantelis, 2018).

For the reasons discussed above, it is important to have a better understanding of how much efficiency summary-statistics-based meta-analysis could potentially lose compared to the IPD method. A series of works has rigorously examined their relative efficiency (Olkin and Sampson, 1998; Mathew and Nordstrom, 1999; Simmonds and Higgins, 2007; Lin and Zeng, 2010; Liu et al., 2015). Specifically, Olkin and Sampson (1998) and Mathew and Nordstrom (1999) focused on analysis of variance (ANOVA). Simmonds and Higgins (2007) examined a special case of linear regression models for continuous responses. Lin and Zeng (2010) reached far beyond those special settings and considered a general likelihood inference setting. Their result was further extended by Liu et al. (2015) to a more complex setting of analyzing heterogenous studies and achieving complex evidence synthesis. These studies are all restricted to common-effects models (also known as fixed-effects models), which assume that the parameter of interest has a common value across all relevant studies. This assumption, however, does not hold when the effect of interest exhibits between-study

variations.

To accommodate between-study variations, random-effects models are often used and they model the study-specific effects as realizations of a (normal) random variable. Although random-effects models are widely used in the literature as well as in practice, the relative efficiency of using IPD versus summary statistics has not been fully studied. Recently, Zeng and Lin (2015) showed a surprising result; that is, summary-statistics-based analysis is always at least as efficient as the IPD analysis. This conclusion relies on a critical assumption that the between-study variability is of order n^{-1}, where n is the median sample size of studies. In other words, the between-study variability will vanish as the size of each study becomes larger. This assumption, however, may not hold in real world problems. In clinical or psychiatric studies, for example, random effects may represent differences between hospitals or other structural difference among subpopulations. The variability of such random effects may be more appropriately assumed to be a constant, rather than a diminishing term as the patient number in each hospital increases. Similarly, in business fields, random effects often reflect the difference between corporations. The heterogeneity will not disappear as the size of data from each corporation increases.

This paper focuses on the standard random-effects model used in meta-analysis, as opposed to the common-effects models. Assuming the between-study variability remains constant, we systematically investigate the relative efficiency of IPD- and summary-statistics-based meta-analyses. Under a general likelihood inference setting, we show theoretically and numerically that summary-statistics-based analysis is *at most* as efficient as IPD analysis, provided that the random effects follow the Gaussian distribution and maximum likelihood estimation is used to obtain summary statistics. More specifically, the two meta-analysis methods are asymptotically equivalent. The asymptotics here refers to that both the study size n and the number of studies K are sufficiently large, and n diverges at a

higher-order rate (i.e., $Kn^{-1/2} \to 0$). On the other hand, given small or moderate n and K, summary-statistics-based analysis may incur an appreciable loss of efficiency.

2 Theoretical Results

We consider a general likelihood inference setting similar to those examined in the literature (Lin and Zeng, 2010; Liu et al., 2015; Zeng and Lin, 2015). Assume that there are K independent studies with n_k individuals in the k-th study ($k = 1, \ldots, K$). For each study, we let $(Y_{ki}, \boldsymbol{X}_{ki})$ ($i = 1, \ldots, n_k$) denote the original individual data, where Y_{ki} and p-dimensional vector $\boldsymbol{X}_{ki}$ may represent a response variable and p explanatory variables, respectively. These variables are allowed to be either continuous or categorical. We assume that in each study, the individual data $(Y_{ki}, \boldsymbol{X}_{ki})$ follow a general random-effects model:

- Across-study level: The random-effects $\boldsymbol{\beta}_k \mid \boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \boldsymbol{T})$.
- Within-study level: Given $\boldsymbol{\beta}_k$ and a nuisance parameter vector $\boldsymbol{\eta}_k$, $(Y_k, \boldsymbol{X}_{ki})$ has density function $f_k(Y_k, \boldsymbol{X}_{ki}; \boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$.

At the across-study level, the parameter $\boldsymbol{\beta}$ represents the mean effect of the random-effects $\boldsymbol{\beta}_k$'s, and $\boldsymbol{T}$, a variance-covariance matrix, represents the between-study variability. Oftentimes, of interest is the inference of $\boldsymbol{\beta}$. At the within-study level, the density function $f_k(Y_{ki}, \boldsymbol{X}_{ki}; \boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ can be derived from parametric models.

In clinical or social research, for instance, often used in each study is a generalized linear model as below:

$$g(E(Y_k)) = \beta_{0k} + \beta_{1k} X_{1k} + \beta_{2k} X_{2k} + \beta_{3k} X_{1k} X_{2k}, \tag{2.1}$$

where $g(\cdot)$ is the link function. Model (2.1) gives out a linear regression model when

$g(\mu) = \mu$, a logistic regression model when $g(\mu) = \log(\mu/(1-\mu))$, and a probit regression model when $g(\mu) = \Phi^{-1}(\mu)$. The coefficients β_{1k}, β_{2k}, and β_{3k} may represent the effects of treatment, a covariate, and their interaction, respectively. In each individual study, researchers may only report in their publication the summary statistics for the treatment effect β_{1k} and in some cases, the interaction effect β_{3k} as well. Then, $\boldsymbol{\eta}_k = (\beta_{0k}, \beta_{2k})'$ are treated as nuisance parameters, for which summary statistics may not be available.

2.1 Summary-statistics-based inference

Let $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\eta}}_k$ be the maximum likelihood estimates of $\boldsymbol{\beta}_k$ and $\boldsymbol{\eta}_k$, respectively, by maximizing the log-likelihood function from the k-th study:

$$\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k) = \log L_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k) = \sum_{i=1}^{n_k} \log f_k(Y_{ki}, \boldsymbol{X}_{ki}; \boldsymbol{\beta}_k, \boldsymbol{\eta}_k).$$

Denote the observed information matrix by

$$\mathcal{I}_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k) = \begin{pmatrix} \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k} \\ \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k} \end{pmatrix} = \begin{pmatrix} -\partial^2\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)/\partial\boldsymbol{\beta}_k^2 & -\partial^2\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)/\partial\boldsymbol{\beta}_k\partial\boldsymbol{\eta}_k \\ -\partial^2\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)/\partial\boldsymbol{\eta}_k\partial\boldsymbol{\beta}_k & -\partial^2\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)/\partial\boldsymbol{\eta}_k^2 \end{pmatrix}.$$

When used for deriving statistics, $\mathcal{I}_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ is evaluated by plugging in an estimate of the parameters. In traditional meta-analyses, we retrieve $\hat{\boldsymbol{\beta}}_k$ and the estimate of its variance:

$$\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_k \mid \boldsymbol{\beta}_k, \boldsymbol{\eta}_k) = \left(\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} - \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}^{-1}\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k}\right)^{-1}_{|\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\eta}}_k}$$

from each study, and use them for summary-statistics based meta-analytic inference. When $(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)'s$ are treated as random effects, $\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_k \mid \boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ is an estimate of $\hat{\boldsymbol{\beta}}_k$'s conditional variance, and we let $\widehat{\mathrm{var}}_C(\hat{\boldsymbol{\beta}}_k) = \widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_k \mid \boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ for notational simplicity.

Specifically, traditional meta-analyses assume, at least approximately, that

$$\hat{\boldsymbol{\beta}}_k \mid \boldsymbol{\beta}_k \sim N\left(\boldsymbol{\beta}_k, \widehat{\mathrm{var}}_C(\hat{\boldsymbol{\beta}}_k)\right).$$

Coupled with the random-effects assumption $\boldsymbol{\beta}_k \mid \boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \boldsymbol{T})$, the unconditional distribution of $\hat{\boldsymbol{\beta}}_k$ is

$$\hat{\boldsymbol{\beta}}_k \sim N\left(\boldsymbol{\beta}, \widehat{\text{var}}_C(\hat{\boldsymbol{\beta}}_k) + \boldsymbol{T}\right).$$

Averaging these $\hat{\boldsymbol{\beta}}_k$'s using the inverse-variance weighting scheme leads to an overall estimator $\hat{\boldsymbol{\beta}}_{SS}$ of $\boldsymbol{\beta}$ as follows

$$\hat{\boldsymbol{\beta}}_{SS} = \left[\sum_{k=1}^{K} \left\{\widehat{\text{var}}_C(\hat{\boldsymbol{\beta}}_k) + \hat{\boldsymbol{T}}\right\}^{-1}\right]^{-1} \sum_{k=1}^{K} \left\{\widehat{\text{var}}_C(\hat{\boldsymbol{\beta}}_k) + \hat{\boldsymbol{T}}\right\}^{-1} \hat{\boldsymbol{\beta}}_k, \tag{2.2}$$

where $\hat{\boldsymbol{T}}$ is a consistent estimator of $\boldsymbol{T}$. Such $\hat{\boldsymbol{T}}$ could be obtained from the method of moments or the likelihood method (Whitehead, 2003, pp.90, 94-96). A consistent estimate of the variance of $\hat{\boldsymbol{\beta}}_{SS}$ is

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{SS}) = \left[\sum_{k=1}^{K} \left(\widehat{\text{var}}_C(\hat{\boldsymbol{\beta}}_k) + \hat{\boldsymbol{T}}\right)^{-1}\right]^{-1}. \tag{2.3}$$

For efficiency comparison, we also consider an *asymptotic variance* of $\hat{\boldsymbol{\beta}}_{SS}$

$$\text{aVar}(\hat{\boldsymbol{\beta}}_{SS}) = \left[\sum_{k=1}^{K} \left\{ \left(\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} - \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\mathcal{I}^{-1}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k}\right)^{-1} + \boldsymbol{T}\right\}^{-1}\right]^{-1},$$

where the variance components $\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k}, \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}, \mathcal{I}^{-1}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}$ and $\boldsymbol{T}$ are all evaluated using the true values of the parameters.

2.2 Relative efficiency to IPD-based inference

When the individual-level data are available from all the K studies, we can perform maximum likelihood inference for $\boldsymbol{\beta}$ by pooling together the log-likelihood function $\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ from each of the studies and integrating out the random effects $\boldsymbol{\beta}_k$. To implement this, we assume that the nuisance parameters $\boldsymbol{\eta}_k$'s are also random effects. More specifically, $\boldsymbol{\beta}_k$

and $\boldsymbol{\eta}_k$ jointly follow a multivariate normal distribution

$$\begin{pmatrix} \boldsymbol{\beta}_k \\ \boldsymbol{\eta}_k \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\eta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{T} & \boldsymbol{\Xi} \\ \boldsymbol{\Xi}^\top & \boldsymbol{\Phi} \end{pmatrix} \right).$$

Taking Model (2.1) as an example, $\boldsymbol{T}$ is a 2×2 variance-covariance matrix for the treatment and interaction effects $\boldsymbol{\beta}_k = (\beta_{1k}, \beta_{3k})'$, $\boldsymbol{\Phi}$ is a 2×2 variance-covariance matrix for the baseline and covariate effects $\boldsymbol{\eta}_k = (\beta_{0k}, \beta_{2k})'$, and $\boldsymbol{\Xi}$ is a 2×2 matrix representing the covariance between $\boldsymbol{\beta}_k$ and $\boldsymbol{\eta}_k$. In this example, the variance-covariance component at the across-study level contains 4 variance parameters and 6 correlation parameters.

Integrating out the random effects $\boldsymbol{\beta}_k$ and $\boldsymbol{\eta}_k$ using their joint distribution, we obtain the log-likelihood function of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ from the k-th study:

$$\begin{aligned} \ell_k(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi}) &= \log \iint L_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k) \begin{vmatrix} \boldsymbol{T} & \boldsymbol{\Xi} \\ \boldsymbol{\Xi}^\top & \boldsymbol{\Phi} \end{vmatrix}^{-1/2} \\ &\quad \exp\left[-\frac{1}{2} \begin{pmatrix} \boldsymbol{\beta}_k - \boldsymbol{\beta} \\ \boldsymbol{\eta}_k - \boldsymbol{\eta} \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{T} & \boldsymbol{\Xi} \\ \boldsymbol{\Xi}^\top & \boldsymbol{\Phi} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\beta}_k - \boldsymbol{\beta} \\ \boldsymbol{\eta}_k - \boldsymbol{\eta} \end{pmatrix} \right] d\boldsymbol{\beta}_k d\boldsymbol{\eta}_k. \quad (2.4) \end{aligned}$$

The overall log-likelihood function is $\ell(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi}) = \sum_{k=1}^{K} \ell_k(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi})$.

In what follows, we show that $\ell_k(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi})$ in (2.4) can be approximated by the logarithm of a normal density function. To achieve this, we expand $\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ in a neighborhood of the maximum likelihood estimates $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\eta}}_k$:

$$\begin{aligned} \ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k) = \ell_k(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\eta}}_k) &- \frac{1}{2} \begin{pmatrix} \boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta}_k - \hat{\boldsymbol{\eta}}_k \end{pmatrix}^\top \begin{pmatrix} \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k} \\ \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k} \end{pmatrix}_{|\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\eta}}_k} \begin{pmatrix} \boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta}_k - \hat{\boldsymbol{\eta}}_k \end{pmatrix} \\ &+ o_p\left(\left\| \begin{pmatrix} \boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta}_k - \hat{\boldsymbol{\eta}}_k \end{pmatrix} \right\|^2 \right). \quad (2.5) \end{aligned}$$

Although (2.5) is a local expansion, Laplace approximation theory ensures that plugging this expansion into the global integral in (2.4) yields the following result.

Lemma 1. *Under the regularity conditions (C1)-(C6) specified in the Supplementary Materials A, we have, with probability* 1*,*

$$\begin{aligned}\ell_k(\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{T},\boldsymbol{\Phi},\boldsymbol{\Xi}) &= -\frac{1}{2}\begin{pmatrix}\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_k\\ \boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_k\end{pmatrix}^\top\left\{\begin{pmatrix}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\\ \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}\end{pmatrix}^{-1}_{|\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\eta}}_k}+\begin{pmatrix}\boldsymbol{T} & \boldsymbol{\Xi}\\ \boldsymbol{\Xi}^\top & \boldsymbol{\Phi}\end{pmatrix}\right\}^{-1}\begin{pmatrix}\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_k\\ \boldsymbol{\eta}-\hat{\boldsymbol{\eta}}_k\end{pmatrix}\\ &\quad -\frac{1}{2}\log\left|\begin{pmatrix}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\\ \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}\end{pmatrix}_{|\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\eta}}_k}+\begin{pmatrix}\boldsymbol{T} & \boldsymbol{\Xi}\\ \boldsymbol{\Xi}^\top & \boldsymbol{\Phi}\end{pmatrix}^{-1}\right|+c_k+O(n_k^{-\frac{1}{2}}), \quad (2.6)\end{aligned}$$

as $n_k\to\infty$, where c_k is a statistic that does not depend on the parameters.

Note that (2.6) provides a point-wise approximation of ℓ_k. With the further assumption that the parameter space is compact, this approximation becomes uniform in the sense that the error term $O(n_k^{-\frac{1}{2}})$ in (2.6) does not rely on the values of the parameters as long as they are in the compact space. As a result, we can discuss the properties of the maximum likelihood estimator for $\boldsymbol{\beta}$ using (2.6). Hereafter, we assume that n is the median of $\{n_1,\ldots,n_K\}$ and $n_k=np_k$ where p_k is a constant within a compact interval in $(0,\infty)$.

Lemma 2. *Under conditions (C1)-(C7), for a fixed K, the maximizer of $\sum_{k=1}^K\ell_k(\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{T},\boldsymbol{\Phi},\boldsymbol{\Xi})$ for given $(\boldsymbol{T},\boldsymbol{\Phi},\boldsymbol{\Xi})$ (satisfying $\begin{pmatrix}\boldsymbol{T} & \boldsymbol{\Xi}\\ \boldsymbol{\Xi}^\top & \boldsymbol{\Phi}\end{pmatrix}$ is positive definite) is*

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}_{IPD}\\ \hat{\boldsymbol{\eta}}_{IPD}\end{pmatrix}=\left\{\left[\sum_{k=1}^K\mathcal{M}_k(\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\eta}}_k,\boldsymbol{T},\boldsymbol{\Phi},\boldsymbol{\Xi})^{-1}\right]^{-1}\left[\sum_{k=1}^K\mathcal{M}_k(\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\eta}}_k,\boldsymbol{T},\boldsymbol{\Phi},\boldsymbol{\Xi})^{-1}\begin{pmatrix}\hat{\boldsymbol{\beta}}_k\\ \hat{\boldsymbol{\eta}}_k\end{pmatrix}\right]\right\}+o_p(1), \quad (2.7)$$

as $n\to\infty$ where

$$\mathcal{M}_k(\boldsymbol{\beta}_k,\boldsymbol{\eta}_k,\boldsymbol{T},\boldsymbol{\Phi},\boldsymbol{\Xi})=\begin{pmatrix}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\\ \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k} & \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}\end{pmatrix}^{-1}+\begin{pmatrix}\boldsymbol{T} & \boldsymbol{\Xi}\\ \boldsymbol{\Xi}^\top & \boldsymbol{\Phi}\end{pmatrix}.$$

Remark 1. *For the approximation in Lemma 2 to hold as both $K,n\to\infty$ and $Kn^{-1/2}\to 0$, further uniformity conditions are needed to guarantee that the sum of the reminder terms in*

(2.6) over k's still converges. Such conditions can be obtained by extending the conditions for Theorem 7 in Kass et al. (1990).

Zeng and Lin (2015) suggested that directly maximizing with respect to $\boldsymbol{T}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Xi}$ might result in inconsistent estimators, which may in turn lead to less efficient estimators for $\boldsymbol{\beta}$. We consider maximizing $\sum_{k=1}^{K} \ell_k(\boldsymbol{\beta}, \boldsymbol{\eta}, \hat{\boldsymbol{T}}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Xi}})$ for some consistent estimators $\hat{\boldsymbol{T}}$, $\hat{\boldsymbol{\Phi}}$, and $\hat{\boldsymbol{\Xi}}$. In that case, the variance of $\hat{\boldsymbol{\beta}}_{IPD}$ can be consistently estimated by

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{IPD}\right) = \left\{ \sum_{k=1}^{K} \mathcal{M}_k(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\eta}}_k, \hat{\boldsymbol{T}}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Xi}})^{-1} \right\}^{-1}_{[\boldsymbol{\beta},\boldsymbol{\beta}]}. \tag{2.8}$$

The asymptotic variance of $\hat{\boldsymbol{\beta}}_{IPD}$ is

$$\text{aVar}\left(\hat{\boldsymbol{\beta}}_{IPD}\right) = \left\{ \sum_{k=1}^{K} \mathcal{M}_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi})^{-1} \right\}^{-1}_{[\boldsymbol{\beta},\boldsymbol{\beta}]}.$$

To examine the relative efficiency of $\hat{\boldsymbol{\beta}}_{IPD}$ and $\hat{\boldsymbol{\beta}}_{SS}$, we compare their estimated variances and asymptotic variances. The results are presented below, where for matrices $A \geq B$ means that $A - B$ is positive semi-definite.

Theorem 1. *Under the conditions (C1)-(C7), we have the following results:*

(a) For any fixed K and n, the variance estimates follow the inequality

$$\widehat{var}\left(\hat{\boldsymbol{\beta}}_{SS}\right) \geq \widehat{var}\left(\hat{\boldsymbol{\beta}}_{IPD}\right)$$

provided that the same estimate $\hat{\boldsymbol{T}}$ of $\boldsymbol{T}$ is used for calculating $\widehat{var}(\hat{\boldsymbol{\beta}}_{SS})$ in (2.3) and $\widehat{var}\left(\hat{\boldsymbol{\beta}}_{IPD}\right)$ in (2.8).

(b) For any fixed K and n, the asymptotic variances follow the inequality

$$aVar\left(\hat{\boldsymbol{\beta}}_{SS}\right) \geq aVar\left(\hat{\boldsymbol{\beta}}_{IPD}\right).$$

The strict inequality may hold even if $\boldsymbol{\Xi} = \mathbf{0}$.

(c) As both $K, n \to \infty$ and $Kn^{-1/2} \to 0$, the asymptotic variances follow

$$\lim_{n\to\infty, K\to\infty} K \cdot aVar\left(\hat{\boldsymbol{\beta}}_{SS}\right) = \lim_{n\to\infty, K\to\infty} K \cdot aVar\left(\hat{\boldsymbol{\beta}}_{IPD}\right).$$

The equality holds even if $\boldsymbol{\Xi} \neq \mathbf{0}$.

Theorem 1 says that (a) for any fixed K and n, if we use the same variance component estimate $\hat{\boldsymbol{T}}$ to evaluate the variances of $\hat{\boldsymbol{\beta}}_{SS}$ and $\hat{\boldsymbol{\beta}}_{IPD}$, the variance estimates follow the inequality $\widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{SS}\right) \geq \widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{IPD}\right)$. The same inequality holds if the true values of all the parameters are used to evaluate the variances of $\hat{\boldsymbol{\beta}}_{SS}$ and $\hat{\boldsymbol{\beta}}_{IPD}$, as stated in Theorem 1(b). In this situation, even if $\boldsymbol{\Xi} = \mathbf{0}$ (i.e., the two random effects $\boldsymbol{\beta}_k$ and $\boldsymbol{\eta}_k$ are independent), the strict inequality in (b) may hold, i.e., $\text{aVar}\left(\hat{\boldsymbol{\beta}}_{SS}\right) > \text{aVar}\left(\hat{\boldsymbol{\beta}}_{IPD}\right)$. The reason is that within each study, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\eta}}_k$ may correlate with each other, and thus $\mathcal{M}_k$ in Lemma 2 may not be a diagonal matrix even if $\boldsymbol{\Xi} = \mathbf{0}$. The two inequalities in (a) and (b) are confirmed in our simulation studies, where $\hat{\boldsymbol{\beta}}_{SS}$ incurs an appreciable loss of efficiency.

When $n \to \infty$ at a higher-order rate of K, Theorem 1(c) shows that the asymptotic variance of $\hat{\boldsymbol{\beta}}_{SS}$ is equal to that of $\hat{\boldsymbol{\beta}}_{IPD}$. The equality holds even if the covariance $\boldsymbol{\Xi}$ between the random effects $\boldsymbol{\beta}_k$ and $\boldsymbol{\eta}_k$ is non-zero. This implies that asymptotically, summary-statistics-based meta-analysis can achieve full efficiency *without requiring the information of nuisance parameters and correlation.* This is different from the findings for the fixed-effects models examined by Lin and Zeng (2010) and Liu et al. (2015). They proved that meta-analysis of summary statistics can incur a substantial loss of efficiency if the effects of interest are correlated with nuisance effects yet the correlation is not reported. Our seemingly counterintuitive finding can be heuristically explained by (2.7), where the IPD estimates $(\hat{\boldsymbol{\beta}}_{IPD}, \hat{\boldsymbol{\eta}}_{IPD})'$ are approximated by a matrix-weighted average of the study-specific estimates $(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\eta}}_k)'$. The weight matrix $\mathcal{M}_k$ is a sum of two components. The first component will diminish as $n \to \infty$, while the second remains constant. Thus, if of interest is merely the inference of $\boldsymbol{\beta}$, the contribution of $\hat{\boldsymbol{\eta}}_k$'s will diminish as well when n diverges at a faster rate than K. Our findings here will be numerically demonstrated in Section 3.

Remark 2. *The fact that* $(\hat{\boldsymbol{\beta}}_{IPD}, \hat{\boldsymbol{\eta}}_{IPD})'$ *can be approximated by a matrix-weighted average*

of the study-specific estimates $(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\eta}}_k)'$ *as seen in (2.7) also implies a weaker result; that is, analyzing summary statistics will have no loss of efficiency asymptotically if (a) there is no nuisance parameter; or (b) the estimates of all nuisance parameters as well as their correlation estimates are reported and multivariate meta-analysis is conducted. A similar conclusion holds for fixed-effects models (Lin and Zeng, 2010; Liu et al., 2015). For random-effects models, our conclusion is stronger as the asymptotic equivalence holds even when nuisance parameters are present yet their summary statistics are not available.*

2.3 A special case

In the Supplementary Materials C, we examine a special case where the likelihood function in (2.5) is exact in the sense that there is no approximation error term $O(n_k^{-\frac{1}{2}})$. Specifically, assume that the k-th log-likelihood function is of the following form

$$\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k) = -\begin{pmatrix} \boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta}_k - \hat{\boldsymbol{\eta}}_k \end{pmatrix}^\top \begin{pmatrix} \mathcal{I}_{k,11} & \mathcal{I}_{k,12} \\ \mathcal{I}_{k,21} & \mathcal{I}_{k,22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta}_k - \hat{\boldsymbol{\eta}}_k \end{pmatrix} + c_k,$$

where c_k is a statistic that does not depend on any parameters. An example is that $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are coefficients in a linear regression model. The result indicates that we can establish an inequality between the *exact* variances of $\hat{\boldsymbol{\beta}}_{SS}$ and $\hat{\boldsymbol{\beta}}_{IPD}$, when each individual log-likelihood function can be written as a quadratic form of $(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ without an approximation error. For the general case where approximation errors may exist, the inequalities in Theorem 1(a)-(b) are established between the variance estimates or between the asymptotic variances.

3 Simulation Studies

We conduct simulation studies to numerically examine the relative efficiency of summary-statistics-based and IPD meta-analyses in random-effects settings. To mimic meta-analyses in the real world, our simulation follows a generalized linear model as in (2.1). This model has been used widely to examine the treatment effect as well as its interaction with a covariate (e.g., Burke et al., 2017; Kontopantelis, 2018). Specifically, we consider two settings where (a) Y_k's are continuous and linear regression models (i.e., $g(\mu) = \mu$) are used to produce summary statistics in each study; and (b) Y_k's are binary and probit regression models (i.e., $g(\mu) = \Phi^{-1}(\mu)$) are used to produce summary statistics. We explore a variety of scenarios by varying the number of studies, the sample size, the correlation between explanatory variables, and the size of signal-to-noise ratio.

3.1 Continuous outcomes

We simulate a continuous outcome from the linear regression model

$$Y_{ki} = \beta_{0k} + \beta_{1k}X_{1ki} + \beta_{2k}X_{2ki} + \beta_{3k}X_{1ki}X_{2ki} + \epsilon_{ki}, \quad k = 1, \cdots, K,\ i = 1, \cdots, n_k, \quad (3.1)$$

where $\epsilon_{ki} \sim N(0, \sigma_\epsilon)$. The binary variable X_{1ki} may represent the treatment status, and the continuous variable X_{2ki} may represent a covariate of interest. They are simulated from the following distribution

$$\begin{aligned} (X^*_{1ki}, X_{2ki}) &\overset{iid}{\sim} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho_{x,k}\\ \rho_{x,k} & 1\end{pmatrix}\right), \\ X_{1ki} &= I(X^*_{1ki} > 0). \end{aligned}$$

The random effects $\boldsymbol{\beta}_k \sim N(\mathbf{1}_k, \Sigma_\beta)$, where Σ_β has 0.25 as its diagonal elements and 0.125 as its off-diagonals. In other words, the standard deviation of each of the random effects

$(\beta_{k0}, \beta_{k1}, \beta_{k2}, \beta_{k3})$ is 0.5, and the correlation between them is 0.5.

Since our theory suggests that the weight matrix $\mathcal{M}_k$ in (2.7) plays a crucial role in determining the relative efficiency, we consider the following settings that specify a variety of its structure. For example, (A) $\mathcal{M}_k$'s are homogeneous across the studies ($\rho_{x,k} = 0$) or heterogeneous ($\rho_{x,k} \sim$ Uniform$(-0.3, 0.3)$ or $\rho_{x,k} \sim$ Uniform$(0, 0.7)$); (B) The first component of $\mathcal{M}_k$ (i.e., the within-study variance-covariance) is comparable to the second component (i.e. the between-study variance-covariance) or significantly smaller ($\sigma_\epsilon = 1, 3$ versus $\sigma_\epsilon = 10, 30$). We also consider various specifications for the number of studies (e.g., $K = 5, 10, 30, 50, 100$) and the sample size in each study (e.g., $n_k = 20, 50, 100, 200, 500$). All the results reported in this section are based on 1000 simulation replications.

We carry out meta-analysis for the interaction effect β_3. Given all the individual-level data, the IPD estimate $\hat{\beta}_{3,IPD}$ is obtained using the function *lmer* in the R package *lme4*. The summary-statistics-based estimate $\hat{\beta}_{3,SS}$ is derived using (2.2), given only the estimate of β_{3k} and its variance from each study. Table 1 reports the bias and variance of the two estimates when $\rho_{x,k} \sim$ Uniform$(-0.3, 0.3)$, assuming the variance components (i.e., σ_ϵ and Σ_β) are known. We observe that when the sample size of each study is small (e.g., $n_k = 20$), both the standard error (SE) and mean squared error (MSE) of $\hat{\beta}_{3,SS}$ are consistently larger than those of $\hat{\beta}_{3,IPD}$. This indicates an appreciable loss of efficiency in estimation using only summary statistics. To better assess such a loss, we calculate the ratio of MSE$(\hat{\beta}_{3,SS})$ and MSE$(\hat{\beta}_{3,IPD})$ as the measure of relative efficiency. We observe in Table 2 that the loss of efficiency is as high as 38% when $n_k = 20$ and $\sigma_\epsilon = 30$. As both n_k and K increase, the relative efficiency approaches one. When $\sigma_\epsilon = 30$, for example, the relative efficiency is 1.03 for $K = 50$ studies, each with a sample size $n_k = 500$. Such an efficiency difference is minimal and likely ignorable in practice. These numerical results have confirmed our theoretical conclusions in Section 2.4. In the Supplementary Materials, Tables 6-7 present

the corresponding results when the variance components σ_ϵ and Σ_β are unknown and their estimates are used. We observe similar patterns as described for Tables 1-2.

To examine the impact of within-study correlations, we vary the size of the correlation between X_1 and X_2. Tables 8-9 in the Supplementary Materials present relative efficiency when $\rho_{x,k} = 0$ and $\rho_{x,k} \sim \text{Uniform}(0, 0.7)$. It appears that the values in Table 8 ($\rho_{x,k} = 0$) are closer to 1. This observation is consistent with our theoretical finding; that is, the smaller the within-study correlation of $\mathcal{M}_k$ in (2.7), the smaller the difference between the two methods. Across Tables 7-9, we also see that the smaller the within-study variance (σ_ϵ), the smaller the difference between the two methods. To demonstrate the inequality of Theorem 1(b) when the between-study correlation $\boldsymbol{\Xi} = \mathbf{0}$, we set $\Sigma_\beta = 0.25\boldsymbol{I}$. In other words, the standard deviation of β_{jk} is 0.5, and the correlation between them is 0. The results are reported in Tables 10-11. The patterns are similar to what we have observed when $\boldsymbol{\Xi} \neq \mathbf{0}$; i.e., summary-statistics-based estimate incurs a significant loss of efficiency when n_k and K are small, but as both n_k and K increase, the loss becomes ignorable.

3.2 Binary outcomes

Using the same simulation settings as the previous section, we obtain a binary outcome $Z_{ki} = \text{I}(Y_{ki} > 0)$ where Y_{ki} follows Model (3.1). As a result, Z_{ki} follows a probit model

$$\Pr(Z_{ki} = 1) = \Phi[(\beta_{0k} + \beta_{1k}X_{1ki} + \beta_{2k}X_{2ki} + \beta_{3k}X_{1ki}X_{2ki})/\sigma_\epsilon].$$

As the dichotomization of the continuous outcome incurs a loss of information, we consider larger sample size n_k $(= 100, 200, 500, 1000, 2000)$ for the efficiency comparison.

Table 3 reports the bias and variance of $\hat{\beta}_{3,SS}$ and $\hat{\beta}_{3,IPD}$ when $\rho_{x,k} = 0$. Similar to the continuous case, we observe that both the SE and MSE of $\hat{\beta}_{3,SS}$ are consistently larger than those of $\hat{\beta}_{3,IPD}$ when $n_k = 100$ or 200. The loss of efficiency is about $7-10\%$ when $n_k = 100$

as seen in Table 4. As the sample size increases to $n_k = 2000$, the difference in MSEs falls well below 1%, which is negligible in practice. As a further study of the relative efficiency, we simulate studies with a mixture of sample sizes. Specifically, 20% of the K studies have sample sizes three times larger than the others. The results are presented in Tables 12-13 when $\rho_{x,k} = 0.7$. We once again observe that summary-statistics-based inference incurs an appreciable loss of efficiency when the sample size is small (e.g., $n_k = 100$). But when both n_k and K are sufficiently large (e.g., $n_k = 2000, K = 50$), the MSEs of the IPD- and summary-statistics-based estimates become comparable. Similar patterns are observed in our simulation of mixed sample sizes for continuous outcomes.

4 Meta-Analysis of the Alcohol Intervention Data

To reduce heavy drinking and related negative consequences, brief motivational interventions have been implemented on college campuses over the last two decades. Huh et al. (2019a) examined the effect of a new intervention using personalized feedback delivered by mail, computer, or the Web (treatment), in comparison with the traditional intervention such as in-person motivational interviews (control). A comprehensive comparison is possible by using the data from nine independent studies (reported in Table 1 of Huh et al., 2019a). The sample sizes of these studies can be found in Table 5. Data availability is described in the Data Availability Section.

To evaluate the effect of the new intervention, we consider the difference between the number of drinks measured at the baseline and followup. It is a continuous outcome, denoted as $Y^{(1)}$. Another outcome of interest is whether or not the difference is negative (i.e. the number of drinks at followup is less than that at baseline). This is a binary outcome, denoted as $Y^{(2)}$, with 1 for "yes" and 0 for "no". We consider three explanatory

variables: the treatment indicator X_1 (1 for treatment and 0 for control), an individual's onset number of drinks X_2, and the interaction between X_1 and X_2.

Assuming that all the regression coefficients are random effects, the IPD meta-analysis fits GLMs (2.1) to the individual data. Specifically, we use linear models for the continuous outcome $Y^{(1)}$ and logistic models for the binary outcome $Y^{(2)}$. The summary-statistics-based inference only uses the estimates of the study-specific coefficients and their variances. The analysis result is reported in Table 5. Overall, the analysis of summary statistics produces similar estimates to those from the IPD analysis. Nevertheless, we should not overlook a notable difference in the estimates of the treatment effect ($\hat{\beta}_{1,IPD} = 0.396$ versus $\hat{\beta}_{1,SS} = 0.187$) for the binary outcome $Y^{(2)}$. As for the estimation of variability, we observe that the two methods produce similar estimates of the standard errors. On the other hand, those based on summary statistics are consistently larger than those based on individual data, with an appreciable loss of efficiency ($6 - 20\%$). These observations confirm our theory, and they are consistent with our findings in simulation studies.

In the section of Discussion, we point out two factors that may alleviate the loss of efficiency: (a) the smaller the ratio of the within-study variance versus the between-study variance; and (b) the smaller the within-study correlation. In this example, the median of such variance ratios is 4.76, meaning that on average the within-study variances are much larger than the between-study variances. The median of the correlations is 0.68, indicating strong a within-study association. These results show that neither of Factors (a) or (b) is present. It is thus not surprising to see a notable loss of efficiency.

As seen in the footnote of Table 5, among the nine studies used for meta-analysis, two studies (Study 8a and 8b) are considerably larger than the others. We therefore carry out a further analysis by excluding these two studies and performing meta-analysis of the remaining studies. The result is presented in Table 15. The conclusions are similar to those

from the analysis of all the studies. In particular, (1) the onset number of drinks (X_2) is significant, and its effect size is similar to that with the two large studies included; and (2) the standard errors based on summary statistics are consistently larger than those based on individual data, with an appreciable loss of efficiency ($5 - 22\%$).

5 Discussion

This paper has examined the relative efficiency of using summary statistics to perform random-effects meta-analysis as compared to the gold standard of using the IPD method. Our theoretical and numerical findings can be summarized as follows:

(i) Asymptotically, summary-statistics-based meta-analysis is as efficient as the IPD analysis. The asymptotics refers to that both the study size n and the number of studies K are sufficiently large, and n diverges at a higher-order rate ($Kn^{-1/2} \to 0$). The attainment of the full efficiency does not require information of nuisance parameters, meaning that needed are summary statistics for the parameter of interest.

(ii) For small or moderate K and n_k, summary-statistics-based meta-analysis may incur an appreciable loss of efficiency. The following factors may generally alleviate the loss of efficiency: (a) the smaller the within-study variance as compared to the between-study variance; and (b) the smaller the within-study correlation.

Our findings are different from those reported in Zeng and Lin (2015) in which they assumed that the between-study variability will diminish as the sample size n increases in each study. We has adopted a more practical assumption that the between-study variability remains constant, regardless of the the sample size n. Practitioners can decide which theory applies contingent upon which assumption is more suitable for the case under consideration.

Our result has implications for the comparison between one-stage and two-stage IPD meta-analyses. The one-stage IPD method analyzes the individual data from all the studies through a hierarchical model with random effects. The two-stage IPD method analyzes the data in each study separately and then combines summary statistics using traditional meta-analysis methods. In the second stage, univariate meta-analysis is often conducted for the parameter of interest. Current literature provides empirical and numerical evidence suggesting that these two IPD methods often give very similar results, and most differences arise because of different modeling assumptions (Burke et al., 2017; Kontopantelis, 2018). Our result theoretically confirms the asymptotic equivalence of the one-stage and two-stage IPD methods *when all the regression coefficients are modeled as random effects*.

On the other hand, when the sample sizes of the studies are not sufficiently large, the traditional two-stage IPD method may incur an appreciable loss of efficiency as evidenced by our numerical results of summary-statistics-based inference. The reason is that when the within-study variance-covariance (i.e., the first component of the weight matrix $\mathcal{M}_k$ in (2.7)) is comparable to the between-study variance-covariance (i.e., the second component of $\mathcal{M}_k$), the nuisance parameters $\boldsymbol{\eta}_k$'s may also contribute to the inference of $\boldsymbol{\beta}$ through the correlations. But this information is not used in the traditional two-stage method, which carries out univariate meta-analysis in its second stage. Our result in (2.7) suggests that multivariate meta-analysis be conducted in the second stage of the two-stage meta-analysis. Under the same simulation setting of Table 4, Table 14 in the Supplementary Materials presents the relative efficiency of multivariate analysis of summary statistics for $(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ versus analysis of individual data. The comparison between Table 4 and Table 14 shows a notable gain of efficiency, as a result of borrowing strength from nuisance parameters.

The correlations in (2.7) may come from multiple sources. For example, the correlation between the treatment variable and the covariate will certainly contribute to the first

component matrix of $\mathcal{M}_k$. The stronger the correlation, the more information the covariate may contribute to the IPD meta-analysis. In randomized trials, although such a correlation should be zero, the effect sizes may still be correlated if the interaction term is included in statistical modeling. This has been observed in our simulation studies. The correlations in the second component matrix of $\mathcal{M}_k$ may result from the link between the treatment effect and the baseline risk. Such a link is not unusual and often modeled in meta-analysis of clinical trials; see McIntosh (1996); Guolo (2013); Ghidey et al. (2013).

Our investigation has focused on the standard random-effects models. Our preliminary study suggests that the conclusion may also hold for mixed-effects models, where the nuisance coefficients are treated as fixed effects. But a further study is needed. In the literature, other meta-analysis methods have been proposed for the case where the number of available studies is moderate or small (Follmann and Proschan, 1999; Liu et al., 2018). For such specific methods, it remains unknown whether or not our conclusion still holds regarding the relative efficiency of using summary statistics versus individual data. Further research is also needed to address a much weaker condition; i.e., the total sample size of all the studies diverge while n_k could be bounded for some studies. For example, there may be a few very large studies and many others are much smaller.

Our discussion has addressed one of many relative benefits of meta-analysis of IPD over summary statistics. Another crucial benefit is that the IPD analysis can avoid the so-called "ecological fallacy" (Reade et al., 2008; Cooper and Patall, 2009). When study-level moderators (e.g., average patient characteristics) are available, meta-regression is commonly used in practice to examine the relationship between the treatment effect and the moderators across studies. This relationship should not be confused with that relationship within studies. The confusion will result in aggregation bias and incorrect conclusions (Thompson and Higgins, 2002). The only way to eliminate this ecological fallacy is to

model the within-study relationship appropriately using individual-level data.

Acknowledgement

Chen's research is partially supported by the South Africa National Research Foundation (NRF) and South Africa Medical Research Council (SAMRC) (South Africa DST-NRF-SAMRC SARChI Research Chair in Biostatistics, Grant number 114613). Zhang's research is partially supported by the National Institutes of Health (R01 MH116527 and R01 HG010171) and the National Science Foundation (DMS-1722544) of the United States of America. The authors are grateful to the Co-Editor, the Associate Editor, and two anonymous referees for their constructive comments.

Data Availability Statement

The data used in Section 4 are available in Mendeley (Huh et al., 2019b) (https://data.mendeley.com/datasets/4dw4kn97fz/2).

References

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

Burke, D. L., Ensor, J., and Riley, R. D. (2017), "Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ," *Statistics in medicine*, 36, 855–875.

Chen, D. G. and Peace, K. E. (2013), *Applied Meta-Analysis Using R.*, Chapman & Hall/CRC: Boca Raton, FL, USA.

Chen, X. and Xie, M.-g. (2014), "A split-and-conquer approach for analysis of extraordinarily large data," *Statistica Sinica*, 24, 1655–1684.

Cheung, M. W.-L. and Jak, S. (2016), "Analyzing big data in psychology: a split/analyze/meta-analyze approach," *Frontiers in Psychology*, 7, 738.

Cooper, H. and Patall, E. (2009), "The relative benefits of meta-analysis conducted with individual participant data versus aggregated data," *Psychological Methods*, 14, 165–176.

dbGaP (2020), The database of Genotypes and Phenotypes for archiveing and distributing the data and results from studies that have investigated the interaction of genotype and phenotype in Humans, https://www.ncbi.nlm.nih.gov/gap (accessed January 1, 2020).

Follmann, D. A. and Proschan, M. A. (1999), "Valid inference in random effects meta-analysis," *Biometrics*, 55, 732–737.

Ghidey, W., Stijnen, T., and van Houwelingen, H. C. (2013), "Modelling the effect of baseline risk in meta-analysis: a review from the perspective of errors-in-variables regression," *Statistical Methods in Medical Research*, 22, 307–323.

Guolo, A. (2013), "Flexibly modeling the baseline risk in meta-analysis," *Statistics in Medicine*, 32, 40–50.

Huh, D., Mun, E., Walters, S. T., Zhou, Z., and Atkins, D. C. (2019a), "A tutorial on individual participant data meta-analysis using Bayesian multilevel modeling to estimate alcohol intervention effects across heterogeneous studies," *Addictive Behaviors*, 94, 162–170.

Jordan, M. I. et al. (2013), "On statistics, computation and scalability," *Bernoulli*, 19, 1378–1390.

Kass, R. E., Tierney, L., and Kadane, J. B. (1990), "The validity of posterior expansions based on Laplace's method," in *Bayesian and Likelihood Methods in Statistics and Econometrics*, eds. Geisser, S., Hodges, J., Press, S., and Zellner, A., Elsevier Science Publishers B.V. (North-Holland), pp. 473–488.

Kontopantelis, E. (2018), "A comparison of one-stage vs two-stage individual patient data meta-analysis methods: A simulation study," *Research Synthesis Methods*, 9, 417–430.

Lee, J. Y., Brown, J. J., and Ryan, L. M. (2017), "Sufficiency revisited: Rethinking statistical algorithms in the big data era," *The American Statistician*, 71, 202–208.

Lin, D. Y. and Zeng, D. (2010), "On the relative efficiency of the using summary statistics versus individual-level data in meta-analysis," *Biometrika*, 97(2), 321–332.

Liu, D., Liu, R. Y., and Xie, M. (2015), "Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness," *Journal of the American Statistical Association*, 110, 326–340.

Liu, S., Tian, L., Lee, S., and Xie, M.-g. (2018), "Exact inference on meta-analysis with generalized fixed-effects and random-effects models," *Biostatistics & Epidemiology*, 2, 1–22.

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

McIntosh, M. W. (1996), "The population risk as an explanatory variable in research synthesis of clinical trials," *Statistics in Medicine*, 15, 1713–1728.

Olkin, I. and Sampson, A. (1998), "Comparison of meta-analysis versus analysis of variance of individual patient data." *Biometrics*, 54, 317–322.

Pigott, T. D. (2012), *Advances in Meta-Analysis*, Springer-Verlag: New York, NY, USA.

Reade, M. C., Delaney, A., Bailey, M. J., and Angus, D. C. (2008), "Bench-to-bedside review: Avoiding pitfalls in critical care meta-analysis - funnel plots, risk estimates, types of heterogeneity, baseline risk and the ecologic fallacy," *Critical Care*, 12, 1–8.

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.

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

Thompson, S. G. and Higgins, J. P. (2002), "How should meta-regression analyses be undertaken and interpreted?" *Statistics in Medicine*, 21, 1559–1573.

Whitehead, A. (2003), *Meta-Analysis of Controlled Clinical Trials*, John Wiley & Sons, Ltd.: Chichester, UK.

Zeng, D. and Lin, D. (2015), "On random-effects meta-analysis," *Biometrika*, 102, 281–294.

Huh, D., Mun, E., Walters, S. T., Zhou, Z., and Atkins, D. C. (2019b), Data and code for: A tutorial on individual participant data meta-analysis using Bayesian multilevel modeling to estimate alcohol intervention effects across heterogeneous studies (2019), Mendeley Data, v2, http://dx.doi.org/10.17632/4dw4kn97fz.2 (accessed January 1, 2020).

Supporting Information

Web Appendices, Proofs, and Tables referenced in Sections 2 and 3 are available with this paper at the Biometrics website on Wiley Online Library.

Table 1: The IPD- and summary-statistics-based meta-analysis estimates of the interaction effect β_3 for continuous outcomes when $\rho_{x,k} \sim \text{Uniform}(-0.3, 0.3)$ (with known variance components $\sigma_\epsilon = 10$ and Σ_β).

		IPD					Summary statistics				
K	n_k	Mean	SE	ESE	$\sqrt{\text{MSE}}$	CP	Mean	SE	ESE	$\sqrt{\text{MSE}}$	CP
5	20	1.060	2.131	2.111	2.131	0.951	1.045	2.384	2.358	2.383	0.949
	50	1.034	1.338	1.314	1.338	0.943	1.030	1.402	1.379	1.402	0.949
	100	1.003	0.879	0.935	0.879	0.962	1.005	0.912	0.957	0.912	0.955
	200	0.962	0.685	0.680	0.686	0.951	0.964	0.689	0.687	0.690	0.954
	500	1.001	0.463	0.461	0.463	0.941	1.000	0.466	0.463	0.466	0.941
10	20	1.040	1.497	1.461	1.496	0.948	0.981	1.701	1.669	1.700	0.949
	50	0.989	0.934	0.922	0.933	0.955	0.990	0.978	0.963	0.977	0.953
	100	1.019	0.647	0.661	0.647	0.961	1.018	0.668	0.676	0.667	0.959
	200	0.990	0.485	0.478	0.485	0.941	0.993	0.489	0.483	0.489	0.941
	500	1.001	0.331	0.326	0.331	0.946	1.001	0.334	0.328	0.333	0.947
30	20	0.984	0.816	0.834	0.816	0.952	1.001	0.947	0.954	0.947	0.950
	50	1.001	0.525	0.530	0.525	0.947	0.997	0.553	0.556	0.552	0.957
	100	0.999	0.390	0.379	0.389	0.942	1.003	0.398	0.389	0.398	0.945
	200	0.994	0.285	0.276	0.285	0.942	0.994	0.288	0.279	0.287	0.950
	500	0.999	0.188	0.188	0.188	0.955	0.999	0.189	0.189	0.189	0.956
50	20	1.008	0.672	0.645	0.672	0.941	1.005	0.771	0.738	0.770	0.936
	50	0.999	0.400	0.411	0.400	0.960	1.008	0.424	0.432	0.424	0.955
	100	1.000	0.296	0.294	0.296	0.944	1.002	0.307	0.302	0.307	0.944
	200	1.007	0.214	0.214	0.214	0.944	1.005	0.214	0.216	0.214	0.948
	500	0.999	0.144	0.146	0.144	0.954	0.999	0.145	0.146	0.145	0.954
100	20	0.984	0.450	0.454	0.450	0.955	0.981	0.525	0.521	0.525	0.945
	50	0.999	0.277	0.290	0.277	0.965	1.004	0.292	0.305	0.292	0.966
	100	1.011	0.217	0.208	0.217	0.941	1.011	0.223	0.213	0.224	0.936
	200	1.003	0.153	0.151	0.153	0.947	1.003	0.155	0.153	0.155	0.946
	500	1.000	0.102	0.103	0.102	0.946	1.000	0.103	0.104	0.103	0.951

Mean– the average of estimates for 1000 simulation replicates; SE–standard error of the estimates from 1000 simulation runs; ESE–estimated standard error of the estimate for each simulation run; MSE–mean squared error; CP–coverage probability of 95% confidence intervals.

Table 2: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for continuous outcomes when $\rho_{x,k} \sim$ Uniform$(-0.3, 0.3)$ (with known variance components σ_ϵ and Σ_β).

σ_ϵ	$n_k \setminus K$	5	10	30	50	100
1	20	1.069	1.069	1.078	1.084	1.107
	50	1.007	1.013	1.010	1.012	1.000
	100	1.001	1.010	1.003	1.002	1.005
	200	1.000	1.001	1.001	1.000	0.999
	500	1.000	1.000	1.000	1.000	1.000
3	20	1.191	1.191	1.221	1.200	1.254
	50	1.043	1.046	1.054	1.054	1.049
	100	1.021	1.034	1.018	1.022	1.029
	200	0.998	1.004	1.005	1.001	1.001
	500	0.999	1.001	1.001	1.002	1.004
10	20	1.251	1.291	1.346	1.315	1.361
	50	1.098	1.096	1.108	1.123	1.110
	100	1.078	1.063	1.045	1.074	1.057
	200	1.012	1.017	1.019	1.005	1.025
	500	1.014	1.013	1.014	1.008	1.018
30	20	1.255	1.313	1.369	1.344	1.379
	50	1.115	1.120	1.122	1.141	1.123
	100	1.102	1.077	1.050	1.104	1.069
	200	1.021	1.029	1.029	1.021	1.035
	500	1.038	1.028	1.034	1.030	1.029

Table 3: The IPD- and summary-statistics-based meta-analysis estimates of the interaction effect β_3 for binary outcomes when $\rho_{x,k} = 0$.

		IPD					Summary statistics				
K	n_k	Mean	SE	ESE	$\sqrt{\text{MSE}}$	CP	Mean	SE	ESE	$\sqrt{\text{MSE}}$	CP
10	100	1.011	0.841	0.929	0.841	0.962	1.021	0.875	0.946	0.875	0.960
	200	0.989	0.590	0.658	0.589	0.960	0.989	0.600	0.664	0.600	0.958
	500	0.990	0.404	0.425	0.404	0.955	0.986	0.406	0.426	0.406	0.954
	1,000	0.998	0.302	0.317	0.302	0.958	0.993	0.302	0.317	0.302	0.957
	2,000	1.004	0.238	0.246	0.238	0.944	1.000	0.239	0.246	0.238	0.946
30	100	1.004	0.488	0.520	0.488	0.973	1.010	0.505	0.531	0.505	0.964
	200	1.017	0.345	0.367	0.346	0.963	1.019	0.352	0.371	0.352	0.962
	500	1.008	0.240	0.238	0.240	0.949	1.003	0.241	0.239	0.241	0.946
	1,000	1.013	0.170	0.179	0.171	0.960	1.006	0.170	0.179	0.170	0.961
	2,000	0.998	0.138	0.139	0.138	0.929	0.991	0.138	0.139	0.138	0.930
50	100	0.985	0.381	0.398	0.381	0.964	0.990	0.396	0.407	0.396	0.959
	200	1.018	0.262	0.281	0.262	0.968	1.018	0.266	0.283	0.266	0.967
	500	1.004	0.183	0.183	0.183	0.948	0.997	0.184	0.183	0.184	0.949
	1,000	1.009	0.133	0.137	0.133	0.957	1.001	0.133	0.137	0.133	0.956
	2,000	0.995	0.110	0.108	0.110	0.952	0.988	0.110	0.108	0.110	0.948
100	100	1.020	0.261	0.276	0.262	0.961	1.031	0.273	0.282	0.275	0.955
	200	1.000	0.193	0.196	0.193	0.954	0.999	0.196	0.198	0.196	0.952
	500	0.997	0.128	0.128	0.128	0.948	0.990	0.128	0.128	0.128	0.949
	1,000	1.001	0.099	0.096	0.099	0.930	0.993	0.100	0.096	0.100	0.933
	2,000	1.001	0.074	0.076	0.074	0.962	0.994	0.073	0.076	0.074	0.965

Mean– the average of estimates for 1000 simulation replicates; SE–standard error of the estimates from 1000 simulation runs; ESE–estimated standard error of the estimate for each simulation run; MSE–mean squared error; CP–coverage probability of 95% confidence intervals.

Table 4: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for binary outcomes.

Correlation	$n_k \setminus K$	10	30	50	100
$\rho_x = 0$	100	1.083	1.071	1.080	1.100
	200	1.036	1.038	1.030	1.037
	500	1.011	1.009	1.010	1.010
	1,000	1.004	0.996	0.994	1.007
	2,000	1.001	1.002	1.003	0.998
$\rho_x \sim$ Unif(-0.3,0.3)	100	1.066	1.092	1.109	1.087
	200	1.037	1.026	1.058	1.066
	500	1.015	1.005	0.989	0.997
	1,000	1.001	1.018	0.999	1.006
	2,000	0.994	0.991	1.003	0.992

Table 5: The IPD- and summary-statistics-based meta-analysis of nine independent studies on the alcohol interventions.

	IPD			Summary Statistics		
Parameter	Estimate	SE	P-value	Estimate	SE	P-value
Linear regression models for $Y^{(1)}$ (the change in the number of drinks)						
β_1	-0.087	0.092	0.346	-0.088	0.095	0.354
β_2	-0.389	0.035	1.60×10^{-28}	-0.401	0.039	2.49×10^{-25}
β_3	0.025	0.033	0.455	0.015	0.037	0.693
Logistic regression models for $Y^{(2)}$ (whether the number of drinks was reduced)						
β_1	0.396	0.206	0.054	0.187	0.218	0.391
β_2	0.586	0.048	5.13×10^{-34}	0.531	0.052	8.49×10^{-25}
β_3	-0.074	0.049	0.128	-0.023	0.052	0.652

The analysis is based on the following nine studies in Huh et al. (2019a) (control and treatment sample sizes given in parentheses): Study 2 (102, 92), Study 8a (519, 512), Study 8b (754, 719), Study 8c (147, 127), Study 9 (91, 92), Study 11 (160, 150), Study 13/14 (24, 27), Study 18 (99, 93), Study 21 (70, 63).

Supporting Information for "Relative efficiency of using summary versus individual data in random-effects meta-analysis" by Ding-Geng Chen, Dungang Liu, Xiaoyi Min and Heping Zhang

A Conditions for the Laplace approximation

For simplicity, we denote $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\eta}^\top)^\top$ and $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k^\top, \boldsymbol{\eta}_k^\top)^\top$, let p_k be the dimension and $\boldsymbol{\Theta}_k$ be the domain of $\boldsymbol{\theta}_k$.

(C1) The parameter $\boldsymbol{\theta}_k$ is identifiable. In other words, for any $\boldsymbol{\theta}_k \neq \boldsymbol{\theta}_k'$ in $\boldsymbol{\Theta}_k$, $f_k(y_k, \boldsymbol{x}_k; \boldsymbol{\theta}_k) \neq f_k(y_k, \boldsymbol{x}_k; \boldsymbol{\theta}_k')$ for some $(y_k, \boldsymbol{x}_k)$;

(C2) For all $(y_k, \boldsymbol{x}_k)$, the function $f_k(y_k, \boldsymbol{x}_k; \boldsymbol{\theta}_k)$ is a three times continuous differentiable function of $\boldsymbol{\theta}_k$ and is positive for all $\boldsymbol{\theta}_k$;

(C3) For all $\boldsymbol{\theta}_k^* \in \boldsymbol{\Theta}_k$, there exist a neighborhood $\mathcal{N}_1(\boldsymbol{\theta}_k^*)$, a positive number n_k, and a random variable Z_1 such that $E_{\boldsymbol{\theta}_k^*}(Z_1) < \infty$ and for all $\boldsymbol{\theta}_k \in \mathcal{N}_1(\boldsymbol{\theta}_k^*)$,

$$\frac{1}{n_k} \sum_{i=1}^{n_k} \log \frac{f_k(y_{ki}, \boldsymbol{x}_{ki}; \boldsymbol{\theta}_k)}{f_k(y_{ki}, \boldsymbol{x}_{ki}; \boldsymbol{\theta}_k^*)} < Z_1;$$

(C4) For all $\boldsymbol{\theta}_k^* \in \boldsymbol{\Theta}_k$, there exist a neighborhood $\mathcal{N}_2(\boldsymbol{\theta}_k^*)$ and a random variable Z_2 such that $E_{\boldsymbol{\theta}_k^*}(Z_2) < \infty$ and for all $\boldsymbol{\theta}_k \in \mathcal{N}_2(\boldsymbol{\theta}_k^*)$, all $1 \leq d \leq 3$, and all $1 \leq j_i, \ldots, j_d \leq p_k$, these is

$$\left| \frac{\partial^d \log f_k(y_k, \boldsymbol{x}_k; \boldsymbol{\theta}_k)}{\partial \theta_{kj_1} \cdots \partial \theta_{kj_d}} \right| < Z_2;$$

(C5) For all $\boldsymbol{\theta}_k^* \in \boldsymbol{\Theta}_k$, define M to be the Hessian matrix of $E_{\boldsymbol{\theta}_k^*}[\log f_k(Y_k, \boldsymbol{X}_k; \boldsymbol{\theta}_k^*) - \log f_k(Y_k, \boldsymbol{X}_k; \boldsymbol{\theta}_k)]$, then $\det(M) > 0$;

(C6) For all $\boldsymbol{\theta}_k^* \in \boldsymbol{\Theta}_k$, the maximum likelihood estimate is strongly consistent.

(C7) $\boldsymbol{\theta}$ lies in the interior of a compact set within the parameter space.

B Proofs

Proof of Lemma 1. According to Theorems 7 and 8 in Kass et al. (1990), under conditions (C1)-(C6) in Appendix A, the sequence of log-likelihood functions $\{\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k), n_k = 1, \ldots\}$ is "Laplace regular" with probability one for any true value $\boldsymbol{\theta}_k^*$ in $\boldsymbol{\Theta}$. The proof of Lemma 1 is then similar to Theorem 1 of Kass et al. (1990). In particular, we only need to consider the integration in equation (2.4) over $B_\delta(\hat{\boldsymbol{\theta}}_k) \subseteq \boldsymbol{\Theta}$ for any $0 < \delta < \delta_0$, where $B_\delta(\hat{\boldsymbol{\theta}}_k)$ is the open ball of radius δ centered at $\hat{\boldsymbol{\theta}}_k$. In equation (2.4), we expand $\ell(\boldsymbol{\theta})$ around $\hat{\boldsymbol{\theta}}_k$ to the third order and keep the other terms unchanged. With arguments similar to those in Kass et al. (1990), the third order expansion from $\ell(\boldsymbol{\theta})$ leads to the following leading term of (2.4) and an error term of $O(n_k^{-1/2})$-order:

$$
\begin{aligned}
&\ell_k(\hat{\boldsymbol{\theta}}_k) - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \log\int_{B_\delta(\hat{\boldsymbol{\theta}}_k)} \exp[-\frac{1}{2}(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k)^\top \mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k) - \frac{1}{2}(\boldsymbol{\theta}_k - \boldsymbol{\theta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_k - \boldsymbol{\theta})] d\boldsymbol{\theta}_k \\
&= \log\int_{B_\delta(\hat{\boldsymbol{\theta}}_k)} \exp\{-\frac{1}{2}(\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_k)^\top [\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k) + \boldsymbol{\Sigma}^{-1}](\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_k) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_k)^\top [\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1} + \boldsymbol{\Sigma}]^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_k)\} d\boldsymbol{\theta}_k \\
&+ \ell_k(\hat{\boldsymbol{\theta}}_k) - \frac{1}{2}\log|\boldsymbol{\Sigma}| \\
&\approx -\frac{1}{2}\log|\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k) + \boldsymbol{\Sigma}^{-1}| - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_k)^\top [\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1} + \boldsymbol{\Sigma}]^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_k) + \ell_k(\hat{\boldsymbol{\theta}}_k) - \frac{1}{2}\log|\boldsymbol{\Sigma}|
\end{aligned}
\tag{B.2}
$$

where

$$
\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{\Xi} \\ \boldsymbol{\Xi}^\top & \boldsymbol{\Phi} \end{pmatrix},
$$

and

$$
\tilde{\boldsymbol{\theta}}_k = [\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k) + \boldsymbol{\Sigma}^{-1}]^{-1}[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)\hat{\boldsymbol{\theta}}_k + \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}].
$$

Since $\tilde{\boldsymbol{\theta}}_k \in B_\delta(\hat{\boldsymbol{\theta}}_k)$ when n_k is large enough, we can expand the range of integration to the whole parameter space in (B.2) as Kass et al. (1990), which leads to the approximation with an error of exponential decreasing order. Lemma 1 is proved. □

Proof of Lemma 2. Denote the leading term in approximation (2.6) as $\tilde{\ell}_k(\boldsymbol{\theta}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi})$, then

$$\begin{aligned}
& \sum_{k=1}^{K} \tilde{\ell}_k(\boldsymbol{\theta}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi}) \\
= \; & -\frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_k)^\top[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1}+\boldsymbol{\Sigma}]^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_k) - \frac{1}{2}\sum_{k=1}^{K}\log|\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)+\boldsymbol{\Sigma}^{-1}| + C \\
= \; & -\frac{1}{2}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})^\top\{\sum_{k=1}^{K}[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1}+\boldsymbol{\Sigma}]^{-1}\}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}) - \frac{1}{2}\sum_{k=1}^{K}\hat{\boldsymbol{\theta}}_k^\top[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1}+\boldsymbol{\Sigma}]^{-1}\hat{\boldsymbol{\theta}}_k \\
+ \; & \frac{1}{2}\{\sum_{k=1}^{K}[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1}+\boldsymbol{\Sigma}]^{-1}\hat{\boldsymbol{\theta}}_k\}^\top\{\sum_{k=1}^{K}[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1}+\boldsymbol{\Sigma}]^{-1}\}^{-1}\{\sum_{k=1}^{K}[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1}+\boldsymbol{\Sigma}]^{-1}\hat{\boldsymbol{\theta}}_k\} \\
- \; & \frac{1}{2}\sum_{k=1}^{K}\log|\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)+\boldsymbol{\Sigma}^{-1}| + C, \qquad \text{(B.3)}
\end{aligned}$$

where $\tilde{\boldsymbol{\theta}} = \{\sum_{k=1}^{K}[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1}+\boldsymbol{\Sigma}]^{-1}\}^{-1}\{\sum_{k=1}^{K}[\mathcal{I}_k(\hat{\boldsymbol{\theta}}_k)^{-1}+\boldsymbol{\Sigma}]^{-1}\hat{\boldsymbol{\theta}}_k\}$. Thus, the leading term in (2.7) maximizes $\sum_{k=1}^{K}\tilde{\ell}_k(\boldsymbol{\theta}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi})$. (2.7) can be proved by noticing that for any $\epsilon > 0$, for any sequence $\{\boldsymbol{\theta}^{(n)}\}$ such that $||\boldsymbol{\theta}^{(n)} - \tilde{\boldsymbol{\theta}}^{(n)}|| > \epsilon$, $\sum_{k=1}^{K}\ell_k(\tilde{\boldsymbol{\theta}}^{(n)}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi}) - \sum_{k=1}^{K}\ell_k(\boldsymbol{\theta}^{(n)}, \boldsymbol{T}, \boldsymbol{\Phi}, \boldsymbol{\Xi})$ is greater than 0 with probability approaching 1. □

Proof of Theorem 1. To prove Part (a), it suffices to show that

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{SS}\right) \geq \widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{IPD}\right), \qquad \text{(B.4)}$$

i.e., the matrix $\widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{SS}\right) - \widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{IPD}\right)$ is positive semi-definite. After some algebraic calculations, we can show that

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{SS}\right) = \left(\sum_{k=1}^{K} A_k^{-1}\right)^{-1}$$

and

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{IPD}\right) = \left\{ \sum_{k=1}^{K} \begin{pmatrix} A_k & B_k \\ B_k^\top & C_k \end{pmatrix}^{-1} \right\}^{-1}_{[A_k]},$$

where

$$A_k = \left(\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} - \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\mathcal{I}^{-1}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k}\right)^{-1}_{|\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\eta}}_k} + \hat{\boldsymbol{T}},$$

$$B_k = -\mathcal{I}^{-1}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\left(\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k} - \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k}\mathcal{I}^{-1}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\right)^{-1}_{|\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\eta}}_k} + \hat{\boldsymbol{\Xi}},$$

$$C_k = \left(\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k} - \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k}\mathcal{I}^{-1}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\right)^{-1}_{|\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\eta}}_k} + \hat{\boldsymbol{\Phi}}.$$

If we let

$$\begin{pmatrix} A_k & B_k \\ B_k^\top & C_k \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{A}_k & \tilde{B}_k \\ \tilde{B}_k^\top & \tilde{C}_k \end{pmatrix},$$

then the inequality (B.4) is equivalent to

$$\left\{ \sum_{k=1}^{K} \left(\tilde{A}_k - \tilde{B}_k\tilde{C}_k^{-1}\tilde{B}_k^\top\right) \right\}^{-1} \geq \left\{ \sum_{k=1}^{K}\tilde{A}_k - \sum_{k=1}^{K}\tilde{B}_k\left(\sum_{k=1}^{K}\tilde{C}_k\right)^{-1}\sum_{k=1}^{K}\tilde{B}_k^\top \right\}^{-1}.$$

The above inequality could be further simplified as

$$\sum_{k=1}^{K}\tilde{B}_k\tilde{C}_k^{-1}\tilde{B}_k^\top \geq \sum_{k=1}^{K}\tilde{B}_k\left(\sum_{k=1}^{K}\tilde{C}_k\right)^{-1}\sum_{k=1}^{K}\tilde{B}_k^\top,$$

which holds according to Lemma 1 in the Appendix A of Lin and Zeng (2010). This in turn establishes the inequality (B.4).

The proof of the inequality in Part (b) is similar to that of Part (a) except that we use the true values of the parameters $\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{T}, \boldsymbol{\Xi}, \boldsymbol{\Phi}$ to evaluate A_k, B_k, C_k. The equality is achieved if and only if

$$\tilde{B}_1\tilde{C}_1^{-1} = \tilde{B}_2\tilde{C}_2^{-1} = \cdots = \tilde{B}_K\tilde{C}_K^{-1}.$$

Since $\tilde{B}_k = -A_k^{-1}B_k(C_k - B_k^\top A_k^{-1}B_k)^{-1}$ and $\tilde{C}_k^{-1} = C_k - B_k^\top A_k^{-1}B_k$, the above condition is equivalent to

$$A_1^{-1}B_1 = A_2^{-1}B_2 = \cdots = A_K^{-1}B_K.$$

Therefore, given $\boldsymbol{\Xi} = \mathbf{0}$, the equality can be achieved only if the quantity

$$\left\{\left(\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} - \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}^{-1}\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k}\right)^{-1} + \boldsymbol{T}\right\}^{-1} \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k}^{-1}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\left(\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k} - \mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k}^{-1}\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\right)^{-1}$$

does not depend on the study index k. This condition will not hold except in special cases. Therefore, the inequality may hold even if $\boldsymbol{\Xi} = \mathbf{0}$.

Following the arguments in the proof of Part (b), the equality in Part (c) holds. To see this, notice that $A_i^{-1}B_i \to \boldsymbol{T}^{-1}\boldsymbol{\Xi}$ as the sample size in each study $n \to \infty$. Moreover, $|A_i^{-1}B_i - A_j^{-1}B_j| = O(1/n)$ for any $i \neq j$. Thus, as $n \to \infty$, $K \to \infty$ and $Kn^{-1/2} \to 0$, $\sum_{i\neq j}|A_i^{-1}B_i - A_j^{-1}B_j| \to 0$. This guarantees that the equality is achieved in the limit. □

C A special case of Section 2.3

We consider a special case where the likelihood function in (2.5) is exact in the sense that there is no approximation error term $O(n_k^{-\frac{1}{2}})$. Specifically, assume that in the k-th study, the log-likelihood function is of the following form

$$\ell_k(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k) = -\begin{pmatrix}\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta}_k - \hat{\boldsymbol{\eta}}_k\end{pmatrix}^\top \begin{pmatrix}\mathcal{I}_{k,11} & \mathcal{I}_{k,12} \\ \mathcal{I}_{k,21} & \mathcal{I}_{k,22}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta}_k - \hat{\boldsymbol{\eta}}_k\end{pmatrix} + c_k,$$

where c_k is a statistic that does not depend on any parameters. For clarity, we assume that all the variance components, including the within-study covariance $\mathcal{I}_k$ and the between-study variance $\boldsymbol{T}$, are known.

For this special case, the summary statistic

$$\hat{\boldsymbol{\beta}}_k \sim N\left(\boldsymbol{\beta}_k, \mathrm{var}_C(\hat{\boldsymbol{\beta}}_k) + \boldsymbol{T}\right),$$

where the exact variance $\mathrm{var}_C(\hat{\boldsymbol{\beta}}_k) = \left(\mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\beta}_k} - \mathcal{I}_{k,\boldsymbol{\beta}_k\boldsymbol{\eta}_k}\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\eta}_k}^{-1}\mathcal{I}_{k,\boldsymbol{\eta}_k\boldsymbol{\beta}_k}\right)^{-1}$. The combined

estimator is

$$\hat{\boldsymbol{\beta}}_{SS} = \left[\sum_{k=1}^{K} \left\{ \mathrm{var}_C(\hat{\boldsymbol{\beta}}_k) + \boldsymbol{T} \right\}^{-1}\right]^{-1} \sum_{k=1}^{K} \left\{ \mathrm{var}_C(\hat{\boldsymbol{\beta}}_k) + \boldsymbol{T} \right\}^{-1} \hat{\boldsymbol{\beta}}_k.$$

The exact variance of $\hat{\boldsymbol{\beta}}_{SS}$ is

$$\mathrm{var}(\hat{\boldsymbol{\beta}}_{SS}) = \left[\sum_{k=1}^{K} \left\{ \left(\mathcal{I}_{k,11} - \mathcal{I}_{k,12}\mathcal{I}_{k,22}^{-1}\mathcal{I}_{k,21}\right)^{-1} + \boldsymbol{T} \right\}^{-1}\right]^{-1}.$$

To compare $\mathrm{var}(\hat{\boldsymbol{\beta}}_{SS})$ with $\mathrm{var}(\hat{\boldsymbol{\beta}}_{IPD})$, we derive the log-likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\eta})$ in (2.6), which simplifies to

$$\ell_k(\boldsymbol{\beta}, \boldsymbol{\eta}) = -\begin{pmatrix} \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_k \end{pmatrix}^{\top} \left[\begin{pmatrix} \mathcal{I}_{k,11} & \mathcal{I}_{k,12} \\ \mathcal{I}_{k,21} & \mathcal{I}_{k,22} \end{pmatrix}^{-1} + \begin{pmatrix} \boldsymbol{T} & \boldsymbol{\Xi} \\ \boldsymbol{\Xi}^{\top} & \boldsymbol{\Phi} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k \\ \boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_k \end{pmatrix} + c_k.$$

Therefore, the IPD estimator is

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{IPD} \\ \hat{\boldsymbol{\eta}}_{IPD} \end{pmatrix} = \left\{ \left[\sum_{k=1}^{K} \mathcal{M}_k^{-1}\right]^{-1} \left[\sum_{k=1}^{K} \mathcal{M}_k^{-1} \begin{pmatrix} \hat{\boldsymbol{\beta}}_k \\ \hat{\boldsymbol{\eta}}_k \end{pmatrix}\right] \right\},$$

where

$$\mathcal{M}_k = \begin{pmatrix} \mathcal{I}_{k,11} & \mathcal{I}_{k,12} \\ \mathcal{I}_{k,21} & \mathcal{I}_{k,22} \end{pmatrix}^{-1} + \begin{pmatrix} \boldsymbol{T} & \boldsymbol{\Xi} \\ \boldsymbol{\Xi}^{\top} & \boldsymbol{\Phi} \end{pmatrix}.$$

The exact variance of $\hat{\boldsymbol{\beta}}_{IPD}$ is

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}_{IPD}\right) = \left\{ \sum_{k=1}^{K} \mathcal{M}_k^{-1} \right\}^{-1}_{[1,1]}.$$

Similar to the proof of Theorem 1, we can show that

$$\mathrm{var}(\hat{\boldsymbol{\beta}}_{SS}) \geq \mathrm{var}\left(\hat{\boldsymbol{\beta}}_{IPD}\right).$$

The equality can be achieved if (a) the within-study correlation $\mathcal{I}_{k,12} = \mathbf{0}$ and the between-study correlation $\boldsymbol{\Xi} = \mathbf{0}$; or (b) $\mathcal{M}_k$ is the same for all k's. The above inequality is

established provided that all variance components are known. Given estimates of variance components, the inequality still holds in the settings considered in our simulation studies.

The result in this subsection indicates that we can establish an inequality between the exact variances of $\hat{\boldsymbol{\beta}}_{SS}$ and $\hat{\boldsymbol{\beta}}_{IPD}$, when each individual log-likelihood function can be written as a quadratic form of $(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$ without an approximation error. For the general case where approximation errors may exist, the inequalities in Theorem 1(a)-(b) are established between the variance estimates or between the asymptotic variances.

D Supplementary tables for simulation studies

Table 6: The IPD- and summary-statistics-based meta-analysis estimates of the interaction effect β_3 for continuous outcomes (with estimated variance components $\sigma_\epsilon = 10$ and Σ_β).

		IPD					Summary statistics				
K	n_k	Mean	SE	ESE	$\sqrt{\text{MSE}}$	CP	Mean	SE	ESE	$\sqrt{\text{MSE}}$	CP
5	20	1.058	2.184	2.295	2.184	0.959	1.043	2.393	2.503	2.392	0.952
	50	1.030	1.354	1.440	1.354	0.958	1.030	1.401	1.499	1.400	0.961
	100	1.003	0.885	1.022	0.885	0.971	1.003	0.913	1.044	0.913	0.971
	200	0.963	0.683	0.743	0.684	0.968	0.965	0.689	0.752	0.689	0.965
	500	1.001	0.466	0.488	0.466	0.950	1.000	0.466	0.491	0.466	0.953
10	20	1.021	1.521	1.593	1.520	0.953	0.966	1.699	1.775	1.699	0.959
	50	0.991	0.939	0.998	0.939	0.961	0.990	0.977	1.037	0.976	0.958
	100	1.020	0.653	0.717	0.653	0.969	1.017	0.669	0.732	0.669	0.969
	200	0.991	0.485	0.516	0.485	0.956	0.993	0.490	0.521	0.489	0.953
	500	1.001	0.333	0.343	0.332	0.947	1.001	0.333	0.345	0.333	0.952
30	20	0.984	0.823	0.881	0.823	0.958	1.000	0.946	0.991	0.945	0.954
	50	1.002	0.525	0.560	0.524	0.962	0.996	0.553	0.584	0.552	0.962
	100	1.001	0.391	0.400	0.390	0.944	1.004	0.398	0.409	0.398	0.954
	200	0.994	0.286	0.289	0.285	0.944	0.994	0.288	0.293	0.288	0.949
	500	1.000	0.188	0.194	0.188	0.951	0.999	0.189	0.195	0.189	0.954
50	20	1.008	0.676	0.675	0.676	0.940	1.005	0.772	0.762	0.771	0.944
	50	0.998	0.403	0.429	0.402	0.961	1.009	0.425	0.450	0.425	0.962
	100	0.999	0.296	0.307	0.296	0.954	1.002	0.307	0.315	0.307	0.949
	200	1.007	0.215	0.222	0.215	0.957	1.005	0.214	0.224	0.214	0.955
	500	0.998	0.145	0.148	0.145	0.951	0.999	0.145	0.148	0.145	0.953
100	20	0.983	0.450	0.470	0.451	0.963	0.980	0.525	0.533	0.525	0.953
	50	1.000	0.278	0.299	0.278	0.968	1.003	0.292	0.314	0.292	0.971
	100	1.010	0.218	0.213	0.218	0.941	1.011	0.223	0.219	0.224	0.943
	200	1.002	0.153	0.155	0.153	0.953	1.003	0.155	0.157	0.155	0.949
	500	1.000	0.102	0.104	0.102	0.948	1.000	0.103	0.105	0.103	0.952

Mean– the average of estimates for 1000 simulation replicates; SE–standard error of the estimates from 1000 simulation runs; ESE–estimated standard error of the estimate for each simulation run; MSE–mean squared error; CP–coverage probability of 95% confidence intervals.

Table 7: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for continuous outcomes (with estimated variance components σ_ϵ and Σ_β) when $\rho_{x,k} \sim \text{Uniform}(-0.3, 0.3)$.

σ_ϵ	$n_k \setminus K$	5	10	30	50	100
1	20	1.033	1.042	1.065	1.080	1.102
	50	0.994	1.004	1.007	1.013	1.001
	100	1.001	1.005	1.003	1.002	1.005
	200	0.998	1.000	1.000	1.000	0.999
	500	1.000	1.000	1.000	1.000	1.000
3	20	1.142	1.142	1.184	1.180	1.254
	50	1.013	1.020	1.053	1.037	1.044
	100	1.010	1.020	1.013	1.012	1.031
	200	0.997	1.000	1.005	0.999	0.995
	500	0.994	0.997	1.000	1.000	1.004
10	20	1.200	1.249	1.319	1.302	1.360
	50	1.070	1.082	1.110	1.113	1.106
	100	1.064	1.049	1.040	1.075	1.056
	200	1.016	1.018	1.017	0.998	1.025
	500	1.000	1.005	1.015	1.003	1.014
30	20	1.208	1.271	1.345	1.328	1.376
	50	1.077	1.109	1.117	1.132	1.122
	100	1.085	1.069	1.045	1.099	1.070
	200	1.018	1.028	1.030	1.021	1.034
	500	1.027	1.022	1.030	1.026	1.029

Table 8: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for continuous outcomes when $\rho_{x,k} = 0$.

σ_ϵ	$n_k \setminus K$	5	10	30	50	100
1	20	1.051	1.024	1.047	1.023	1.039
	50	1.005	1.001	1.011	1.003	1.010
	100	1.000	1.005	1.001	1.003	1.003
	200	0.999	0.999	1.002	1.000	1.001
	500	1.000	1.000	1.000	1.000	1.000
3	20	1.164	1.134	1.174	1.134	1.141
	50	1.021	1.034	1.053	1.054	1.033
	100	1.007	1.010	0.989	1.017	1.004
	200	1.006	0.998	1.006	1.002	1.001
	500	1.002	1.002	0.998	1.002	1.002
10	20	1.227	1.260	1.288	1.269	1.257
	50	1.047	1.089	1.112	1.109	1.084
	100	1.034	1.022	1.018	1.028	1.024
	200	1.011	1.002	1.014	1.022	1.008
	500	1.002	1.004	0.994	1.013	1.003
30	20	1.229	1.276	1.309	1.303	1.282
	50	1.052	1.093	1.130	1.122	1.105
	100	1.042	1.035	1.036	1.043	1.033
	200	1.019	1.011	1.021	1.030	1.017
	500	1.002	1.010	0.995	1.015	1.006

Table 9: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for continuous outcomes when $\rho_{x,k} \sim$ Uniform$(0, 0.7)$.

σ_ϵ	$n_k \setminus K$	5	10	30	50	100
1	20	1.019	1.027	1.086	1.065	1.088
	50	0.997	1.015	1.014	1.015	1.000
	100	1.002	1.002	1.007	1.007	1.006
	200	1.000	0.998	1.001	0.999	0.999
	500	1.000	0.999	1.000	1.001	1.000
3	20	1.106	1.139	1.278	1.153	1.213
	50	1.031	1.047	1.075	1.056	1.031
	100	1.010	1.012	1.022	1.020	1.038
	200	0.998	0.999	0.998	0.983	1.000
	500	0.992	0.992	1.000	1.000	1.004
10	20	1.198	1.212	1.418	1.260	1.316
	50	1.096	1.094	1.133	1.117	1.096
	100	1.047	1.050	1.061	1.086	1.062
	200	1.010	1.005	1.024	1.005	1.022
	500	1.004	1.013	1.020	1.003	1.011
30	20	1.218	1.230	1.435	1.294	1.334
	50	1.113	1.115	1.145	1.129	1.113
	100	1.075	1.066	1.066	1.108	1.065
	200	1.020	1.022	1.041	1.025	1.034
	500	1.028	1.035	1.045	1.030	1.037

Table 10: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for continuous outcomes when the between-study correlation $\mathbf{\Xi} = \mathbf{0}$ and $\rho_{x,k} \sim$ Uniform$(-0.3, 0.3)$ (with known variance components $\sigma_\epsilon = 10$ and Σ_β).

σ_ϵ	$n_k \backslash K$	5	10	30	50	100
1	20	1.055	1.061	1.074	1.056	1.075
	50	1.006	1.008	1.006	1.006	1.000
	100	1.000	1.004	1.003	1.001	1.005
	200	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000
3	20	1.192	1.201	1.235	1.204	1.245
	50	1.038	1.043	1.055	1.051	1.048
	100	1.014	1.026	1.019	1.020	1.029
	200	0.998	1.001	1.002	0.997	1.004
	500	1.000	1.001	1.000	1.000	1.003
10	20	1.253	1.300	1.357	1.325	1.364
	50	1.097	1.104	1.115	1.126	1.115
	100	1.074	1.065	1.047	1.078	1.062
	200	1.009	1.017	1.015	1.003	1.023
	500	1.012	1.011	1.012	1.007	1.017
30	20	1.255	1.316	1.371	1.346	1.379
	50	1.115	1.122	1.124	1.141	1.124
	100	1.100	1.078	1.050	1.105	1.071
	200	1.021	1.030	1.028	1.022	1.032
	500	1.040	1.029	1.032	1.030	1.030

Table 11: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for continuous outcomes when the between-study correlation $\boldsymbol{\Xi} = \mathbf{0}$ and $\rho_{x,k} \sim \text{Uniform}(-0.3, 0.3)$ (with estimated variance components σ_ϵ and Σ_β).

σ_ϵ	$n_k \backslash K$	5	10	30	50	100
1	20	1.024	1.036	1.056	1.055	1.071
	50	0.992	1.000	1.005	1.008	1.001
	100	0.998	1.001	1.003	1.001	1.004
	200	1.000	0.999	1.000	1.000	1.001
	500	1.000	1.000	1.000	1.000	1.000
3	20	1.154	1.171	1.187	1.185	1.236
	50	0.999	1.021	1.053	1.038	1.042
	100	1.001	1.007	1.018	1.019	1.029
	200	0.996	0.997	1.000	0.996	1.002
	500	0.994	1.000	0.999	0.999	1.002
10	20	1.200	1.264	1.331	1.311	1.365
	50	1.063	1.088	1.115	1.114	1.109
	100	1.067	1.055	1.039	1.076	1.060
	200	1.014	1.018	1.016	1.001	1.022
	500	1.007	1.002	1.013	1.002	1.011
30	20	1.209	1.274	1.344	1.328	1.376
	50	1.079	1.109	1.118	1.131	1.123
	100	1.085	1.071	1.045	1.098	1.072
	200	1.019	1.031	1.028	1.021	1.034
	500	1.030	1.025	1.029	1.027	1.032

Table 12: The IPD- and summary-statistics-based meta-analysis estimates of the interaction effect β_3 for binary outcomes when 20% of the K studies have sample sizes $3n_k$.

		IPD					Summary statistics				
K	n_k	Mean	SE	ESE	$\sqrt{\text{MSE}}$	CP	Mean	SE	ESE	$\sqrt{\text{MSE}}$	CP
10	100	0.999	0.874	0.947	0.874	0.967	1.012	0.890	0.968	0.889	0.963
	200	0.969	0.652	0.681	0.652	0.958	0.966	0.655	0.691	0.655	0.961
	500	1.018	0.430	0.441	0.430	0.961	1.014	0.425	0.447	0.425	0.961
	1,000	1.031	0.312	0.339	0.314	0.958	1.030	0.319	0.345	0.320	0.956
	2,000	1.013	0.252	0.271	0.252	0.947	1.006	0.247	0.276	0.247	0.949
30	100	0.969	0.500	0.531	0.500	0.966	0.976	0.507	0.542	0.508	0.968
	200	0.989	0.353	0.379	0.353	0.954	0.987	0.358	0.384	0.358	0.962
	500	1.007	0.243	0.253	0.243	0.968	1.002	0.247	0.257	0.247	0.961
	1,000	1.005	0.176	0.194	0.176	0.968	0.998	0.176	0.198	0.176	0.970
	2,000	0.997	0.142	0.161	0.142	0.978	0.992	0.142	0.164	0.142	0.973
50	100	1.016	0.376	0.406	0.376	0.965	1.018	0.385	0.414	0.385	0.963
	200	1.015	0.274	0.290	0.274	0.960	1.012	0.277	0.294	0.277	0.964
	500	1.007	0.184	0.195	0.184	0.962	0.998	0.187	0.198	0.187	0.968
	1,000	1.012	0.140	0.149	0.140	0.964	1.007	0.141	0.152	0.141	0.967
	2,000	1.005	0.110	0.124	0.110	0.962	0.999	0.111	0.126	0.110	0.965
100	100	1.005	0.269	0.283	0.269	0.958	1.006	0.279	0.289	0.279	0.952
	200	1.013	0.196	0.203	0.196	0.962	1.008	0.198	0.205	0.198	0.962
	500	1.005	0.126	0.137	0.126	0.970	0.998	0.126	0.139	0.126	0.972
	1,000	1.000	0.098	0.105	0.098	0.959	0.993	0.098	0.107	0.098	0.958
	2,000	0.997	0.081	0.088	0.081	0.962	0.991	0.080	0.090	0.081	0.966

Mean– the average of estimates for 1000 simulation replicates; SE–standard error of the estimates from 1000 simulation runs; ESE–estimated standard error of the estimate for each simulation run; MSE–mean squared error; CP–coverage probability of 95% confidence intervals.

Table 13: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for binary outcomes when 20% of the K studies have sample sizes $3n_k$.

$n_k \setminus K$	10	30	50	100
100	1.036	1.030	1.051	1.077
200	1.010	1.027	1.026	1.019
500	0.978	1.028	1.025	0.992
1,000	1.040	0.993	1.011	1.018
2,000	0.960	0.995	1.004	0.994

Table 14: Relative efficiency of $\hat{\beta}_{3,SS}$ versus $\hat{\beta}_{3,IPD}$ for binary outcomes under the same setting of Table 4. But $\hat{\beta}_{3,SS}$ is obtained from a multivariate random-effects meta-analysis using the summary statistics for $(\boldsymbol{\beta}_k, \boldsymbol{\eta}_k)$.

Correlation	$n_k \setminus K$	10	30	50	100
$\rho_x = 0$	100	0.930	0.940	0.944	0.934
	200	0.969	0.962	0.958	0.976
	500	0.988	0.983	0.984	0.993
	1,000	0.991	0.986	0.988	0.992
	2,000	0.995	0.997	0.993	0.991
$\rho_x \sim$ Unif(-0.3,0.3)	100	0.932	0.929	0.941	0.945
	200	0.974	0.971	0.966	0.956
	500	0.987	0.981	0.983	0.975
	1,000	0.992	0.988	0.987	0.989
	2,000	0.996	0.993	0.995	0.991

Table 15: The IPD- and summary-statistics-based meta-analysis of the alcohol intervention data excluding two large studies (8a and 8b).

	IPD			Summary Statistics		
Parameter	Estimate	SE	P-value	Estimate	SE	P-value
Linear regression models for $Y^{(1)}$ (the change in the number of drinks)						
β_1	0.065	0.183	0.722	0.012	0.188	0.950
β_2	-0.370	0.049	4.80×10^{-14}	-0.401	0.054	1.06×10^{-13}
β_3	-0.015	0.046	0.744	-0.030	0.052	0.559
Logistic regression models for $Y^{(2)}$ (whether the number of drinks was reduced)						
β_1	0.404	0.223	0.070	0.264	0.235	0.260
β_2	0.538	0.066	2.35×10^{-16}	0.493	0.069	1.02×10^{-12}
β_3	-0.072	0.050	0.153	-0.034	0.055	0.535

E Supplementary example

Beta-blockade was a major drug to reduce mortality after myocardial infarction in the treatment of patients with myocardial infarction. Yusuf et al. (1985) presented an overview of the effectiveness of beta-blockade during and after myocardial infarction from 22 clinical trial centers. Since its publication, it has been widely cited for meta-analysis and other applications. For example, Freemantle et al. (1999) used it for meta-analysis to assess the effectiveness of beta-blockade in short term treatment for acute myocardial infarction and in longer term secondary prevention.

As shown in Table 16, each center reported the number of deaths, along with the total number of patients, in the "Control" and the "Treated" group (with beta-blocker). Table

16 essentially provides individual-level data for a total of $20,290$ patients. Taking Center 1 for example, there are 3 deaths among the 39 patients in the "Control" group. This corresponds to 39 observations with a treatment indicator variable being 0 ("Control") and a binary response variable being 1 for 3 subjects and 0 for the remaining 36. Following this argument, we can "reconstruct" individual observations (IPD data) of all the patients.

To conduct IPD analysis of the treatment effect, we model the probability of "Death" $P(Y = 1)$ in terms of the treatment indicator variable $X(= 0 \text{ or } 1)$. Specifically, we use a logistic regression model

$$P(Y_{ki} = 1|X_{ki}) = \frac{\exp(\alpha_k + \beta_k X_{ki})}{1 + \exp(\alpha_k + \beta_k X_{ki})} \quad (k = 1, \ldots, 22; i = 1, \ldots, n_k).$$

To allow possible heterogeneity among the centers, we assume that both the intercepts α_k's (i.e., the baseline effects) and the treatment effects β_k's are random and they follow a bivariate normal distribution

$$\begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \tau_\alpha^2 & \rho\tau_\alpha\tau_\beta \\ \rho\tau_\alpha\tau_\beta & \tau_\beta^2 \end{pmatrix} \right).$$

The maximum likelihood inference yields $\hat{\beta}_{IPD} = -0.247$ (i.e., the log odds ratio) with an estimated standard error of 0.057. This implies that the use of beta-blockade reduces the probability of death, and this effect is significant with a p-value of 1.65×10^{-5}.

For summary-statistics-based meta-analysis, we assume that only the center-specific treatment effect estimate $\hat{\beta}_k$ and its variance estimate $\widehat{\text{var}}(\hat{\beta}_k \mid \beta_k)$ are given from each center. These summary statistics are presented in Table 17. Plugging these statistic into (2.2), we carry out meta-analysis. The estimated treatment effect is $\hat{\beta}_{SS} = -0.250$ with a estimated standard error of 0.058, and we conclude that the treatment effect is significant with a p-value of 1.44×10^{-5}. The analysis results here are quite similar to those obtained from the IPD analysis. The similarity implies that our summary-statistics-based analysis

virtually does not lose any efficiency, even in the situation where we are not given any information of the nuisance parameters α_k's. Our further examination shows that α_k's are, in fact, correlated with the treatment effects β_k's with a correlation estimate $\hat{\rho} = -0.45$. This result confirms our finding in Theorem 1(c), which says that meta-analysis using merely summary statistics of β_k's is fully efficient without the knowledge of the nuisance parameters α_k's even if the two random effects α_k's and β_k's are correlated.

Table 16: Beta-blocker data collected from 22 clinical centers

	Control		Treated			Control		Treated	
Center	Deaths	Total	Deaths	Total	Center	Deaths	Total	Deaths	Total
1	3	39	3	38	12	47	266	45	263
2	14	116	7	114	13	16	293	9	291
3	11	93	5	69	14	45	883	57	858
4	127	1520	102	1533	15	31	147	25	154
5	27	365	28	355	16	38	213	33	207
6	6	52	4	59	17	12	122	28	251
7	152	939	98	945	18	6	154	8	151
8	48	471	60	632	19	3	134	6	174
9	37	282	25	278	20	40	218	32	209
10	188	1921	138	1916	21	43	364	27	391
11	52	583	64	873	22	39	674	22	680

Table 17: Summary statistics used in meta-analysis of the Beta-blocker data

Center	$\hat{\beta}_k$	$\widehat{\text{var}}(\hat{\beta}_k \mid \beta_k)$	Center	$\hat{\beta}_k$	$\widehat{\text{var}}(\hat{\beta}_k \mid \beta_k)$
1	0.028	0.723	12	-0.039	0.053
2	-0.741	0.233	13	-0.593	0.181
3	-0.541	0.319	14	0.282	0.042
4	-0.246	0.019	15	-0.321	0.089
5	0.069	0.079	16	-0.135	0.068
6	-0.584	0.457	17	0.141	0.133
7	-0.512	0.019	18	0.322	0.305
8	-0.079	0.042	19	0.444	0.514
9	-0.424	0.075	20	-0.218	0.068
10	-0.335	0.014	21	-0.591	0.066
11	-0.213	0.038	22	-0.608	0.074

References

Freemantle, N., J. Cleland, P. Young, J. Mason, and J. Harrison (1999). Beta-blockade after myocardial infarction: systematic review and meta regression analysis. *BMJ 318*, 1730–1737.

Yusuf, S., R. Peto, J. Lewis, R. Collins, and P. Sleight (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases. 27*, 335–371.