Original Article

Research Synthesis Methods

Received 21 December 2012, Revised 26 February 2014, Accepted 23 April 2014 Published online 9 June 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1119

Meta-analysis of a continuous outcome combining individual patient data and aggregate data: a method based on simulated individual patient data

Yusuke Yamaguchi,[a*] Wataru Sakamoto,[b] Masashi Goto,[c] Jan A. Staessen,[d,e] Jiguang Wang,[f] Francois Gueyffier[g] and Richard D. Riley[h]

When some trials provide individual patient data (IPD) and the others provide only aggregate data (AD), meta-analysis methods for combining IPD and AD are required. We propose a method that reconstructs the missing IPD for AD trials by a Bayesian sampling procedure and then applies an IPD meta-analysis model to the mixture of simulated IPD and collected IPD. The method is applicable when a treatment effect can be assumed fixed across trials. We focus on situations of a single continuous outcome and covariate and aim to estimate treatment–covariate interactions separated into within-trial and across-trial effect. An illustration with hypertension data which has similar mean covariates across trials indicates that the method substantially reduces mean square error of the pooled within-trial interaction estimate in comparison with existing approaches. A simulation study supposing there exists one IPD trial and nine AD trials suggests that the method has suitable type I error rate and approximately zero bias as long as the available IPD contains at least 10% of total patients, where the average gain in mean square error is up to about 40%. However, the method is currently restricted by the fixed effect assumption, and extension to random effects to allow heterogeneity is required. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: meta-analysis; individual patient data; treatment–covariate interaction; statistical simulation

1. Introduction

Meta-analysis is a statistical methodology to synthesise results of several trials for the purpose of summarising the evidence, for example, about a treatment effect. Interest is growing in using meta-analysis to estimate variation of treatment effect according to patient characteristics rather than just to estimate an overall effect size across trials (Thompson and Higgins, 2005). For example, when the benefit of a drug treatment is expected to depend on a

[a]*Division of Mathematical Science, Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan*
[b]*Division of Human Ecology, Graduate School of Environmental and Life Science, Okayama University, 3-1-1, Tsushima-naka, Kita-ku, Okayama 700-8530, Japan*
[c]*NPO, Biostatistical Research Association, 2-22-10-A411, Kamishinden, Toyonaka, Osaka 560-0085, Japan*
[d]*Studies Coordinating Centre, Division of Hypertension and Cardiovascular Rehabilitation, Department of Cardiovascular Diseases, University of Leuven, Campus Gasthuisberg, Herestraat 49/702, B-3000 Leuven Belgium*
[e]*Department of Epidemiology, Maastricht University, Maastricht Netherlands*
[f]*Centre for Epidemiological Studies and Clinical Trials, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Ruijin 2nd Road 197, Shanghai 200025, China*
[g]*INSERM, CIC201, Lyon F-69000, France*
[h]*Public Health, Epidemiology and Biostatistics, Public Health Building, University of Birmingham, Edgbaston B15 2TT, UK*
**Correspondence to: Yusuke Yamaguchi, Division of Mathematical Science, Graduate School of Engineering Science, Osaka University, 1–3, Machikaneyama-cho, Toyonaka, Osaka, 560–8531, Japan.*
E-mail: yamaguti@sigmath.es.osaka-u.ac.jp

patient characteristic such as age, race or gender of patients, it could be clinically important to identify the patient characteristic as a predictor for the treatment effect and an effect modifier (i.e. a variable that interacts with the drug treatment) in the context of meta-analysis. Meta-regression is one choice to relate the treatment effect to one or more trial-level covariates. When interest lies in the interaction between the treatment effect and the patient characteristics, the meta-regression incorporates covariates as summary statistics on background factor of patients for each trial, such as a mean age and a proportion of male patients. However, using only such aggregate data (AD) causes an issue referred to as ecological bias (Morgenstern, 1982). The 'across-trial interaction' between the treatment effect estimates and mean covariate values may not reflect the more pertinent 'within-trial interaction' between individual outcomes and individual covariate values (Riley and Steyerberg, 2010). The discrepancy between associations at the aggregation-level and the individual-level has been discussed in a variety of disciplines including political science (King, 1997), sociology (Robinson, 1950), spatial epidemiology (Elliott *et al*., 2000) and public health (Morgenstern, 1995). In particular, Wakefield (2008) gave illustrative explanations to describe how the ecological bias occurs in geographical correlation study (e.g. an association between stomach cancer mortality and infant mortality, each measured in several countries) and concluded that the aggregation process reduces information, and this information loss usually prevents identification of individual-level association. The relationship described by the meta-regression is also an observational association so that the across-trial interaction estimate from the meta-regression often suffers from the bias by confounding (Thompson and Higgins, 2002). In addition, when meta-analyses involve trials with very similar mean covariate values, a test to detect the treatment–covariate interaction using the meta-regression has seriously lower statistical power (Lambert *et al*., 2002; Simmonds and Higgins, 2007). Berlin *et al*. (2002) conducted two types of meta-analyses by using individual patient-level data and trial-level data from five trials in their clinical research and showed that the meta-analysis based on the trial-level data failed to detect the treatment–covariate interaction.

As a solution to this problem, meta-analysis methods based on individual patient data (IPD), which synthesise the raw data from each trial, have been advocated by many researchers (Riley *et al*., 2010; Simmonds *et al*., 2005). We here focus only on a one-stage meta-analysis approach, which combines all the IPD from all the trials by a single meta-analysis model. Use of the original IPD including patient-level covariates allows one to implement much more meaningful evaluation of the treatment–covariate interaction. In particular, the IPD meta-analysis simultaneously provide both the across-trial and the within-trial interaction effect estimate, which removes the potential risk associated with ecological bias or trial-level confounding. Moreover, even when involving trials with very similar mean covariate values, the IPD meta-analysis can preserve higher statistical power for the treatment–covariate interaction by using the within-trial effect estimate as long as the covariate values of patients within the same trial vary (Simmonds and Higgins, 2007).

However, the IPD meta-analysis may have a disadvantage related to resource, such as substantial time and costs to obtain and process the IPD from all trials, which can cause a situation where meta-analysts obtain IPD from only a proportion of trials. And also, the meta-analysts cannot always collect the IPD from all trials because the IPD might have been lost or damaged, or trial practitioners may not be willing to collaborate (Riley *et al*., 2010). If the availability of IPD is associated with the results in each trial, a meta-analysis based only on the collected IPD may be biassed (Stewart and Tierney, 2002). For these reasons, it has become increasingly important to consider situations where some trials provide IPD and the others provide only AD. Some researchers have already investigated how to combine IPD and AD in meta-analysis, especially when treatment–covariate interaction is of interest (Riley and Steyerberg, 2010; Riley *et al*., 2008). The one-step meta-analysis approach combining IPD and AD is flexible enough to fit a fully hierarchical model with random effects (Sutton *et al*., 2008; Goldstein *et al*., 2000) and thus is useful to investigate the influence of covariates on heterogeneity of the treatment effect both within and across trials. There are some papers that present models to combine IPD and AD in mixed treatment comparison meta-analysis (Donegan *et al*., 2013; Jansen, 2012; Saramago *et al*., 2012). Such approaches have also been developed for ecological study in social sciences (Jackson *et al*., 2006; Haneuse and Wakefield, 2008; Wakefield, 2004; Wakefield *et al*., 2011). Wakefield *et al*. (2011) advocated that the only reliable approach for removing ecological bias is to supplement the ecological data with individual-level information. We build on these articles here.

In this paper, we propose a novel method for combining IPD and AD, in which missing IPD for trials providing only AD are multiply simulated, and then a patient-specific statistical model is applied to the mixture of each simulated IPD (SIPD) and the collected IPD. Once the SIPD are generated, we show how existing IPD meta-analysis approaches can be applied, and we demonstrate the benefits of incorporating the SIPD. We here consider a Bayesian sampling scheme to generate the SIPD, which is based on multiple imputation applied in the analysis of incomplete data with missing outcomes and covariates (Rubin, 1987). In particular, we generate the SIPD from a posterior predictive distribution of the missing IPD given AD and the collected IPD. This Bayesian sampling structurally takes into account uncertainty in model parameters and hence leads a valid inference for the quantities of interest in a similar way to the multiple imputation justified within Bayesian framework. Multiple estimates of a parameter of interest (e.g. within-trial treatment–covariate interaction effect) produced by repeatedly conducting an IPD meta-analysis for the mixture of each SIPD and the collected IPD are summarised in accordance with a rule by Rubin (1987). Note that we only consider fixed effect meta-analysis models through this paper, so that the treatment effect and the

treatment–covariate interaction effects are assumed to be common across trials. This assumption may not be plausible in situations where between-trial heterogeneity exists.

In Section 2, we introduce a data set in hypertension, where IPD are available from five trials, with blood pressure as the continuous outcome of interest. In Section 3, we describe existing models for combing IPD and AD and also discuss within-trial and across-trial relationships. In Section 4, we describe our new method based on SIPD. In Section 5, we illustrate the proposed method with the hypertension data, where we assess how a covariate modifies treatment effectiveness. In Section 6, we conduct a simulation study to examine the performance of our proposed method in comparison with existing methods. Finally, in Section 7, we conclude this paper with some discussion.

2. Motivating example

Wang *et al*. (2005) performed a quantitative overview of trials in hypertension to investigate to what extent lowering of systolic blood pressure (SBP) and diastolic blood pressure contributed to cardiovascular prevention. They selected randomised controlled trials that tested active antihypertensive drugs against placebo or no treatment. For their analyses, IPD was sought from trials in the individual data analysis of antihypertensive intervention trials data set (Gueyffier *et al*., 1995) or at the Studies Coordinating Centre in Leuven (Belgium) (Liu *et al*., 1998; Staessen *et al*., 1997; Amery *et al*., 1985). Ten trials were ultimately included, and these provided IPD for a total of 28 592 patients. To illustrate our method, we will carry out a meta-analysis of five (12 603 patients) of these 10 trials, which are sufficiently homogeneous across trials with respect to a treatment effect and a trial-level covariate. These five trials were chosen as they were conducted in populations with a similar mean age of around 70 years. This is because we here consider models with a common covariate effect across trials rather than with trial-specific covariate effects. In some cases, this assumption is acceptable for trials with similar mean covariates. Moreover, we supposed a situation where the across-trial treatment–covariate interaction was estimated poorly. When trials involved have similar mean covariates, estimation of the across-trial treatment–covariate interaction is subject to ecological bias and/or trial-level confounding and often suffers from large standard error. The mean change in SBP (follow-up minus baseline) for each treatment group in each trial are shown in Table 1, with negative values indicating a beneficial effect. The treatment effect is shown in the rightmost column in Table 1, with negative values indicating that the treatment is effective. Table 1 also shows the mean age, and the groups appear to be well balanced in each trial at baseline. In every trial, the active treatment reduces SBP more than placebo on average, but it is also clinically important to assess how age modifies the treatment effect. The hypertension data will be used in this paper to demonstrate and critically assess the method developed; those interested in more clinical conclusions are referred elsewhere (Wang *et al*., 2005).

3. Existing models for combining IPD and AD

Consider a meta-analysis of N' randomised trials with two groups (treatment or control) and a single continuous outcome and a single continuous covariate observed for each patient in each group. Suppose that N AD trials ($i = 1, \ldots, N$) and $N' - N$ IPD trials ($i = N+1, \ldots, N'$) are collected. Let y_{ij} and z_{ij} be the patient-level outcome and the covariate observed for the jth patient ($j = 1, \ldots, n_i$) in the ith trial, and let x_{ij} be coded 0/1 to denote control/treatment

Table 1. Summary of the five trials included in the meta-analysis of Wang *et al*. (2005).

	Number of patients		Age (years)		SBP (follow-up minus baseline)		
Trial name*	Control	Treatment	Control mean (s.d.)	Treatment mean (s.d.)	Control mean (s.d.)	Treatment mean (s.d.)	Estimate of mean difference (s.e.)
HEP	199	150	69.57 (5.39)	69.71 (5.18)	−11.65 (23.30)	−24.88 (21.11)	−13.23 (2.39)
EWPHE	82	90	74.11 (8.69)	72.64 (7.99)	−7.78 (22.76)	−20.46 (19.80)	−12.68 (3.27)
MRC-2	1337	1314	70.43 (2.75)	70.39 (2.77)	−17.55 (21.95)	−28.20 (21.78)	−10.65 (0.85)
SHEP	2371	2365	71.54 (6.68)	71.64 (6.72)	−13.88 (19.90)	−25.39 (18.42)	−11.51 (0.56)
Sy-Eur	2297	2398	70.20 (6.68)	70.25 (6.75)	−8.70 (15.04)	−18.89 (16.15)	−10.18 (0.46)

SBP, systolic blood pressure; s.d., standard deviation; s.e., standard error.
*Trial names are consistent with Wang *et al*. (2005), where further details and trial publications can be found.

group. The IPD for trials $i = N+1, \ldots, N'$ consist of their patient-level observations; that is, (y_{ij}, x_{ij}, z_{ij}) for all patients $j = 1, \ldots, n_i$. On the other hand, the AD for trials $i = 1, \ldots, N$ consist only of sample means and sample variances in each group, which are obtained by summarising their patient-level observations; that is, $(\bar{y}_{iT}, s^2_{y_iT}, \bar{z}_{iT}, s^2_{z_iT}, \bar{y}_{iC}, s^2_{y_iC}, \bar{z}_{iC}, s^2_{z_iC})$ for $i = 1, \ldots, N$. Here,

$$\bar{y}_{iT} = \frac{\sum_{j\in T} y_{ij}}{n_{iT}}, s^2_{y_iT} = \frac{\sum_{j\in T}\left(y_{ij} - \bar{y}_{iT}\right)^2}{n_{iT} - 1}, \bar{z}_{iT} = \frac{\sum_{j\in T} z_{ij}}{n_{iT}}, s^2_{z_iT} = \frac{\sum_{j\in T}\left(z_{ij} - \bar{z}_{iT}\right)^2}{n_{iT} - 1},$$
$$\bar{y}_{iC} = \frac{\sum_{j\in C} y_{ij}}{n_{iC}}, s^2_{y_iC} = \frac{\sum_{j\in C}\left(y_{ij} - \bar{y}_{iC}\right)^2}{n_{iC} - 1}, \bar{z}_{iC} = \frac{\sum_{j\in C} z_{ij}}{n_{iC}}, s^2_{z_iC} = \frac{\sum_{j\in C}\left(z_{ij} - \bar{z}_{iC}\right)^2}{n_{iC} - 1}$$

where T and C indicate treatment and control group, respectively. Also, n_{iT} and n_{iC} denote the number of patients for each group in the ith trial.

If just meta-analysing the IPD trials, we apply the following one-stage model derived by modifying a model by Riley *et al*. (2008), which accounts for the clustering of patients within trials by a trial-specific intercept (ϕ_i) and estimates a pooled treatment–covariate interaction (γ_W) on the basis of within-trial information separated from the across-trial interaction (γ_A):

$$\begin{aligned} y_{ij} &= \phi_i + \theta x_{ij} + \mu z_{ij} + \gamma_A x_{ij}\bar{z}_i + \gamma_W x_{ij}\left(z_{ij} - \bar{z}_i\right) + \varepsilon_{ij}, \\ \varepsilon_{ij} &\sim N\left(0, \sigma^2_y\right). \end{aligned} \tag{1}$$

Here, ϕ_i is the fixed intercept for the ith trial (which essentially accounts for clustering of patients within trials), θ is a fixed hypothetical treatment effect in a trial with $\bar{z}_i = 0$, μ is a mean change in control group response for a one-unit increase in z_{ij}, γ_A and γ_W are respectively the across-trial and the within-trial effect of treatment–covariate interaction. $\bar{z}_i = \sum_{j=1}^{n_i} z_{ij}/n_i$ denotes a mean covariate value in the ith trial. Note that θ, μ, γ_A and γ_W are considered fixed effects here, and σ^2_y is assumed common across trials. In contrast to Riley *et al*. (2008), model (1) assumes the fixed hypothetical treatment effect θ and includes a common μ and σ^2_y across trials rather than a trial-specific μ_i and $\sigma^2_{y_i}$ for each trial. A random treatment effect is more plausible to take into account heterogeneity across trials, and the assumptions of μ and σ^2_y may lead biassed estimates of the treatment effect and the treatment–covariate interaction effect (Riley *et al*., 2008; Higgins *et al*., 2001); however, these assumptions are necessary to build the proposed method (Section 4). According to a recommendation by Riley *et al*. (2008), the treatment–covariate interaction is separated into across-trial and within-trial effects. The separation of across-trial and within-trial interactions has been discussed by many researchers (Neuhaus and Kalbfleisch, 1998; Mancl *et al*., 2000; Begg and Parides, 2003; Simmonds, 2005) and is clinically important to avoid making a wrong conclusion about the treatment–covariate interaction, which might occur if wrongly amalgamating within-trial and across-trial effects (Riley and Steyerberg, 2010).

When a mixture of IPD and AD trials are available, model (1) must be modified. The simplest solution is to reduce the collected IPD to AD and treat all the data as AD (Section 3.1), so that any information on the individual-level associations from the IPD trials is lost. Alternately, one could use only the collected IPD, so that available information from the AD trials is thrown away. In contrast, Riley *et al*. (2008) proposed a model for combing IPD and AD (Section 3.2). This model framework simultaneously estimates the within-trial effect (using just the IPD trials) and the across-trial effect (using both IPD and AD trials). All these approaches are now described.

3.1. *Meta-regression model that uses only AD from all trials*

Once the IPD for trials $i = N+1, \ldots, N'$ are summarised to the AD, a meta-regression model

$$\begin{aligned} d_i &= \theta + \gamma_A \bar{z}_i + \varepsilon_i, \\ \varepsilon_i &\sim N\left(0, \sigma^2_{d_i}\right) \end{aligned} \tag{2}$$

can be applied to the AD for all trials $i = 1, \ldots, N'$, where

$$d_i = \bar{y}_{iT} - \bar{y}_{iC}$$

denotes a mean difference between groups estimated from the ith trial, and the error variance is assumed to be known as

$$\sigma^2_{d_i} = V(d_i) = \frac{s^2_{y_iT}}{n_{iT}} + \frac{s^2_{y_iC}}{n_{iC}}.$$

As in model (1), we also assume that θ is a fixed effect in the meta-regression model (2).

3.2. Model that uses both IPD and AD trials

We apply the following model for combining IPD and AD, derived by modifying a model by Riley *et al*. (2008):

$$\begin{aligned} y_{ij}^{*} &= D_i\phi_i + \theta x_{ij} + D_i\mu z_{ij}^{*} + \gamma_A x_{ij}\bar{z}_i + D_i\gamma_W x_{ij}(z_{ij}^{*} - \bar{z}_i) + \varepsilon_{ij}^{*}, \\ \varepsilon_{ij}^{*} &\sim N\left(0, V_i^{*}\right) \end{aligned} \tag{3}$$

where D_i is a dummy variable to distinguish IPD trials from AD trials. With similar definitions by Riley *et al*. (2008), for each IPD trial, $D_i = 1, y_{ij}^{*} = y_{ij}, V_i^{*} = \sigma_y^2$ and $z_{ij}^{*} = z_{ij}$. For each AD trial, there is only one response ($j = 1$) and $D_i = 0$, $x_{i1} = 1$, $y_{i1}^{*} = d_i$, $V_i^{*} = V(d_i)$ (assumed known) and $z_{i1}^{*} = \bar{z}_i$. This model framework ensures that the AD from trials $i = 1, \ldots, N$ help to estimate only the parameters of across-trial relationships (θ and γ_A), whereas the IPD from trials $i = N+1, \ldots, N'$ help to estimate all the parameters (Riley *et al*., 2008). That is, only the collected IPD contributes to the estimation of the parameters of within-trial relationships (μ and γ_W). As in model (1), we again assume that θ, μ, γ_A and γ_W are fixed effects, and σ_y^2 is common across trials.

4. A new method based on simulated IPD (SIPD) for AD trials

We now introduce our new proposed method for combining IPD and AD. The proposed method takes the following procedures for inference of parameters:

(1) Generate multiple sets of SIPD for each trial providing only AD.
  (1-1) Get a posterior distribution of unknown model parameters using the collected IPD and AD and then draw samples of the parameters from the posterior distribution.
  (1-2) Get the SIPD as random samples from a posterior predictive distribution of the missing IPD with known parameters drawn in step (1–1).
(2) Fit a standard meta-analysis model to each set of SIPD combined with the collected IPD.
(3) Suitably summarise resulting estimates from the set of meta-analyses from step (2).

We refer to this whole process as SIPD method, and each step is now described in more detail.

4.1. Step (1): generating SIPD

We first consider a model for generating the SIPD, which is based on the IPD meta-analysis model (1), and then construct the posterior distribution of the unknown model parameters using the collected IPD and AD. Once getting a density function of the posterior distribution explicitly, we can draw samples of the parameters from the posterior distribution using an algorithm of Markov chain Monte Carlo (MCMC) method. And then, we approximately construct the posterior predictive distribution of the missing IPD with known parameters drawn in the previous step. We finally get the SIPD for trials providing only AD by drawing samples from the posterior predictive distribution.

Now, let

$$\begin{aligned} Y_{\text{miss-IPD}} &= \left\{ \left(y_{ij}, x_{ij}, z_{ij}\right) : i = 1, \ldots, N; j = 1, \ldots, n_i \right\}, \\ Y_{\text{AD}} &= \left\{ \left(\bar{y}_{iT}, s_{y_iT}^2, \bar{z}_{iT}, s_{z_iT}^2, \bar{y}_{iC}, s_{y_iC}^2, \bar{z}_{iC}, s_{z_iC}^2\right) : i = 1, \ldots, N \right\}, \\ Y_{\text{IPD}} &= \left\{ \left(y_{ij}, x_{ij}, z_{ij}\right) : i = N+1, \ldots, N'; j = 1, \ldots, n_i \right\}. \end{aligned}$$

Here, $Y_{\text{miss-IPD}}$ is the missing IPD for trials $i = 1, \ldots, N$, Y_{AD} is the AD summarised from them and Y_{IPD} is the collected IPD for trials $i = N+1, \ldots, N'$, respectively. We firstly consider the IPD meta-analysis model (1) for $Y_{\text{miss-IPD}}$ and Y_{IPD}, and in addition, assume that the covariates of patients assigned to the treatment (or control) group in the ith trial follow a normal distribution with mean m_{z_iT} (or m_{z_iC}) and variance $\sigma_{z_iT}^2$ (or $\sigma_{z_iC}^2$); i.e.

$$z_{ij} \sim \begin{cases} N\left(m_{z_iT}, \sigma_{z_iT}^2\right), & j \in T \\ N\left(m_{z_iC}, \sigma_{z_iC}^2\right), & j \in C \end{cases},$$

Here, $\bar{z}_i = \sum_{j=1}^{n_i} z_{ij}/n_i$ is considered to be approximately constant. If we also assume that z_{ij} and ε_{ij} are independent of each other, we have the conditional distribution of y_{ij} given z_{ij} as follows:

$$y_{ij}|z_{ij} \sim \begin{cases} N\left(\phi_i + \theta + \mu z_{ij} + \gamma_A \bar{z}_i + \gamma_W (z_{ij} - \bar{z}_i), \sigma_y^2\right), & j \in T \\ N\left(\phi_i + \mu z_{ij}, \sigma_y^2\right), & j \in C \end{cases} \tag{4}$$

for $i = 1, \ldots, N'$ and $j = 1, \ldots, n_i$. Here, parameters in the covariate distributions are estimated by a standard method as $\hat{m}_{z_i T} = \bar{z}_{iT}$, $\hat{\sigma}^2_{z_i T} = s^2_{z_i T}$, $\hat{m}_{z_i C} = \bar{z}_{iC}$, $\hat{\sigma}^2_{z_i C} = s^2_{z_i C}$, and now let

$$\hat{\xi} = \left\{ \left(\hat{m}_{z_i T}, \hat{\sigma}^2_{z_i T}, \hat{m}_{z_i C}, \hat{\sigma}^2_{z_i C} \right) : i = 1, \ldots, N' \right\}.$$

And also, let $f(Y_{\text{miss-IPD}}, Y_{\text{IPD}} | \hat{\xi}, \eta)$ be the density function of the distribution (4) with parameter

$$\eta = \left(\phi_1 \ldots, \phi_{N'}, \theta, \mu, \gamma_A, \gamma_W, \sigma_y^2 \right).$$

Then, we draw the SIPD (say $Y^*_{\text{miss-IPD}}$) from the posterior predictive distribution of $Y^*_{\text{miss-IPD}}$ given Y_{IPD} and Y_{AD}; that is,

$$Y^*_{\text{miss-IPD}} \sim f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$$

where

$$f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi}) = \int f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, \hat{\xi}, \eta) f(\eta | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi}) d\eta \tag{5}$$

and $f(\eta | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ is the posterior distribution of η given Y_{AD} and Y_{IPD}. Because of the integration in (5), $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ cannot be expressed in a closed form. Also, it is impossible to draw samples from this distribution directly; however, once getting samples of the parameter η, we can draw $Y_{\text{miss-IPD}}$ from $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ approximately. If R sets of parameter values (say $\eta^{[r]}$ for $r = 1, \ldots, R$) are drawn from the posterior distribution $f(\eta | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ in (5), then the posterior predictive distribution of $Y_{\text{miss-IPD}}$ can be approximated as

$$f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi}) \approx \frac{1}{R} \sum_{r=1}^{R} f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, \hat{\xi}, \eta^{[r]}) \tag{6}$$

where $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, \hat{\xi}, \eta^{[r]})$ is the conditional distribution of $Y_{\text{miss - IPD}}$ given Y_{AD} and the rth drawn parameter $\eta^{[r]}$. This numerical integration indicates that the random sampling from $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ can be alternated by two-stage random sampling from $f(\eta | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ and then $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, \hat{\xi}, \eta)$; that is, one random sample from $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, \hat{\xi}, \eta^{[r]})$ corresponds to one random sample from $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$. The repetition of this drawing yields R sets of SIPD (say $Y^{[r]}_{\text{miss-IPD}}$ for $r = 1, \ldots, R$). Here, the rth set of SIPD consists of the outcome and covariate values from patients involved in the AD trials; that is,

$$Y^{[r]}_{\text{miss-IPD}} = \left\{ \left(y^{[r]}_{ij}, x^{[r]}_{ij}, z^{[r]}_{ij} \right) : i = 1, \ldots, N; j = 1, \ldots, n_i \right\}. \tag{7}$$

Considering the example in Section 2, for an AD trial in hypertension which provides only summary statistics for each treatment group (number of patients, sample mean and standard deviation of change in SBP, and sample mean and standard deviation of age, shown in Table 1), we now have R sets of simulated patient-level observations on the change in SBP and the age. These simulated patient-level observations (i.e. SIPD) can be used for applying model (1), which enables us to assess how the age modifies the treatment effect in terms of the patient-level association (i.e. the within-trial treatment–covariate interaction).

Thus, we generate the simulated IPD for the AD trials by using the approximation (6), which requires us to draw from $f(\eta | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ and then draw from $f(Y_{\text{miss-IPD}} | Y_{\text{AD}}, \hat{\xi}, \eta^{[r]})$. We now describe how to draw from these distributions in more detail.

4.1.1. Step (1–1): drawing from $f(\eta | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$. Drawing samples of parameter η from $f(\eta | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi})$ is straightforward to achieve by the MCMC method; in particular, we use Metropolis–Hastings algorithm. Now, because of between-trial independence, the posterior distribution of η can be rewritten as

$$f(\eta | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi}) \propto f(Y_{\text{AD}} | \hat{\xi}, \eta) f(Y_{\text{IPD}} | \hat{\xi}, \eta) f(\eta) \tag{8}$$

where $f(\eta)$ is the density function for a prior distribution of η and we use a vague prior for this; that is, $f(\eta) \propto \sigma_y^{-2}$. Then, we are required to specify the explicit forms of $f(Y_{\text{AD}} | \hat{\xi}, \eta)$ and $f(Y_{\text{IPD}} | \hat{\xi}, \eta)$ in order to apply the MCMC

method. For the former, we cannot write down $f(Y_{AD}|\hat{\xi},\eta)$ directly because we do not assume the separation model for the collected IPD and AD such as the existing model (3). We therefore derive this by marginalising $f(Y_{\text{miss-IPD}}|\hat{\xi},\eta)$ with respect to $Y_{\text{miss-IPD}}$. Supposing that Y_{AD} is given as a function of $Y_{\text{miss - IPD}}$, say $h(Y_{\text{miss - IPD}}) = Y_{AD}$ where $h(\cdot)$ is the function to summarise individual outcome and covariate values from patients assigned to each group in each trial to sample means and sample variances, we have the following density function under an assumption known as coarsening at random by Heitjan and Rubin (1991):

$$f(Y_{AD}|\hat{\xi},\eta) = \int_{h(Y_{\text{miss-IPD}})=Y_{AD}} f(Y_{\text{miss-IPD}}|\hat{\xi},\eta)dY_{\text{miss-IPD}} \tag{9}$$

where the integration with respect to $Y_{\text{miss - IPD}}$ is taken over the region that satisfies $h(Y_{\text{miss-IPD}}) = Y_{AD}$. We are here assuming availability of IPD is not related to the missing IPD, which is similar to an assumption of missing at random (Little and Rubin, 2002). In particular, this is calculated as follows except for terms unrelated to the parameters:

$$\begin{aligned}\prod_{i=1}^{N}\Bigg[&\left(\sigma^2_{y_iT}\right)^{-n_{iT}/2} exp\left\{-\frac{n_{iT}}{2\sigma^2_{y_iT}}\left(\left(\bar{y}_{iT}-m_{y_iT}\right)^2+s^2_{y_iT}\right)\right\}\\ &\times\left(\sigma^2_{y_iC}\right)^{-n_{iC}/2} exp\left\{-\frac{n_{iC}}{2\sigma^2_{y_iC}}\left(\left(\bar{y}_{iC}-m_{y_iC}\right)^2+s^2_{y_iC}\right)\right\}\Bigg]\end{aligned} \tag{10}$$

where

$$\begin{aligned} m_{y_iT} &= \phi_i + \theta + \mu\hat{m}_{z_iT} + \gamma_A\bar{z}_i + \gamma_W(\hat{m}_{z_iT} - \bar{z}_i),\\ \sigma^2_{y_iT} &= (\mu+\gamma_W)^2\hat{\sigma}^2_{z_iT} + \sigma^2_y \end{aligned}$$

and

$$\begin{aligned} m_{y_iC} &= \phi_i + \mu\hat{m}_{z_iC},\\ \sigma^2_{y_iC} &= \mu^2\hat{\sigma}^2_{z_iC} + \sigma^2_y. \end{aligned}$$

The calculation to get (10) from (9) is outlined in Appendix A. Here, if we assume a random treatment effect as $\theta_i \sim N(\theta,\tau^2)$ in place of the fixed treatment effect θ in model (1), then we must modify the expression (10) to include the parameter τ^2; however, it may be impossible to derive this density explicitly by a standard approach. This is the reason why the random effects approach cannot be adopted in this article. In Section 7, we discuss how the current approach could be extended to allow for the random effects.

We can also derive $f(Y_{\text{IPD}}|\hat{\xi},\eta)$ in (8) by the normal density as follows:

$$\prod_{i=N+1}^{N'}\prod_{j=1}^{n_i}\frac{1}{\sqrt{2\pi}\sigma_y}\ exp\left\{-\frac{1}{2\sigma^2_y}\left(y_{ij}-\phi_i-\theta x_{ij}-\mu z_{ij}-\gamma_A x_{ij}\bar{z}_i-\gamma_W x_{ij}\left(z_{ij}-\bar{z}_i\right)\right)^2\right\}. \tag{11}$$

By using (10) and (11), the Metropolis–Hastings algorithm allows us to draw R sets of values of the parameter η from $f(\eta|Y_{AD}, Y_{\text{IPD}})$, which are denoted as

$$\eta^{[r]} = \left(\phi_1^{[r]},\ldots,\phi_N^{[r]},\theta^{[r]},\mu^{[r]},\gamma_A^{[r]},\gamma_W^{[r]},\sigma_y^{2[r]}\right)$$

for $r = 1,\ldots,R$. Implementing steps of the algorithm to get $\eta^{[r]}$ are detailed in Appendix B.

4.1.2. Step (1–2): drawing from $f(Y_{\text{miss-IPD}}|Y_{AD},\hat{\xi},\eta^{[r]})$. We can derive $f(Y_{\text{miss-IPD}}|\hat{\xi},\eta^{[r]})$ as the normal density with known parameter $\eta^{[r]}$ in a similar way to (11), while $f(Y_{\text{miss-IPD}}|Y_{AD},\hat{\xi},\eta^{[r]})$ is difficult to derive exactly because its sample space is defined on the region that satisfies $h(Y_{\text{miss-IPD}}) = Y_{AD}$. This means that sample means and sample variances of outcome and covariate for each group in each trial, which is computed by using individual outcome and covariate values from each patient drawn from $f(Y_{\text{miss-IPD}}|Y_{AD},\hat{\xi},\eta^{[r]})$, must be equivalent to the corresponding sample means and sample variances in Y_{AD}. This difficulty is associated with some issues on the conditional distribution given sufficient statistics (Cheng, 1984; Engen and Lillegard, 1997; Lindqvist and Taraldsen, 2005). Lindqvist and Taraldsen (2005) suggested a general formula in order to calculate the conditional expectation on the basis of this conditional distribution by Monte Carlo approximation. We here brief how the sampling technique by Lindqvist and Taraldsen (2005) draws $Y_{\text{miss - IPD}}$ from $f(Y_{\text{miss-IPD}}|Y_{AD},\hat{\xi},\eta^{[r]})$.

Now, we represent the observation vectors for each group in the ith AD trial as follows:

$$\mathbf{y}_{iT} = \{y_{ij} : j \in T\}, \quad \mathbf{y}_{iC} = \{y_{ij} : j \in C\}, \quad \mathbf{z}_{iT} = \{z_{ij} : j \in T\}, \quad \mathbf{z}_{iC} = \{z_{ij} : j \in C\}.$$

Recall that $Y_{\text{miss-IPD}}$ denotes the missing IPD from trials $i = 1, \ldots, N$; that is, (y_{ij}, x_{ij}, z_{ij}) for $j = 1, \ldots, n_i$. Because of between-trial and between-group independence, we have the density function of $Y_{\text{miss-IPD}}$ given Y_{AD} and $\eta^{[r]}$ as follows:

$$\begin{aligned} f\left(Y_{\text{miss-IPD}} | Y_{\text{AD}}, \hat{\xi}, \eta^{[r]}\right) = \prod_{i=1}^{N} & f\left(\mathbf{z}_{iT} | \bar{z}_{iT}, s^2_{z_i T}, \hat{\xi}\right) f\left(\mathbf{y}_{iT} | \bar{y}_{iT}, s^2_{y_i T}, \mathbf{z}_{iT}, \eta^{[r]}\right) \\ & \times f\left(\mathbf{z}_{iC} | \bar{z}_{iC}, s^2_{z_i C}, \hat{\xi}\right) f\left(\mathbf{y}_{iC} | \bar{y}_{iC}, s^2_{y_i C}, \mathbf{z}_{iC}, \eta^{[r]}\right). \end{aligned} \tag{12}$$

Then, the rth SIPD for the ith AD trial; that is, $(y^{[r]}_{ij}, x^{[r]}_{ij}, z^{[r]}_{ij})$ for $j = 1, \ldots, n_i$, are generated as random samples drawn from the corresponding conditional distribution in (12), which indicates the following sequential sampling procedures:

$$\begin{aligned} \mathbf{z}^{[r]}_{iT} &\sim f\left(\mathbf{z}_{iT} | \bar{z}_{iT}, s^2_{z_i T}, \hat{\xi}\right), \\ \mathbf{y}^{[r]}_{iT} | \mathbf{z}^{[r]}_{iT} &\sim f\left(\mathbf{y}_{iT} | \bar{y}_{iT}, s^2_{y_i T}, \mathbf{z}^{[r]}_{iT}, \eta^{[r]}\right) \end{aligned}$$

and

$$\begin{aligned} \mathbf{z}^{[r]}_{iC} &\sim f\left(\mathbf{z}_{iC} | \bar{z}_{iC}, s^2_{z_i C}, \hat{\xi}\right), \\ \mathbf{y}^{[r]}_{iC} | \mathbf{z}^{[r]}_{iC} &\sim f\left(\mathbf{y}_{iC} | \bar{y}_{iC}, s^2_{y_i C}, \mathbf{z}^{[r]}_{iC}, \eta^{[r]}\right). \end{aligned}$$

This means that, for the ith AD trial, we firstly draw $\mathbf{z}^{[r]}_{iT}$ and then $\mathbf{y}^{[r]}_{iT}$ by using $\mathbf{z}^{[r]}_{iT}$, which are applied for drawing $\mathbf{z}^{[r]}_{iC}$ and $\mathbf{y}^{[r]}_{iC}$.

Here, $\mathbf{z}^{[r]}_{iT}$ denotes samples from the conditional normal distribution given sample mean $\bar{z}_{iT}$ and sample variance $s^2_{z_i T}$. A result by Lindqvist and Taraldsen (2005) outlined in Appendix C allows us to achieve this drawing as follows:

$$\mathbf{z}^{[r]}_{iT} = \left\{ \bar{z}_{iT} + \frac{u_{ij} - \bar{u}_i}{s_{u_i}} s_{z_i T} : j \in T \right\} \tag{13}$$

where $\{u_{ij} : j \in T\}$ denotes n_{iT} random samples from the standard normal distribution. Also, $\bar{u}_i$ and $s^2_{u_i}$ are a sample mean and a sample variance for them, respectively. Furthermore, we can get $\mathbf{y}^{[r]}_{iT}$ in a similar way to (13) as follows:

$$\mathbf{y}^{[r]}_{iT} = \left\{ \bar{y}_{iT} + \left(\mu^{[r]} + \gamma^{[r]}_W\right)\left(z^{[r]}_{ij} - \bar{z}_{iT}\right) + \delta_i\left(v_{ij} - \bar{v}_i\right) : j \in T \right\} \tag{14}$$

where $\{v_{ij} : j \in T\}$ denotes random numbers from the standard normal distribution. Also, $\bar{v}_i$ and $s^2_{v_i}$ are a sample mean and a sample variance for them, respectively, and

$$\begin{aligned} \delta_i &= \frac{-\left(\mu^{[r]} + \gamma^{[r]}_W\right) s_{z_i, v_i} + \sqrt{\left(\mu^{[r]} + \gamma^{[r]}_W\right)^2 s^2_{z_i, v_i} - s^2_{v_i} s^2_{z_i T} + s^2_{v_i} s^2_{y_i T}}}{s^2_{v_i}}, \\ s_{z_i, v_i} &= \frac{1}{n_{iT} - 1} \sum_{j \in T} \left(z^{[r]}_{ij} - \bar{z}_{iT}\right)\left(v_{ij} - \bar{v}_i\right). \end{aligned}$$

We can use (13) and (14) for $j \in C$ in the similar way, in which $\mu^{[r]} + \gamma^{[r]}_W$ is replaced by $\mu^{[r]}$. Finally, we have R sets of the SIPD, which are denoted by a vector form of (7) as

$$Y^{[r]}_{\text{miss-IPD}} = \left\{ \left(\mathbf{y}^{[r]}_{iT}, \mathbf{y}^{[r]}_{iC}, \mathbf{z}^{[r]}_{iT}, \mathbf{z}^{[r]}_{iC}\right) : i = 1, \ldots, N \right\} \tag{15}$$

for $r = 1, \ldots, R$.

4.2. *Step (2): fitting IPD meta-analysis model to each set of SIPD*

Step (1) produces R sets of SIPD for the AD trials; that is, $Y^{[r]}_{\text{miss-IPD}}$ for $r = 1, \ldots, R$ in (15). We can now fit the IPD meta-analysis model (1) to each of them combined with the collected IPD. This produces R sets of maximum likelihood estimates for parameters of interest and their variance estimates; for instance, the within-trial treatment–covariate interaction effect, $(\hat{\gamma}^{[r]}_W, V(\hat{\gamma}^{[r]}_W))$ for $r = 1, \ldots, R$.

4.3. Step (3): summarising the results for each SIPD

In step (3), resulting estimates for each set of SIPD combined with the collected IPD are suitably summarised according to the rule by Rubin (1987), which is often used in multiple imputation. For example, Rubin's rule allows us to compute an overall estimate $\hat{\gamma}_W$ and its variance $V(\hat{\gamma}_W)$ as

$$\hat{\gamma}_W = \frac{1}{R}\sum_{r=1}^{R}\hat{\gamma}_W^{[r]} \tag{16}$$

and

$$V(\hat{\gamma}_W) = \frac{1}{R}\sum_{r=1}^{R}V\left(\hat{\gamma}_W^{[r]}\right) + \frac{1+R^{-1}}{R-1}\sum_{r=1}^{R}\left(\hat{\gamma}_W^{[r]} - \hat{\gamma}_W\right)^2 \tag{17}$$

where $\hat{\gamma}_W^{[r]}$ and $V(\hat{\gamma}_W^{[r]})$ are an estimate of γ_W and its variance from the rth SIPD, respectively. The calculations to get (16) and (17) are detailed in Appendix D. For frequentist inference, we can use (16) and (17) to construct a confidence interval of γ_W or to test a null hypothesis $H_0 : \gamma_W = 0$.

In summary, Figure 1 shows a step-by-step instruction for estimating the within-trial treatment–covariate interaction by the SIPD method. We use a free R software to conduct the SIPD method. The source code is shown in Appendix E.

5. Application to hypertension data

Consider now application to the hypertension data. Before carrying out the meta-analysis for combining IPD and AD, we examined the pooled treatment effect and the extent of heterogeneity in the treatment effect across the five trials. Figure 2 shows the forest plot of the five trials in hypertension. The black square represents the mean difference estimate from each trial, with horizontal line as its 95% CI. The size of the square is proportional to the precision of the trial. Under a fixed effect model, the pooled mean difference is estimated by -10.77, with 95% CI of $[-11.40, -10.14]$. This indicates that the treatment is significantly effective in reducing SBP by, on average, 10.77 mmHg more than placebo. The Q test for the pooled mean difference estimate, which is often applied in meta-analysis for determining whether there is heterogeneity in treatment effects (Cochran, 1954), gave a Q statistic and its p-value of 4.849 and 0.303, respectively, and thus, there is no strong evidence of heterogeneity. Further, I^2, the percentage of total variation in the estimates of treatment effect that is due to heterogeneity across trials (Higgins and Thompson, 2002), was 17.50%. This again indicates potential low heterogeneity across the five trials. Therefore, we concluded that a fixed treatment effect could be assumed as for models in Sections 3 and 4.

To imitate situations involving IPD for some trials and only AD for others, we generated scenarios where we assumed that only a limited number of trials (from one to four of the five trials) provided IPD, and the other trials just provided AD as presented in Table 1, which is typical of the AD available to meta-analysts in practice. In each scenario, we focused on how age modifies the treatment effect on change in SBP (follow-up minus baseline), and carried out analyses by (i) fitting the meta-regression model (2) to AD from all five trials, (ii) fitting the IPD meta-analysis model (1) to IPD from only IPD trials available (when the number of IPD trials is one, model (1) must be modified to exclude the parameter of the across-trial treatment–covariate interaction effect because one cannot estimate the across-trial interaction effect with a single trial), (iii) fitting model (3) to the mixture of IPD and AD from all five trials and (iv) applying the SIPD method described in Section 4 to the mixture of IPD and AD from all five trials. In all parts from (ii) to (iv), the analyses were run for each possible combination of IPD and AD trials. For example, in the scenario that two trials provide IPD (i.e. two IPD trials and three AD trials), we performed 10 analyses, one for each combination of which two trials provide IPD and three provide AD. In each scenario, we compared the results with those from a meta-analysis of IPD from all five trials (i.e. full IPD analysis), allowing us to empirically assess the performance of each method and identify the value of combining IPD and AD in practice.

The results of estimates and their standard errors for the within-trial and across-trial treatment–covariate interaction effects, averaged across all possible combinations of IPD and AD trials in each scenario (e.g. we averaged across results from all 10 analyses in the scenario of two IPD trials), are shown in Table 2.

5.1. Within-trial and across-trial interaction effect

We firstly discuss the results of the full IPD analysis. In the situation where IPD are available from all five trials, estimates of the across-trial and the within-trial interaction effect were $\hat{\gamma}_A = -0.662$ (s.e. = 0.464 and p-value = 0.154) and $\hat{\gamma}_W = 0.087$ (s.e. = 0.055 and p-value = 0.114), respectively. The across-trial interaction effect was substantially different from the within-trial interaction effect on the point estimates by visual inspection. This shows the importance of separating the within-trial interaction from the across-trial interaction, as chance, confounding and/or ecological bias is causing the across-trial effect to act in the opposite direction of the within-trial effect here. Figure 3 also highlights the difference between $\hat{\gamma}_A$ and $\hat{\gamma}_W$. In Figure 3, the within-trial

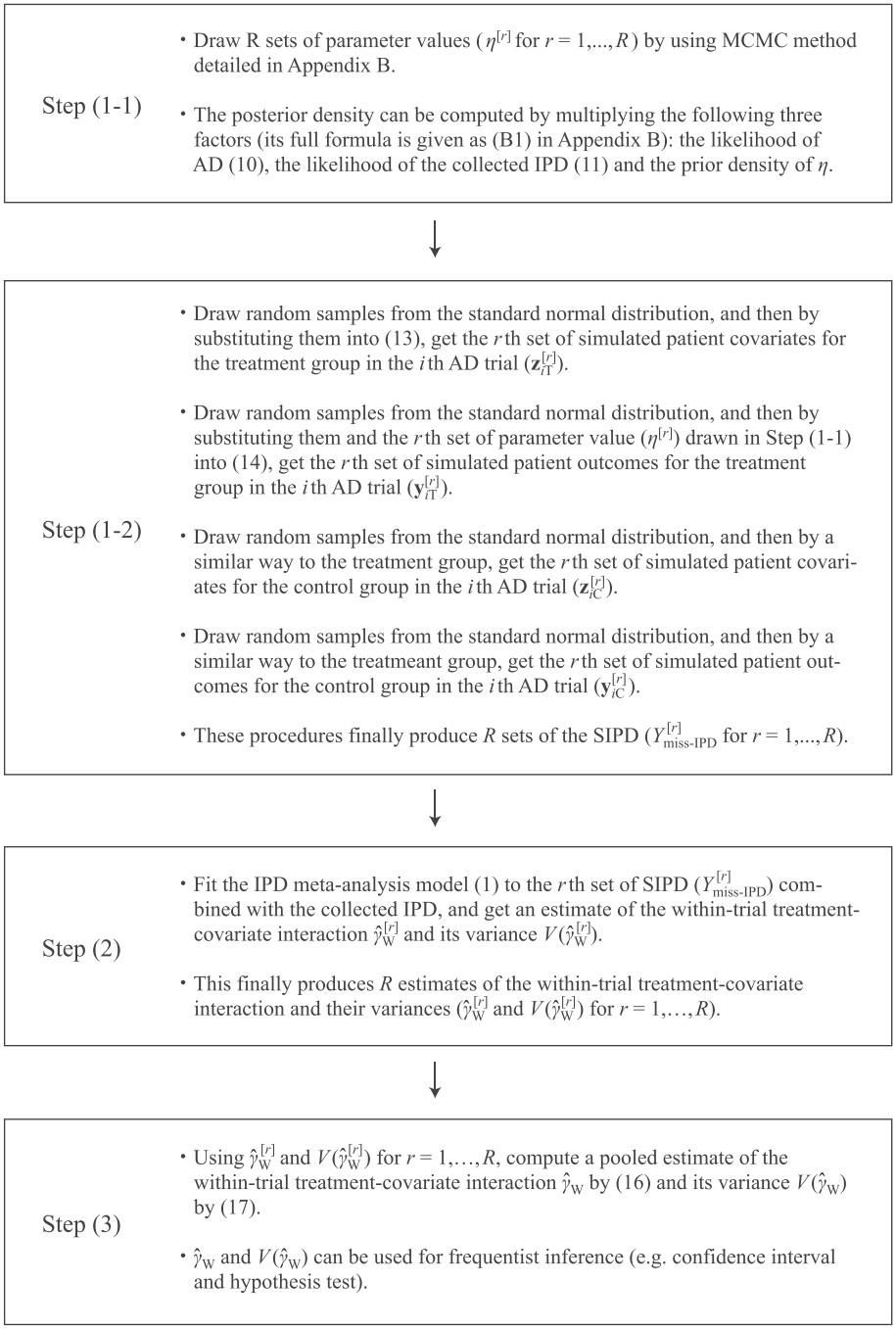

Figure 1. Step-by-step instruction for estimating within-trial treatment–covariate interaction by the SIPD method.

interaction estimated from trials with larger sample size have almost flat gradients (dashed lines), whereas the across-trial interaction has a negative gradient (solid line). If we used a model without separation of the across-trial and the within-trial interaction, we would get a potentially wrongly amalgamated result on the interaction between treatment and age. The standard error of $\hat{\gamma}_A$ was also much larger than that of $\hat{\gamma}_W$, because the number of trials was small, and the mean ages were fairly homogeneous across the five trials. There was no observed between-study heterogeneity in the within-trial interaction ($I^2 = 0\,\%$), and thus, the fixed effect assumption is also plausible for this parameter.

5.2. The gains from SIPD method

Now, consider situations involving a mixture of IPD and AD. In terms of the across-trial interaction, the SIPD method produced estimates closer to the full IPD analysis compared with model (3) or an only IPD analysis regardless of the number of AD trials. For each scenario in Table 2, we also found an important difference between

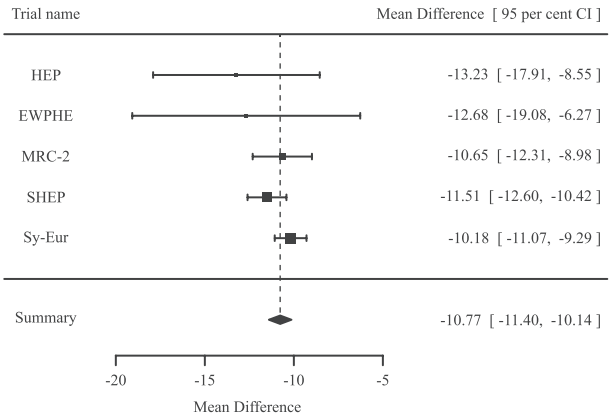

Figure 2. Forest plot of five trials in hypertension.

Table 2. Average of estimates and their standard errors for treatment–covariate interaction effect when analysing change in systolic blood pressure (follow-up minus baseline) from hypertension data, where estimates are averaged across all combinations of trials providing individual patient data.

		Average of estimate			Average of standard error		
	The number of trials providing IPD‡	Only IPD	Model (3)	SIPD	Only IPD	Model (3)	SIPD
Across-trial	5/5*	−0.662		—	0.464		—
interaction	4/5	−0.569	−0.766	−0.662	0.590	0.468	0.464
effect γ_A	3/5	−0.199	−0.766	−0.663	0.920	0.463	0.464
	2/5	0.293	−0.768	−0.665	1.831	0.464	0.464
	1/5	NA	−0.781	−0.667	NA	0.464	0.463
	0/5†	—	−0.766	—	—	0.466	—
Within-trial	5/5*	0.087		—	0.055		—
interaction	4/5	0.092	0.091	0.090	0.063	0.063	0.063
effect γ_W	3/5	0.116	0.117	0.096	0.084	0.084	0.076
	2/5	0.165	0.166	0.103	0.131	0.130	0.097
	1/5	0.252	0.244	0.105	0.259	0.258	0.131

IPD, individual patient data; SIPD, simulated IPD.
Only-IPD: fit model (1) to IPD from only the IPD trials available.
Model (3): fit model (3) to the mixture of IPD and AD.
SIPD: apply the SIPD method to the mixture of IPD and AD.
*Results by fitting model (1) to the full IPD from all five trials.
†Results by fitting the meta-regression model (2) to the AD from all five trials.
‡The numbers of combinations of trials providing IPD are five, ten, ten and five in the scenarios of one, two, three and four IPD trials, respectively.

results for the within-trial interaction effect from model (3) and the SIPD method. When comparing $\hat{\gamma}_W$ from model (3) with that from the full IPD analysis, model (3) provided point estimates located in a positive direction on average, with large standard errors. This is because model (3) allows only the IPD trials to estimate the within-trial interaction effect, and thus, the estimates and their standard error for γ_W by fitting model (3) got close to those from the full IPD analysis as the available number of IPD trials increases. The SIPD method improved both the estimates and standard errors of γ_W to be closer to the correct (full IPD) estimates, especially when the number of IPD trials was small. The most benefit comes in the scenario of one IPD trial, in which estimates of γ_W were $\hat{\gamma}_W = 0.244$ (s.e. = 0.258) from model (3) and $\hat{\gamma}_W = 0.105$ (s.e. = 0.131) from the SIPD method; the latter is much closer to the full IPD analysis result of $\hat{\gamma}_W = 0.087$ (s.e. = 0.055). This shows that the SIPD method allows both of the AD and IPD trials to estimate the within-trial interaction effect, and this adjustment on the basis of the AD trials is useful especially when the number of IPD trials is small.

Table 2 also shows that the differences between results from model (3) and the SIPD method become smaller when increasing the number of IPD trials. In particular, the results were similar in the scenario of three IPD trials and two AD trials, and almost equivalent in the scenario of four IPD trials and one AD trials.

For the scenarios of one or two IPD trials, Figure 4 shows estimates and their standard errors for γ_W obtained by the three methods for each combination, where the horizontal axis represents the name of each IPD trial with

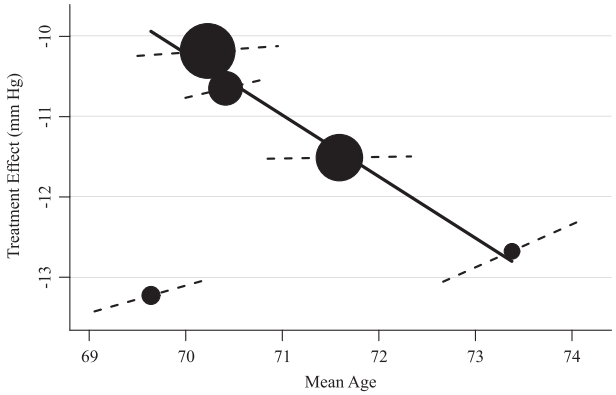

Figure 3. Scatter plot for the five trials in hypertension with across-trial and within-trial interaction effect estimates, in which a solid line represents the across-trial interaction ($\hat{\gamma}_A$) between mean age ($\bar{z}_i$) and treatment effect estimated by fitting model (2); dashed lines represent the within-trial interaction ($\hat{\gamma}_W$) between age and treatment effect estimated separately within each trial using IPD and model (1) without γ_A; the gradient of each dashed line indicates the change in treatment effect for a 1-year increase in age within each trial; the width of the dashed line about the centre of each circle is defined by 1 times the standard deviation of age in each trial; each circle represents a trial and is centred at $\bar{z}_i$ in each trial, and the circle size is proportional to the sample size in each trial.

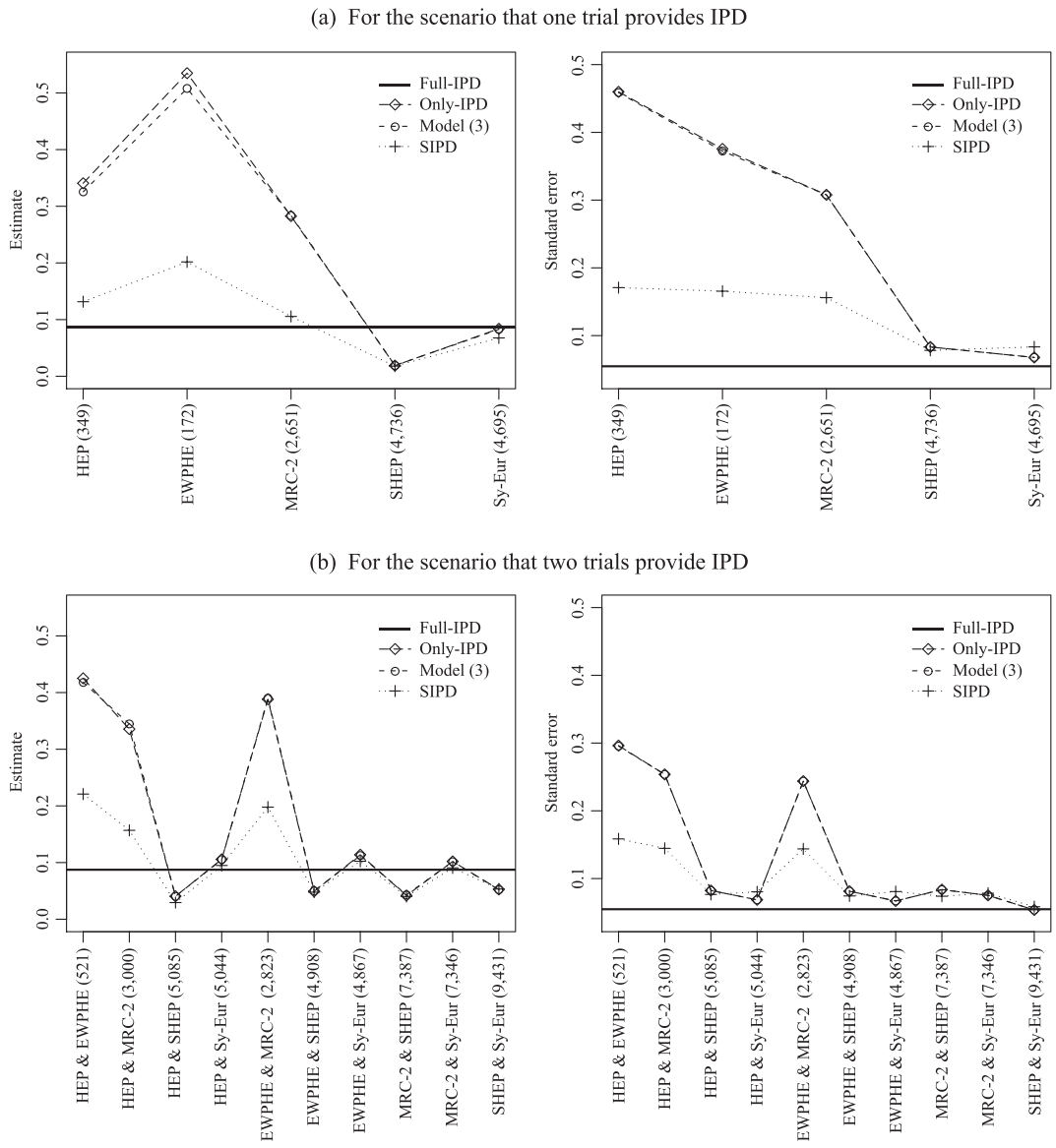

Figure 4. Estimates and their standard errors for within-trial treatment–covariate interaction effect when analysing change in SBP (follow-up minus baseline) from hypertension data in the scenarios that (a) one trial provides IPD and (b) two trials provide IPD.

sample size in parentheses (in the scenarios of two IPD trials, names of two IPD trials and sum of sample sizes from two IPD trials are shown). The heavy solid line represents the results from the full IPD analysis, and thus, the closer results to this line are regarded as superior ones in the sense of matching the full IPD analysis. The results from model (3) were equivalent to those from an only IPD analysis for each scenario, as both only use IPD available to estimate γ_W. For almost all combinations of one IPD trial, the SIPD method provided estimates and their standard errors for γ_W, which were located closer to those from the full IPD analysis than model (3). These were particularly considerable when the number of patients included in the IPD trial was small (e.g. HEP, EWPHE and MRC-2). Similar findings are seen for the scenario of two IPD trials, although the results by the three methods were closer.

The difference between the results from model (3) and the SIPD method is clearly dependent on the proportion of available IPD in all patients, not just the number of IPD trials, because the difference between methods decreased in the case of large sample size of IPD trials in Figure 4 (e.g. SHEP and Sy-Eur). For this viewpoint, we computed the number of patients involved in the IPD trials for all the 30 combinations from Table 2. Figure 5 shows estimates and their standard errors for γ_W from model (3) and the SIPD method for all the scenarios, which includes 30 results sorted by the proportion of patients involved in the IPD trials. As before, the difference between the results from model (3) and the SIPD method became larger when the proportion of patients involved in the IPD trials became smaller; in addition, the differences rapidly diminished when IPD for over 40% of patients was available. Thus, the SIPD method had most notable benefits when the proportion of patients involved in the IPD trials was low.

6. Simulation study

We performed a simulation study to verify the observed performance of the SIPD method. From the results in the application to the hypertension data, it was shown that the SIPD method provided estimates of the within-trial interaction effect closer to those from the full IPD analysis than model (3) when the number of IPD trials is small and when the number of patients involved in the IPD trials is small. To check this finding, we focus here on the within-trial interaction effect and compare some statistical properties of $\hat{\gamma}_W$ obtained by the SIPD method with those obtained by fitting model (3) under some settings of controlled parameters and the number of patients involved in one IPD trial and nine AD trials.

6.1. Design

We considered that the true models for generating individual outcome and covariate values from patients in each trial were written as follows:

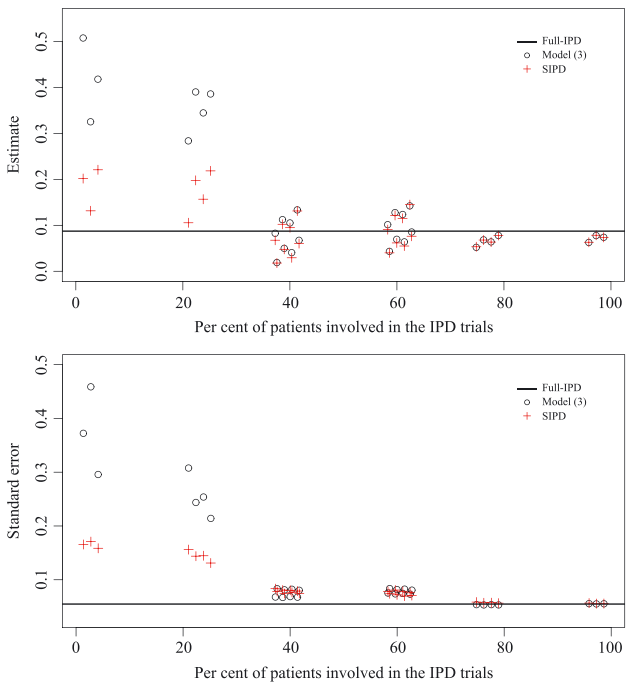

Figure 5. Estimates and their standard errors for within-trial treatment–covariate interaction effect sorted by the proportion of available IPD in all patients when analysing change in SBP (follow-up minus baseline) from hypertension data.

$$
\begin{aligned}
&x_{ij} = \begin{cases} 0, & j \in C \\ 1, & j \in T \end{cases}, \\
&z_{ij} | m_{zi} \sim N\left(m_{zi}, \sigma_z^2\right), \quad m_{zi} \sim N\left(\hat{m}_z, \hat{\sigma}_m^2\right), \\
&y_{ij} | x_{ij}, z_{ij}, \bar{z}_i \sim N\left(\hat{\phi}_i + \hat{\theta} x_{ij} + \hat{\mu} z_{ij} + \hat{\gamma}_A x_{ij} \bar{z}_i + \gamma_W x_{ij} \left(z_{ij} - \bar{z}_i\right), \sigma_y^2\right), \\
&i = 1, \ldots, 10; j = 1, \ldots, n_i
\end{aligned}
\tag{18}
$$

where the true parameters except for γ_W and σ_y^2 were given as estimates by fitting model (1) to IPD from 10 trials originally reported in Wang *et al.* (2005), with change in SBP as an outcome; for example, $\hat{\theta} = -4.958$, $\hat{\mu} = -0.042$ and $\hat{\gamma}_A = -0.079$. The total number of patients was given by $\sum_{i=1}^{10} n_i = 6000$, and each group had the same sample size as $n_{iT} = n_{iC} = n_i/2$. $\hat{m}_z$ and $\hat{\sigma}_m^2$ are the mean covariate value across all the 10 trials and its variance, based on the fact that Wang's data gave $\hat{m}_z = \sum_{i=1}^{10} \bar{z}_i/10 = 62.69$ and $\hat{\sigma}_m^2 = \sum_{i=1}^{10} (\bar{z}_i - \hat{m}_z)^2/(10-1) = 180.8$. We supposed that only one trial provided IPD (the other nine trials provided AD) and controlled the number of patients involved in the one IPD trial by six scenarios of 60, 300, 600, 1200, 2400 and 4800 so that the proportions of patients with available IPD were given by 1%, 5%, 10%, 20%, 40% and 80%, respectively. The nine AD trials involved almost the same number of patients for each scenario.

In the true model (18), inference of $\hat{\gamma}_W$ is mainly affected by: the true value of γ_W, the variance parameter in within-trial covariate distributions, σ_z^2, and the variance parameter in conditional distributions of y_{ij} given z_{ij}, σ_y^2. We here considered $\sigma_z^2 = 100$ for the nine AD trials and controlled that for the one IPD trial by three scenarios of $\sigma_z^2 \in \{25, 50, 100\}$. The mean square error (MSE) and the standard error of $\hat{\gamma}_W$ are expected to become smaller as σ_z^2 increases (Simmonds and Higgins, 2007). Thus, these scenarios for σ_z^2 lead us to a situation that the one IPD trial provides information on γ_W less than or equal to the other AD trials. We also gave $\gamma_W = -0.2$ and $\sigma_y^2 = 200$ so that the power to detect γ_W estimated from the full IPD analysis becomes high enough for each scenario.The implementing procedure was as follows. Firstly, we set the number of patients involved in the IPD trial and σ_z^2 for each of 18 scenarios and then generated 5000 sets of meta-analysis data according to (18) for each scenario. More specifically, we generated mean covariate for the ith trial, m_{zi}, from $N\left(\hat{m}_z, \hat{\sigma}_m^2\right)$ and covariate values for patients in the ith trial from $N\left(m_{zi}, \sigma_z^2\right)$ given m_{zi} and σ_z^2, and then outcome values for patients given covariate values. Secondly, for each set in each scenario, we summarised IPD for nine of the 10 trials to AD. Finally, we analysed the mixture of IPD and AD by two methods: model (3) and the SIPD method. In each analysis, we computed MSE, mean bias and mean standard error for $\hat{\gamma}_W$. We also computed sample mean of absolute differences between estimates of γ_W obtained by fitting model (3) or the SIPD method and those obtained from the full IPD analysis, which was intended to evaluate how far the point estimate obtained by fitting model (3) or the SIPD method is apart from that obtained from the full IPD analysis on average. Moreover, we estimated the type I error rate and the power with one-sided hypothesis test at 5% level of significance for $H_0 : \gamma_W = 0$ and $H_1 : \gamma_W < 0$. The settings of MCMC sampling in the SIPD method were as described in Appendix B.

6.2. *Result*

The results of MSE, mean bias and mean standard error for each scenario are shown in Table 3. Note that σ_z^2 in Table 3 represents the scenarios of variance parameter in the within-trial covariate distributions for patients in one IPD trial, and the true value of σ_z^2 in all AD trials was equal to 100. In each scenario, the SIPD method provided substantially smaller MSEs and mean standard errors in comparison with model (3), especially when the proportion of patients with available IPD was low (e.g. 1% or 5%) and σ_z^2 was small (e.g. $\sigma_z^2 = 25$). The results of the absolute differences also show that the point estimates of γ_W from the SIPD method were on average located closer to those from a full IPD analysis (of all 10 trials) than model (3). The difference between the results from model (3) and the SIPD method was the largest for the scenario of 1% of patients with available IPD and $\sigma_z^2 = 25$, and became smaller as the proportion of patients with available IPD was higher and σ_z^2 increased. These indicate that the SIPD method could adjust the estimate and its standard error of γ_W from the IPD-only analysis closer to those from the full IPD analysis using additional information from the AD trials, especially when the sample size of the IPD trial was small and the variation in patients covariate within the IPD trial was small. For example, in the scenario of 5% of patients in the IPD trial and $\sigma_z^2 = 25$, the MSE was reduced by 50% using the SIPD method (MSE = 0.051) rather than model (3) (MSE = 0.111); similarly, the standard error was reduced considerably by using SIPD (mean s.e. = 0.248) rather than model (3) (mean s.e. = 0.328). In addition, all of the results indicate that the SIPD method could be robust against a situation where the mean covariate values moderately vary across trials.

However, in the scenarios of 1% and 5% of patients with available IPD, the estimates of γ_W from the SIPD method were subject to a positive bias. For example, the mean bias from the SIPD method for the scenario of 1% of patients with available IPD and $\sigma_z^2 = 25$ was 0.107, and thus, $\hat{\gamma}_W$ was larger than the true value of $\gamma_W = -0.2$ on average. This is due to the influence of the information on the within-trial relationships from the AD trials. The

Table 3. Mean square errors, mean biases and mean standard errors for estimator of within-trial treatment-covariate interaction effect, and sample means of absolute differences between estimates from model (3) and the simulated individual patient data (IPD) method (for IPD from one trial and aggregate data from nine trials) and those from the full IPD analysis (for IPD from all 10 trials).

Per cent of total patients in the IPD trial	σ_z^2	Mean square error of $\hat{\gamma}_W$			Mean bias of $\hat{\gamma}_W$			Mean standard error of $\hat{\gamma}_W$			Mean of absolute difference	
		Model ((3))	SIPD	(Full-IPD)*	Model ((3))	SIPD	(Full-IPD)*	Model ((3))	SIPD	(Full-IPD)*	Model ((3))	SIPD
1%	25	0.612	0.098	(0.001)	−0.005	0.107	(0.000)	0.744	0.411	(0.037)	0.622	0.252
	50	0.303	0.088	(0.001)	0.011	0.082	(0.000)	0.526	0.350	(0.037)	0.434	0.235
	100	0.147	0.067	(0.001)	0.002	0.050	(0.000)	0.374	0.281	(0.037)	0.303	0.200
5%	25	0.111	0.051	(0.001)	−0.008	0.041	(0.000)	0.328	0.248	(0.037)	0.264	0.176
	50	0.054	0.035	(0.001)	−0.001	0.020	(0.000)	0.232	0.192	(0.037)	0.184	0.143
	100	0.026	0.020	(0.001)	0.001	0.008	(0.000)	0.164	0.145	(0.037)	0.127	0.108
10%	25	0.055	0.033	(0.001)	0.002	0.020	(0.000)	0.231	0.191	(0.038)	0.184	0.141
	50	0.027	0.020	(0.001)	−0.001	0.007	(0.001)	0.164	0.145	(0.038)	0.128	0.109
	100	0.013	0.011	(0.001)	0.000	0.003	(0.000)	0.116	0.108	(0.037)	0.087	0.080
20%	25	0.027	0.021	(0.002)	0.002	0.011	(0.000)	0.163	0.145	(0.040)	0.125	0.109
	50	0.014	0.012	(0.001)	0.000	0.003	(0.000)	0.116	0.109	(0.039)	0.088	0.080
	100	0.007	0.006	(0.001)	−0.001	0.001	(0.000)	0.082	0.080	(0.037)	0.059	0.055
40%	25	0.013	0.011	(0.002)	0.002	0.004	(0.000)	0.116	0.110	(0.044)	0.086	0.079
	50	0.007	0.006	(0.002)	0.000	0.002	(0.001)	0.082	0.080	(0.041)	0.055	0.053
	100	0.003	0.003	(0.001)	−0.001	0.000	(−0.001)	0.058	0.057	(0.037)	0.035	0.034
80%	25	0.007	0.006	(0.003)	0.000	0.001	(0.000)	0.082	0.081	(0.058)	0.046	0.045
	50	0.003	0.003	(0.002)	0.001	0.001	(0.000)	0.058	0.058	(0.047)	0.026	0.026
	100	0.002	0.002	(0.001)	0.000	0.001	(0.000)	0.041	0.041	(0.037)	0.014	0.014

IPD, individual patient data; SIPD, simulated IPD.
Model (3): fit model (3) to the mixture of IPD and AD, SIPD: apply the SIPD method to the mixture of IPD and AD.
σ_z^2: Variance parameter in within-trial covariate distributions for patients in one IPD trial.
*Results by fitting model (1) to the full IPD from all trials.

SIPD method allows us to extract the information on the within-trial relationships from the AD trial by using (10), and thus, we gain substantially smaller MSEs and mean standard errors for $\hat{\gamma}_W$ in comparison with model (3). On the other hand, this information from the AD trials also pulls the estimates of γ_W in a positive direction when the proportion of patients with available IPD is extremely low. Therefore, in scenarios of less 10% of patients with available IPD, there is a trade-off; the large gain in MSE and standard error comes at the expense of a bias. The bias is negligible in all methods for 10% or over, and the SIPD method still has gain in MSE (up to about 40%) and standard error (up to about 20%) in situations between 10% and 40% patients with available IPD.

Figure 6 also shows the type I error rates and the power for $\hat{\gamma}_W$ estimated by the three methods for each scenario. In the scenarios of 1% and 5% of patients with available IPD, the type I error rates from the SIPD method were highly conservative. Further, when the true within-study interaction was zero, the SIPD method did not produce biassed estimates of γ_W unlike when γ_W was -0.2 (results not shown in Table 3). Therefore, the conservative type I error rates for the SIPD method are likely due to overestimated standard errors of $\hat{\gamma}_W$, even though the standard errors were smaller than those from model (3). In the scenarios of over 10% of available IPD, the SIPD method had better type I error rates close to 5%. The powers of model (3) and the SIPD method to detect the true negative interaction were very similar. The SIPD method was marginally better when 10% to 40% of patients were in the IPD trial.

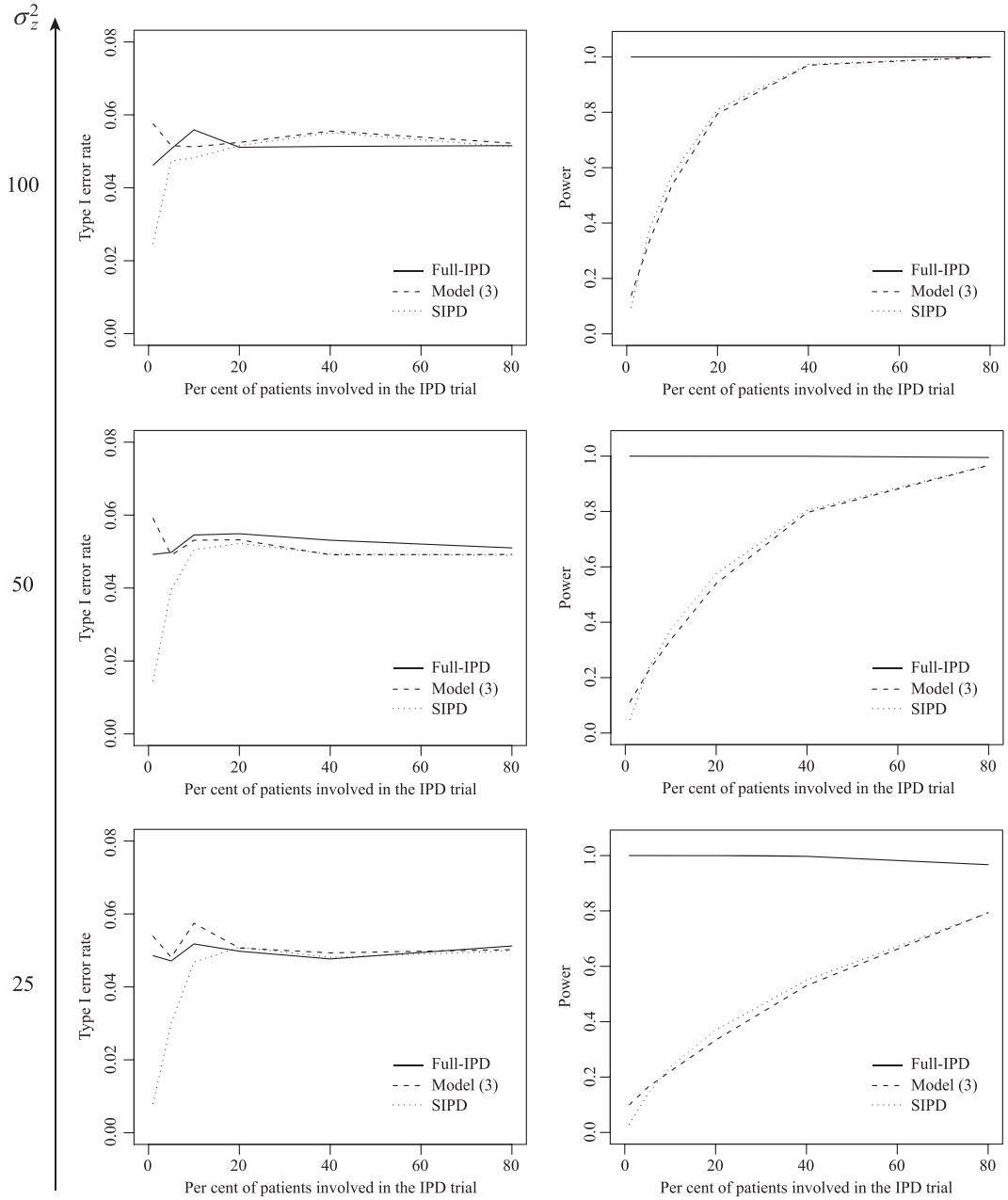

Figure 6. Type I error rates (three panels on the left side) and powers (three panels on the right side) for within-trial treatment–covariate interaction effect.

7. Discussion

We proposed a meta-analysis method for combining IPD and AD in the simple situation of a single continuous outcome and covariate. IPD meta-analysis has been advocated by many researchers, while the methodological development for combining IPD and AD becomes increasingly important because practitioners cannot always collect the IPD for all trials involved (Riley *et al*., 2007; Ahmed *et al*., 2012). Reducing available IPD to AD and focusing on just the across-trial relationship leads to a loss of information and potential bias, and it is important to focus on the within-trial relationship as much as possible. Indeed, the within-trial and the across-trial relationships were estimated in opposite directions in the hypertension data, emphasising the importance of separating them out (Riley *et al*., 2008).

The SIPD method proposed offers a new framework for using meta-analysis to estimate treatment–covariate interactions while separating within-trial and across-trial relationships. Through the application to the hypertension data including five trials with similar mean ages, we demonstrated that the SIPD method provided results for the within-trial interaction effect closer to those from the full IPD analysis than model (3) by Riley *et al*., (2008). The most beneficial results were given for the cases when the number of trials providing IPD was small or the proportion of patients with available IPD was low. In such situations, the collected IPD trials may offer very little information on the within-trial relationships, causing model (3) to yield estimates of the within-trial interaction with large standard errors. By contrast, the SIPD method utilises additional information from the AD trials and, in comparison with model (3), can provide estimates and standard errors closer to those from a full IPD analysis. This is particularly true when given over 10% and under 40% of patients in the IPD trials, as the adjusted estimators from the SIPD method were unbiased and had smaller MSEs and standard errors for these situations in our simulation. The simulation study also suggested that the SIPD method was acceptable to meta-analyses of trials with different mean covariate values, which is a more common situation than the hypertension data with the similar mean covariates across trials. In future research, we need to apply and compare the methods to other examples, in particular, where more than one IPD trial is to be combined and where the mean covariate values vary across trials.

However, the simulation study revealed some limitations of the SIPD method. In particular, the adjustment by using the AD trials gave a bias in estimator for the within-trial interaction effect in the cases when the proportion of patients with available IPD was under 10%. And also, in the same situations, the SIPD method suffered from the conservative type I error rates of the within-trial interaction effect, because the standard errors from the SIPD method were overestimated. However, in situations with over 10% of patients in the IPD trials, the SIPD method performed well.

The SIPD method assumes that the AD trials are similar (exchangeable) to the IPD trials apart from having a separate baseline (intercept). This strong exchangeability assumption means that, conditional on the AD available (means and standard deviations of the treatment groups), the missing IPD can be simulated by drawing from the predictive distribution of the missing information (such as the within-trial interaction, γ_W, and the covariate effect, μ) as informed by the IPD trials. Also, the SIPD method can incorporate information on the within-trial relationships from the AD trials in the predictive distribution of the missing IPD. This is because we first assume an IPD meta-analysis model (1) for not only the collected IPD but also the missing IPD and then marginalise their density with respect to the missing IPD. By doing so, the simulated IPD can be combined with the collected IPD, across multiple simulations using Rubin's rules, to provide estimates that allow stronger inferences to be made, especially with less than 40% of the patients available in the IPD obtained. Of course, we need further simulation studies to verify the performance of the SIPD method in more common circumstances (e.g. more than one IPD trial).

However, we recognise that our SIPD method makes strong exchangeability assumptions and, as it stands, is only applicable to a narrow range of situations. In particular, it assumes that the treatment effect is fixed across trials. It would be useful to extend the method to random effects models to allow for heterogeneity if possible (Higgins *et al*., 2009) and also allow a trial-specific covariate effect (μ_i) and a trial-specific error variance (σ^2_{yi}). Indeed, a meta-analysis of the full 10 trials in the hypertension data originally reported by Wang *et al*. (2005) would potentially require this kind of modelling (Riley *et al*., 2008). Especially when nonrandomised and/or observational evidence are considered in combination with the randomised evidence, these assumptions should be relaxed for modelling the difference of evidence types (e.g. Bayesian hierarchical models of observational and randomised evidence). Our methods therefore need extension and investigation for this situation. We expect that the model used here for generating the SIPD could be extended to include the random treatment effect by another way for drawing from the posterior predictive distribution of the missing IPD, which is known as data augmentation (Tanner and Wong, 1987; van Dyk and Meng, 2001). The data augmentation takes an iterative sampling procedure while alternately updating for the missing IPD and the parameters. This approach might offer an easier derivation of the approximate posterior distribution than the method proposed here. Riley *et al*. (2013) notes that when there is baseline imbalance, a meta-analysis of randomised trials with a continuous outcome should use analysis of covariance, and we welcome consideration of the SIPD method to this situation. Some technical issues are caused when our focus is on assessing how patient characteristics modify the treatment effect while adjusting for baseline imbalance; in particular, this requires more complicated calculation for deriving the

posterior distribution of parameters. We expect that some numerical methods or simulation-based calculations could be used for this. Once the posterior distribution of parameters can be derived, it is not difficult to apply the framework of SIPD method in the same manner. Moreover, in the SIPD method proposed here, we assume that the covariate is normally distributed. It would be necessary to discuss how sensible the results from the SIPD method are with respect to this assumption. The inclusion of some other trial-level covariates will also be considered in further research. Trial-level covariates can also be included into model (5), and then one can use the same implementing procedure shown in Figure 1 for estimating parameters of interest. In addition, the SIPD method has a practical limitation on account that we only consider models for estimating one interaction. Of course in practice, multiple interactions might be of interest. Modelling multiple covariates and multiple interactions would cause a computational complexity as is the case in considering the analysis of covariance model. Nonetheless, where the assumed criteria are considered plausible or worth consideration in a sensitivity analysis, the SIPD approach is a promising method for meta-analysts faced with combining IPD and AD.

For our future research, the proposed method might be most useful as a sensitivity analysis, to examine if the conclusions about the interaction effect remain robust to the inclusion of SIPD. Of course, in order to use the proposed method as the sensitivity analysis in practice, we need to extend the method to allow for more common circumstances (e.g. the random treatment effects). Furthermore, the method could be used for a subgroup analysis in the context of systematic reviews. When we aim to explore beneficial effects in specific patient subgroups; for example, subgroups of very old patients (more than 80 years of age) or not in hypertension, we can allocate patients in the IPD trials and simulated patients in the AD trials to each subgroup and then meta-analyse these subgroups within each trial. We also expect that the method could be extended to allow for different types of data. Especially when considering a situation where a single binary outcome and a single binary covariate are observed for each patient in each trial, we could reconstruct the missing binary outcome and covariate by using a likelihood function on the basis of convolution of binomial distributions.

Appendix A

The difficulty when getting (10) from (9) is to integrate the density over a restricted sample space. Using the notations of $\left(m_{y_iT}, \sigma^2_{y_iT}, m_{y_iC}, \sigma^2_{y_iC}\right)$ and $(\mathbf{y}_{iT}, \mathbf{y}_{iC}, \mathbf{z}_{iT}, \mathbf{z}_{iC})$ in the same way as those in Sections 4.1.1 and 4.1.2, respectively, we can expand the equation (9) as

$$\begin{aligned} f(Y_{\mathrm{AD}}|\hat{\xi}, \eta) &= \int_{h(Y_{\text{miss-IPD}})=Y_{\mathrm{AD}}} f(Y_{\text{miss-IPD}}|\hat{\xi}, \eta) dY_{\text{miss-IPD}} \\ &= \prod_{i=1}^{N} \left[\int_{h(\mathbf{y}_{iT})=\left(\bar{y}_{iT}, s^2_{y_iT}\right)} \left(\int f(\mathbf{y}_{iT}|\mathbf{z}_{iT}, \eta) f(\mathbf{z}_{iT}|\hat{\xi}) d\mathbf{z}_{iT} \right) d\mathbf{y}_{iT} \right. \\ &\quad \left. \times \int_{h(\mathbf{y}_{iC})=\left(\bar{y}_{iC}, s^2_{y_iC}\right)} \left(\int f(\mathbf{y}_{iC}|\mathbf{z}_{iC}, \eta) f(\mathbf{z}_{iC}|\hat{\xi}) d\mathbf{z}_{iC} \right) d\mathbf{y}_{iC} \right] \\ &= \prod_{i=1}^{N} \left[\int_{h(\mathbf{y}_{iT})=\left(\bar{y}_{iT}, s^2_{y_iT}\right)} \left(\prod_{j \in T} \frac{1}{\sqrt{2\pi}\sigma_{y_iT}} exp\left\{ -\frac{\left(y_{ij} - m_{y_iT}\right)^2}{2\sigma^2_{y_iT}} \right\} \right) d\mathbf{y}_{iT} \right. \\ &\quad \left. \times \int_{h(\mathbf{y}_{iC})=\left(\bar{y}_{iC}, s^2_{y_iC}\right)} \left(\prod_{j \in C} \frac{1}{\sqrt{2\pi}\sigma_{y_iC}} exp\left\{ -\frac{\left(y_{ij} - m_{y_iC}\right)^2}{2\sigma^2_{y_iC}} \right\} \right) d\mathbf{y}_{iC} \right] \end{aligned} \tag{A1}$$

where $h(\mathbf{y}_{iT}) = (\bar{y}_{iT}, s^2_{y_iT})$ and $h(\mathbf{y}_{iC}) = (\bar{y}_{iC}, s^2_{y_iC})$ are the functions to summarise patient-level observations to sample means and sample variances for each group. Because of between-trial and between-group independence, the problem to be considered here is reduced to integrate the normal density over the region with the fixed sample mean and sample variance; for example, for the treatment group in the *i*th trial:

$$\int_{h(\mathbf{y}_{iT})=\left(\bar{y}_{iT}, s^2_{y_iT}\right)} \left(\prod_{j=1}^{n_{iT}} \frac{1}{\sqrt{2\pi}\sigma_{y_iT}} exp\left\{ -\frac{\left(y_{ij} - m_{y_iT}\right)^2}{2\sigma^2_{y_iT}} \right\} \right) d\mathbf{y}_{iT}. \tag{A2}$$

Tsiatis (2006) gave a general calculation to solve this problem. We now brief how to calculate (A2) explicitly. The calculation can be applied to the integration for the other trials and the control group in the same way.

Assume there exists a $(n_{iT} - 2)$-dimensional variable $g(\mathbf{y}_{iT})$ that

$$\mathbf{y}_{iT} \leftrightarrow \left\{ \bar{y}_{iT}, s^2_{y_i T}, g(\mathbf{y}_{iT}) \right\}$$

is one-to-one for all $(\bar{y}_{iT}, s^2_{y_i T})$, and let $f(\bar{y}_{iT}, s^2_{y_i T}, g(\mathbf{y}_{iT}))$ denote a probability density of $(\bar{y}_{iT}, s^2_{y_i T}, g(\mathbf{y}_{iT}))$. Then, from a result by Tsiatis (2006), the integration (A2) is equivalent to an integration of $f(\bar{y}_{iT}, s^2_{y_i T}, g(\mathbf{y}_{iT}))$ with respect to $g(\mathbf{y}_{iT})$; that is,

$$(A2) = \int f(\bar{y}_{iT}, s^2_{y_i T}, g(\mathbf{y}_{iT})) dg(\mathbf{y}_{iT}). \tag{A3}$$

This indicates that once we find the transformation $g(\mathbf{y}_{iT})$ and the density $f(\bar{y}_{iT}, s^2_{y_i T}, g(\mathbf{y}_{iT}))$, we can get (A2) explicitly. This calculation is achieved using a result by Pullin (1979), in which Helmert's transformation of $\mathbf{y}_{iT}$ and furthermore sine and cosine transformations are required. Finally, we have the desired integration as

$$(A3) \propto \frac{1}{(\sigma^2_{y_i T}) n_{iT}/2} \exp\left\{ -\frac{n_{iT}}{2\sigma^2_{y_i T}} \left\{ (\bar{y}_{iT} - m_{y_i T})^2 + s^2_{y_{iT}} \right\} \right\}. \tag{A4}$$

Then, the equation (10) is given by substituting (A4) and those for the other trials and the control group into (A1).

Appendix B

The Metropolis–Hastings algorithm to get $\eta^{[r]}$ in the SIPD method takes the following implementing steps (Gelman *et al.*, 1995):

(1) Set a starting value $\eta^{[0]}$, and iterate steps (2)–(4) for $r = 1, \ldots, R$.
(2) Draw a sample η^* from a proposal distribution with density function $q(\eta|\eta^{[r-1]})$; that is,

$$\eta^* \sim q(\eta|\eta^{[r-1]}).$$

(3) We here assume $q(\eta|\eta^{[r-1]})$ as a normal density centred at $\eta^{[r-1]}$.
(4) Compute

$$\alpha = min\left\{ 1, \frac{f\left(\eta^*|Y_{\text{AD}}, \hat{\xi}\right) f\left(\eta^*|Y_{\text{IPD}}, \hat{\xi}\right)}{f\left(\eta^{[r-1]}|Y_{\text{AD}}, \hat{\xi}\right) f\left(\eta^{[r-1]}|Y_{\text{IPD}}, \hat{\xi}\right)} \right\}.$$

(5) Set $\eta^{[r]} = \eta^*$ with probability α, otherwise set $\eta^{[r]} = \eta^{[r-1]}$.

Here, the posterior density of $f(\eta|Y_{\text{AD}}, \hat{\xi}) f(\eta|Y_{\text{IPD}}, \hat{\xi})$ can be written in full as

$$\begin{aligned} f(\eta|Y_{\text{AD}}, \hat{\xi}) f(\eta|Y_{\text{IPD}}, \hat{\xi}) &= f\left(\phi_1, \ldots, \phi_{N'}, \theta, \mu, \gamma_A, \gamma_W, \sigma^2_y | Y_{\text{AD}}, Y_{\text{IPD}}, \hat{\xi}\right) \\ &\propto \prod_{i=1}^{N} \left[\left((\mu + \gamma_W)^2 \hat{\sigma}^2_{z_i T} + \sigma^2_y \right)^{-n_{iT}/2} \exp\left\{ -\frac{n_{iT}}{2\left((\mu + \gamma_W)^2 \hat{\sigma}^2_{z_i T} + \sigma^2_y \right)} \right. \right. \\ &\quad \left. \times \left((\bar{y}_{iT} - \phi_i - \theta - \mu \hat{m}_{z_i T} - \gamma_A \bar{z}_i - \gamma_W (\hat{m}_{z_i T} - \bar{z}_i))^2 + s^2_{y_i T} \right) \right\} \\ &\quad \left. \times \left(\mu^2 \hat{\sigma}^2_{z_i C} + \sigma^2_y \right)^{-n_{iC}/2} \exp\left\{ -\frac{n_{iC}}{2\left(\mu^2 \hat{\sigma}^2_{z_i C} + \sigma^2_y \right)} \left((\bar{y}_{iC} - \phi_i - \mu \hat{m}_{z_i C})^2 + s^2_{y_i C} \right) \right\} \right] \\ &\quad \times \prod_{i=N+1}^{N'} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{ -\frac{1}{2\sigma^2_y} \left(y_{ij} - \phi_i - \theta x_{ij} - \mu z_{ij} - \gamma_A x_{ij} \bar{z}_i - \gamma_W x_{ij} \left(z_{ij} - \bar{z}_i \right) \right)^2 \right\}. \end{aligned} \tag{B1}$$

For the iterative process of the algorithm to draw $R = 500$ values of parameter, we discarded the first 5000 samples (so-called burn-in samples) in order to prevent dependence on the starting values. Moreover, we took a sample at only

every 1000th iteration in order to avoid autocorrelation between the samples taken. In total, 505 000 samples of parameter were drawn. We used variance estimates from only-IPD meta-analyses (fitting the IPD meta-analysis model to the collected IPD) as the variances of the normal candidate-generating distribution. And also, we just graphically checked whether the burn-in samples were sufficient or the MCMC chain converged. In the application to hypertension data, for each possible combination of IPD and AD trials, we repeatedly displayed trace plots of parameter values drawn by the MCMC while controlling some MCMC settings (e.g. the number of burn-in samples and the thinning interval) and then decided the appropriate MCMC settings.

Appendix C

In the SIPD method, the missing IPD must be drawn from the conditional distribution given AD and the collected IPD; however, the density of this conditional distribution is difficult to be expressed exactly. This is associated with some issues on the conditional distribution given the sufficient statistics, discussed by Cheng (1984), Engen and Lillegard (1997) and Lindqvist and Taraldsen (2005). They gave general formula to calculate the conditional expectation on the basis of the conditional distribution given the sufficient statistics. Especially, an issue of sampling from the conditional distributions was considered. We here brief this approach, proposed by Lindqvist and Taraldsen (2005), for a simple case of univariate normal distribution.

Now, let $X=(X_1,\ldots,X_K)$ denote random variables following a normal distribution with mean μ and variance σ^2. Here, $T=\left(\overline{X},S_X\right)$ is the sufficient statistics for $\theta=(\mu,\sigma)$, where $\overline{X}=n^{-1}\sum_{k=1}^{K}X_k$ and $S_X^2=(K-1)^{-1}\sum_{k=1}^{K}\left(X_k-\overline{X}\right)^2$. Let $U=(U_1,\ldots,U_K)$ denote random variables following standard normal distribution, and two functions of χ and τ are defined by

$$\chi(U,\theta)\equiv(\mu+\sigma U_1,\ldots,\mu+\sigma U_K),$$
$$\tau(U,\theta)\equiv(\mu+\sigma\overline{U},\sigma S_U)$$

where $\overline{U}$ and S_U^2 stand for the mean and the variance similarly defined to $\overline{X}$ and S_X^2. Then, there exists unique χ and τ so that the joint distribution of $(\chi(U,\theta),\tau(U,\theta))$ is equivalent to those of (X,T) under the parameter θ. This means that, for given $t=(\overline{x},s_x)$ and U, $\hat{\theta}\equiv\hat{\theta}(U,t)$ in which $\tau(U,\hat{\theta})$ is held is uniquely determined as follows:

$$\hat{\theta}(U,t)\equiv(\hat{\mu}(U,t),\hat{\sigma}(U,t))=\left(\overline{x}-\frac{\overline{U}}{S_U}s_x,\frac{1}{S_U}s_x\right).$$

Thus, the random variable following the conditional distribution of X given $T=t$ is provided as follows:

$$X_t=\chi\left(U,\hat{\theta}(U,t)\right)=\left(\overline{x}+\frac{U_1-\overline{U}}{S_U}s_w,\ldots,\overline{w}+\frac{U_K-\overline{U}}{S_U}s_x\right). \tag{C1}$$

It is easily shown that the probability distribution of X_t is actually equivalent to the conditional distribution of X given $T=t$. Finally, sampling procedure from the conditional distribution given the mean and the variance is as follows: (i) generate random numbers $u=(u_1,\ldots,u_K)$ of U, (ii) substituting u and t to equation (C1), we get

$$x_t=\left(\overline{x}+\frac{u_1-\overline{u}}{s_u}s_x,\ldots,\overline{x}+\frac{u_K-\overline{u}}{s_u}s_x\right).$$

This result naturally leads the SIPD written by (13) and (14).

Appendix D

In step (3), we use the rule by Rubin (1987) to get the overall estimate $\hat{\gamma}_W$ and its variance $V(\hat{\gamma}_W)$. We here brief the calculations for them. In regard to γ_W, the posterior distribution of γ_W, say $\pi(\gamma_W|Y_{AD},Y_{IPD})$, is approximated by using simulated values of $Y_{miss\text{-}IPD}$ combined with Y_{IPD} as follows:

$$\pi(\gamma_W|Y_{AD},Y_{IPD})\approx\frac{1}{R}\sum_{r=1}^{R}\pi\left(\gamma_W|Y_{miss\text{-}IPD}^{[r]},Y_{IPD}\right). \tag{D1}$$

This indicates that the posterior distribution of γ_W can be simulated by first drawing $Y_{miss\text{-}IPD}^{[r]}$ from $f(Y_{miss\text{-}IPD}|Y_{AD},Y_{IPD})$ and then drawing γ_W from $\pi(\gamma_W|Y_{miss\text{-}IPD}^{[r]},Y_{IPD})$. Therefore, once we get R sets of SIPD, the posterior mean and variance can be approximated using (D1) as follows (Little and Rubin, 2002):

$$E(\gamma_W|Y_{AD}, Y_{IPD}) = E[E(\gamma_W|Y_{miss\text{-}IPD}, Y_{IPD})|Y_{AD}, Y_{IPD}]$$
$$\approx \frac{1}{R}\sum_{r=1}^{R}\hat{\gamma}_W^{[r]} \quad \text{(D2)}$$

and

$$Var(\gamma_W|Y_{AD}, Y_{IPD}) = E[Var(\gamma_W|Y_{miss\text{-}IPD}, Y_{IPD})|Y_{AD}, Y_{IPD}] + Var[E(\gamma_W|Y_{miss\text{-}IPD}, Y_{IPD})|Y_{AD}, Y_{IPD}]$$
$$\approx \frac{1}{R}\sum_{r=1}^{R} V\left(\hat{\gamma}_W^{[r]}\right) + \frac{1+R^{-1}}{R-1}\sum_{r=1}^{R}\left(\hat{\gamma}_W^{[r]} - \hat{\gamma}_W\right)^2 \quad \text{(D3)}$$

where $\hat{\gamma}_W^{[r]}$ and $V(\hat{\gamma}_W^{[r]})$ are the estimate of γ_W and its variance from the rth SIPD, respectively, and $\hat{\gamma}_W$ is given by (D2). For frequentist inferences, we use (D2) as the overall estimate $\hat{\gamma}_W$, and (D3) as its variance estimate $V(\hat{\gamma}_W)$.

Appendix E

The R source code to implement the SIPD method is shown here. Users can apply the SIPD method to their own problems by running the following code from R. We now ask the users to prepare some R objects: raw.ipd is a data frame for IPD trials given as Table E1, raw.ad is a data frame for AD trials given as Table E2, initial is a vector of initial values of parameters to be drawn by the MCMC, candvar is a variance–covariance matrix of the normal candidate-distribution for the Metropolis–Hastings algorithm, mcnum is the number of parameter samples drawn by the MCMC, thin is the thinning interval used in the MCMC (a sample at only every thin-th iteration is taken) and burnin is the number of burn-in samples in the MCMC. And also, the users need to precompile a C code

Table E1. Example of R object raw.ipd to be prepared for IPD trials.

STUDY	GROUP	OUT	COV
1	0	−14	62
1	0	18	61
1	0	−10	65
1	1	−57	61
1	0	12	75
1	1	−46	77
⋮	⋮	⋮	⋮

raw.ipd: STUDY indicates the trial-specific number which is assigned to all trials uniquely, GROUP indicates the indicator value which should be coded 0/1 to denote control/treatment group, OUT indicates the continuous outcome value observed for each patient and COV indicates the continuous covariate observed for each patient.

Table E2. Example of R object and raw.ad to be prepared for AD trials.

STUDY	GROUP	N	OUTM	OUTV	COVM	COVV
3	0	1337	−17.55	481.80	70.43	7.56
3	1	1314	−28.20	474.37	70.39	7.67
4	0	2371	−13.88	396.01	71.54	44.62
4	1	2365	−25.39	339.30	71.64	45.16
⋮	⋮	⋮	⋮	⋮	⋮	⋮

raw.ad: STUDY indicates the trial specific number which is assigned to all trials uniquely, GROUP indicates the indicator value which should be coded 0/1 to denote control/treatment group, N indicates the number of patients in each group and trial, OUTM indicates the mean of the continuous outcome from each group and trial, OUTV indicates the variance of the continuous outcome from each group and trial, COVM indicates the mean of the continuous covariate from each group and trial, and COVV indicates the variance of the continuous covariate from each group and trial.

appropriately, which is provided behind the R code. Calling a C function in R is just due to the computational advantage of C. A manual 'textitWriting R Extensions' provides many details on incorporating C code into R, which is available from the Comprehensive R Archive Network website.

```
#### Once run the following R code, and then use a function "SIPDmethod" with your R objects

#### Load a package

library(doBy)

#### Call a C function

dyn.load("filename")
# Load a DLL file to use the C function in R

MCMC_MH <- function(a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s)
{
    .Call("MCMC_MH", as.double(a), as.double(b), as.double(c), as.double(d),
    as.double(e), as.double(f), as.double(g), as.double(h), as.double(i), as.double(j),
    as.double(k), as.double(l), as.double(m), as.double(n), as.double(o), as.double(p),
    as.double(q), as.double(r), as.double(s))
    # The first argument is the name of the C function to be called
}

#### Function for generating SIPD, which returns multiple sets of the simulated outcome value and covariate
#### for patients in the AD trials

SIPDgenerator <- function(AD.ad, para, np)
{
    judge <- 0

    while(judge == 0)
    {
        nrand <- rnorm(np)
        rnmean <- mean(nrand)
        rnvar <- var(nrand)

        sample.cov <- AD.ad[3] + (nrand-rnmean)/sqrt(rnvar)*sqrt(AD.ad[4])

        nrand <- rnorm(np)
        rnmean <- mean(nrand)
        rnvar <- var(nrand)
        B <- para*var(nrand,sample.cov)

        tmp <- B2-rnvar*(para2*AD.ad[4]-AD.ad[2])

        if(tmp > 0)
        {
            delta <- (-B + sqrt(tmp))/rnvar
            sample.y <- AD.ad[1] + para*(sample.cov-AD.ad[3]) + delta*(nrand-rnmean)
            judge <- 1
        }
    }

    return(cbind(sample.y,sample.cov))
}

#### Function for SIPD method, which returns parameter estimates and their SEs

SIPDmethod <- function(raw.ipd, raw.ad, initial, candvar, mcnum, thin, burnin)
{
```

```
# Data processing

studid.ipd < − unique(raw.ipd$STUDY)
studid.ad < − unique(raw.ad$STUDY)
studnum.ipd < − length(studid.ipd)
studnum.ad < − length(studid.ad)
studnum < − studnum.ipd + studnum.ad

tmp1 < − summaryBy(OUTSTUDY + GROUP,data = raw.ipd,FUN = c(mean,sd,length))
tmp2 < − reshape(tmp1,idvar = c("STUDY"),timevar = "GROUP",direction = "wide")
tmp3 < − reshape(raw.ad,idvar = c("STUDY"),timevar = "GROUP",direction = "wide")

n < − array(0,dim = c(studnum,2))
n[tmp2$STUDY,] < − as.matrix(tmp2[,c("OUT.length.0","OUT.length.1")])
n[tmp3$STUDY,] < − as.matrix(tmp3[,c("N.0","N.1")])
n[tmp2$STUDY,] < − as.matrix(tmp2[,c("OUT.length.0","OUT.length.1")])

AD < − array(0,dim = c(studnum,2,4))
AD[tmp2$STUDY,,1] < − as.matrix(tmp2[,c("OUT.mean.0","OUT.mean.1")])
AD[tmp2$STUDY,,2] < − as.matrix(tmp2[,c("OUT.sd.0","OUT.sd.1")])2
AD[tmp3$STUDY,,1] < − as.matrix(tmp3[,c("OUTM.0","OUTM.1")])
AD[tmp3$STUDY,,2] < − as.matrix(tmp3[,c("OUTV.0","OUTV.1")])

tmp1 < − summaryBy(COVSTUDY + GROUP,data = raw.ipd,FUN = c(mean,sd))
tmp2 < − reshape(tmp1,idvar = c("STUDY"),timevar = "GROUP",direction = "wide")
AD[tmp2$STUDY,,3] < − as.matrix(tmp2[,c("COV.mean.0","COV.mean.1")])
AD[tmp2$STUDY,,4] < − as.matrix(tmp2[,c("COV.sd.0","COV.sd.1")])2
AD[tmp3$STUDY,,3] < − as.matrix(tmp3[,c("COVM.0","COVM.1")])
AD[tmp3$STUDY,,4] < − as.matrix(tmp3[,c("COVV.0","COVV.1")])

cov.mean < − apply(AD[,,3],1,mean)

tmp1 < − data.frame(STUDY = studid.ipd)
tmp2 < − data.frame(COVM = cov.mean[studid.ipd])
tmp3 < − cbind(tmp1,tmp2,diag(studnum.ipd))
tmp4 < − merge(raw.ipd,tmp3)
tmp5 < − transform(tmp4,GCOV1 = tmp4$GROUP*tmp4$COVM,GCOV2 = tmp4$GROUP*(tmp4$COV-tmp4
$COVM))

d1 < − subset(tmp5,select = OUT)[,1]
d2 < − subset(tmp4,select = −c(STUDY,GROUP,OUT,COV,COVM))
d3 < − subset(tmp5,select = c(GROUP,COV))
d4 < − subset(tmp5,select = c(GCOV1,GCOV2))

IPD < − as.matrix(cbind(d1,d2,d3,d4))
dimnames(IPD) < − NULL

XX.tmp < − matrix(0,studnum + 4,studnum + 4)
XX.tmp[1:studnum,1:studnum] < − diag(apply(n,1,sum))
XX.tmp[studnum + 1,1:(studnum + 1)] < − c(n[,2],sum(n[,2]))
XX.tmp[studnum + 2,1:(studnum + 2)] < − c(apply(n*AD[,,3],1,sum),sum(n[,2]*AD[,2,3]),
                                          sum(n*(AD[,,3]2 + AD[,,4])))
XX.tmp[studnum + 3,1:(studnum + 3)] < − c(n[,2]*cov.mean,sum(n[,2]*cov.mean),
                                          sum(n[,2]*cov.mean*AD[,2,3]),sum(n[,2]*cov.mean2))
XX.tmp[studnum + 4,1:(studnum + 4)] < − c(n[,2]*(AD[,2,3]-cov.mean),sum(n[,2]*(AD[,2,3]-cov.mean)),
                                          sum(n[,2]*(AD[,2,4] + AD[,2,3]2-AD[,2,3]*cov.mean)),
                                          sum(n[,2]*cov.mean*(AD[,2,3]-cov.mean)),
                                          sum(n[,2]*(AD[,2,4] + (AD[,2,3]-cov.mean)2)))

XX.all < − XX.tmp + t(XX.tmp)
diag(XX.all) < − diag(XX.all)/2
```

```
Xy.all <- c(apply(n*AD[,,1],1,sum),sum(n[,2]*AD[,2,1]),0,sum(n[,2]*cov.mean*AD[,2,1]),0)

XX.ipd <- t(IPD[,-1])%*%IPD[,-1]
Xy.ipd <- t(IPD[,-1])%*%IPD[,1]
yy.ipd <- t(IPD[,1])%*%IPD[,1]

n.ad <- as.vector(n[studid.ad,])
AD.ad1 <- as.vector(AD[studid.ad,,1])
AD.ad2 <- as.vector(AD[studid.ad,,2])
AD.ad3 <- as.vector(AD[studid.ad,,3])
AD.ad4 <- as.vector(AD[studid.ad,,4])

# Sample parameters from its posterior distribution by MCMC

para.ord <- c(studid.ipd,studid.ad,(studnum + 1):(studnum + 5))
candvar.chol <- as.vector(t(chol(candvar[para.ord,para.ord])))

mugam <- numeric(mcnum*(studnum + 5))
mugam <- MCMC_MH(
    studnum, studnum.ipd, sum(n[studid.ipd,]),
    yy.ipd[[1]], as.vector(Xy.ipd), as.vector(XX.ipd), studnum.ad, n.ad,
    AD.ad1, AD.ad2, AD.ad3, AD.ad4, cov.mean[studid.ad],
    initial[para.ord], candvar.chol, mcnum, thin, burnin, mugam)

mugam.ac <- matrix(1:(mcnum*(studnum + 5)),nrow = mcnum,ncol = studnum + 5,byrow = T)
mu <- mugam[mugam.ac[,studnum + 2]]
gam <- mugam[mugam.ac[,studnum + 4]]
SIPD.para <- cbind(mu,mu + gam)

# Generate the SIPD and then estimate parameters of the IPD meta-analysis model

SIPD.yz <- array(0,dim = c(studnum,2))
SIPD.result <- array(0,dim = c(mcnum,2,studnum + 5))
SIPD.Xy <- Xy.all

for(r in 1:mcnum)
{
    for(i in studid.ad)
    {
        for(k in 1:2)
        {
            SIPD <- SIPDgenerator(AD[i,k,], SIPD.para[r,k], n[i,k])
            SIPD.yz[i,k] <- sum(SIPD[,1]*SIPD[,2])
        }
    }

    SIPD.Xy[studnum + 2] <- Xy.ipd[studnum.ipd + 2] + sum(SIPD.yz[studid.ad,])
    SIPD.Xy[studnum + 4] <- Xy.ipd[studnum.ipd + 4] + sum(SIPD.yz[studid.ad,2]
                                                          -n[studid.ad,2]*AD[studid.ad,2,1]*cov.mean[studid.ad])

    SIPD.beta <- solve(XX.all)%*%SIPD.Xy
    SIPD.RSS <- sum(n[,1]*(AD[,1,2] + AD[,1,1]2) + n[,2]*(AD[,2,2] + AD[,2,1]2))
                        -2*t(SIPD.Xy)%*%SIPD.beta + t(SIPD.beta)%*%XX.all%*%SIPD.beta

    SIPD.sigma <- (SIPD.RSS/(sum(n)-length(SIPD.beta)))[[1]]
    AA <- (-1)/SIPD.sigma*XX.all
    AB <- (-1)/SIPD.sigma2*(SIPD.Xy-XX.all%*%SIPD.beta)*2
    BB <- sum(n)/(2*SIPD.sigma2)-1/(SIPD.sigma3)*SIPD.RSS
    SIPD.Vbeta <- solve(-rbind(cbind(AA,AB), c(AB,BB)))
    SIPD.result[r,1,] <- c(SIPD.beta,SIPD.sigma)
    SIPD.result[r,2,] <- diag(SIPD.Vbeta)
```

```
    }

    # Summarise parameter estimates for each SIPD

    Rubin.mean <- apply(SIPD.result[,1,],2,mean)
    Rubin.var <- (1 + 1/mcnum)*apply(SIPD.result[,1,],2,var) + apply(SIPD.result[,2,],2,mean)

    res <- data.frame(Estimate = Rubin.mean,Std.Error = sqrt(Rubin.var))
    rownames(res) <- c(paste("phi",as.character(1:studnum)),"theta","mu","gammaA","gammaW","sigma2")

    return(res)
}

#### End of R code

#### C code for a function of sampling parameters by MCMC, which shoule be stored
#### as a C source file separately from the R source file and precompiled before
#### running the R code

#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <time.h>
#include <Rinternals.h>

double min(double, double);
double urand(void);
double nrand(double, double);

SEXP MCMC_MH(
    SEXP ns_c, SEXP ns_ipd_c, SEXP np_ipd_c,
    SEXP yy_c, SEXP Xy_c, SEXP XX_c, SEXP ns_ad_c, SEXP np_ad_c,
    SEXP AD_ad1_c, SEXP AD_ad2_c, SEXP AD_ad3_c, SEXP AD_ad4_c, SEXP covm_ad_c,
    SEXP initial_c, SEXP candvar_chol_c, SEXP mcnum_c, SEXP thin_c, SEXP burnin_c, SEXP mugam)
{

    int ns = REAL(ns_c)[0], ns_ipd = REAL(ns_ipd_c)[0], np_ipd = REAL(np_ipd_c)[0];
    int nb = ns_ipd + 4;
    double yy = REAL(yy_c)[0];
    double Xy[nb], XX[nb][nb];
    int p1, p2;
    for(p1 = 0; p1 < nb; p1++)
    {
        Xy[p1] = REAL(Xy_c)[p1];
        for(p2 = 0; p2 < nb; p2++) XX[p1][p2] = REAL(XX_c)[p1*nb + p2];
    }

    int ns_ad = REAL(ns_ad_c)[0];
    int np_ad[ns_ad][2];
    double AD_ad1[ns_ad][2], AD_ad2[ns_ad][2], AD_ad3[ns_ad][2], AD_ad4[ns_ad][2], covm_ad[ns_ad];

    int i, k;
    for(i = 0; i < ns_ad; i++)
    {
        covm_ad[i] = REAL(covm_ad_c)[i];
        for(k = 0; k < 2; k++)
        {
            np_ad[i][k] = REAL(np_ad_c)[i + ns_ad*k];
            AD_ad1[i][k] = REAL(AD_ad1_c)[i + ns_ad*k];
            AD_ad3[i][k] = REAL(AD_ad3_c)[i + ns_ad*k];
            AD_ad2[i][k] = REAL(AD_ad2_c)[i + ns_ad*k];
```

```
        AD_ad4[i][k] = REAL(AD_ad4_c)[i + ns_ad*k];
    }
}

int mcnum = REAL(mcnum_c)[0], thin = REAL(thin_c)[0], burnin = REAL(burnin_c)[0];
int mcite = mcnum*thin + burnin + 1;

srand((unsigned)time(NULL));
double tmp = urand();

int leng_p = length(initial_c);
double pcurr[leng_p], pnew[leng_p], ptmp[leng_p], pvar[leng_p][leng_p];
for(p1 = 0; p1 < leng_p; p1++)
{
    pcurr[p1] = REAL(initial_c)[p1];
    pnew[p1] = 0;
    ptmp[p1] = 0;
    for(p2 = 0; p2 < leng_p; p2++) pvar[p2][p1] = REAL(candvar_chol_c)[p1*leng_p + p2];
}

double beta[nb];
for(i = 0; i < ns_ipd; i++) beta[i] = pcurr[i];
beta[ns_ipd] = pcurr[ns];
beta[ns_ipd + 1] = pcurr[ns + 1];
beta[ns_ipd + 2] = pcurr[ns + 2];
beta[ns_ipd + 3] = pcurr[ns + 3];

double lcurrent, Xyb = 0, bXXb_tmp[nb], bXXb = 0;
for(p1 = 0; p1 < nb; p1++)
{
    Xyb += beta[p1]*Xy[p1];
    tmp = 0;
    for(p2 = 0; p2 < nb; p2++) tmp += beta[p2]*XX[p2][p1];
    bXXb_tmp[p1] = tmp;
}

for(p1 = 0; p1 < nb; p1++) bXXb += bXXb_tmp[p1]*beta[p1];
lcurrent = -np_ipd*log(pcurr[ns + 4])/2-(yy-2*Xyb + bXXb)/(2*pcurr[ns + 4]);

double m, sig;
tmp = 0;
for(i = 0; i < ns_ad; i++)
{
    for(k = 0; k < 2; k++)
    {
        m = pcurr[ns_ipd + i] + k*pcurr[ns] + pcurr[ns + 1]*AD_ad3[i][k]
            +k*pcurr[ns + 2]*covm_ad[i] + k*pcurr[ns + 3]*(AD_ad3[i][k]-covm_ad[i]);
        sig = pow(pcurr[ns + 1] + k*pcurr[ns + 3],2)*AD_ad4[i][k] + pcurr[ns + 4];
        tmp += - np_ad[i][k]*log(sig)/2-np_ad[i][k]*(pow(AD_ad1[i][k]-m,2) + AD_ad2[i][k])/(2*sig);
    }
}

lcurrent += tmp;

int j, count = 0;
double r, lnew;
for(j = 0; j < mcite; j++)
{
    for(p1 = 0; p1 < leng_p; p1++) ptmp[p1] = nrand(0,1);

    for(p1 = 0; p1 < leng_p; p1++)
```

```
        {
            tmp = 0;
            for(p2 = 0; p2 < leng_p; p2++) tmp += pvar[p1][p2]*ptmp[p2];
            pnew[p1] = pcurr[p1] + tmp;
        }

        for(i = 0; i < ns_ipd; i++) beta[i] = pnew[i];
        beta[ns_ipd] = pnew[ns];
        beta[ns_ipd + 1] = pnew[ns + 1];
        beta[ns_ipd + 2] = pnew[ns + 2];
        beta[ns_ipd + 3] = pnew[ns + 3];

        Xyb = 0;
        for(p1 = 0; p1 < nb; p1++)
        {
            Xyb += beta[p1]*Xy[p1];

            tmp = 0;
            for(p2 = 0; p2 < nb; p2++) tmp += beta[p2]*XX[p2][p1];
            bXXb_tmp[p1] = tmp;
        }

        bXXb = 0;
        for(p1 = 0; p1 < nb; p1++) bXXb += bXXb_tmp[p1]*beta[p1];

        lnew = -np_ipd*log(pnew[ns + 4])/2-(yy-2*Xyb + bXXb)/(2*pnew[ns + 4]);

        tmp = 0;
        for(i = 0; i < ns_ad; i++)
        {
            for(k = 0; k < 2; k++)
            {
              m = pnew[ns_ipd + i] + k*pnew[ns] + pnew[ns + 1]*AD_ad3[i][k]
                    +k*pnew[ns + 2]*covm_ad[i] + k*pnew[ns + 3]*(AD_ad3[i][k]-covm_ad[i]);
              sig = pow(pnew[ns + 1] + k*pnew[ns + 3],2)*AD_ad4[i][k] + pnew[ns + 4];
              tmp += - np_ad[i][k]*log(sig)/2-np_ad[i][k]*(pow(AD_ad1[i][k]-m,2) + AD_ad2[i][k])/(2*sig);
            }
        }

        lnew += tmp;

        r = urand();
        if(r < min(1,exp(lnew-lcurrent)))
        {
            lcurrent = lnew;
            for(p1 = 0; p1 < leng_p; p1++) pcurr[p1] = pnew[p1];
        }

        if(j > burnin && ((j-burnin)%thin)==0)
        {
            for(p1 = 0; p1 < leng_p; p1++) REAL(mugam)[count*leng_p + p1] = pcurr[p1];
            count += 1;
        }
    }

    return mugam;
}

double min(double x, double y)
{
    return (x < y) ? x: y;
```

```
}
double nrand(double mu, double sigma)
{
    double x, y, z;

    x = urand();
    y = urand();
    z = sqrt(−2.0 * log(x)) * sin(2.0 * M_PI * y);

    return mu + sigma * z;
}

double urand(void)
{
    return rand() / (RAND_MAX + 1.0);
}

#### End of C code
```

Acknowledgements

While undertaking this work, the first author was supported by funding from the MRC Midlands Hub for Trials Methodology Research, at the University of Birmingham (Medical Research Council grant ID G0800808). We thank the two anonymous reviewers for their extremely helpful comments and suggestions, which have greatly improved this article. The INDANA data base was assembled through grants from PHRC national 1997 #97054, PHRC national 2002, BIOMED2 #BMH4-CT98-3291.

References

Thompson SG, Higgins JPT. 2005. Can meta-analysis help target interventions at individuals most likely to benefit? *Lancet* **365**: 341–346.

Morgenstern H. 1982. Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health* **72**: 1336–1344.

Riley RD, Steyerberg EW. 2010. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods* **1**: 2–19.

King G. 1997. A Solution to the Ecological Inference Problem. Princeton University Press, Princeton, New Jersey.

Robinson WS. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* **15**: 351–357.

Elliott P, Wakefield JC, Best NG, Briggs DJ. 2000. Spatial Epidemiology: Methods and Applications, Oxford University Press, Oxford.

Morgenstern H. 1995. Ecologic studies in epidemiology: concepts, principles, and methods. *Annual Review of Public Health* **16**: 61–81.

Wakefield JC. Ecologic studies revisited. 2008. *Annual Review of Public Health* **29**: 75–90.

Thompson SG, Higgins JPT. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* **21**: 1559–1573.

Lambert PC, Sutton AJ, Abrams KR, Jones DR. 2002. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* **55**: 86–94.

Simmonds MC, Higgins JPT. 2007. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine* **26**: 2982–2999.

Berlin JA, Santanna J, Scmid CH, Szczech LA, Feldman HI. 2002. Individual patient versus group-level data meta-regression for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine* **21**: 371–387.

Riley RD, Lambert PC, Abo-Zaid G. 2010. Meta-analysis of individual participant data: rationale, conduct, and reporting. *British Medical Journal* **340**: c221.

Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ and Thompson S.G. 2005. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* **2**: 209–217.

Stewart LA, Tierney JF. 2002. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and the Health Professions* **25**: 76–97.

Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Boutitie F. 2008. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine* **27**: 1870–1893.

Sutton AJ, Kendrick D, Coupland CAC. 2008. Meta-analysis of individual- and aggregate-level data. *Statistics in Medicine* **27**: 651–669.

Goldstein H, Yang M, Omar RZ, Turner RM, Thompson SG. 2000. Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society, Series C* **49**: 399–412.

Donegan S, Williamson P, D'Alessandro U, Garner P, Tudur Smith C. 2013. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. *Statistics in Medicine* **32**: 914–930.

Jansen JP. 2012. Network meta-analysis of individual and aggregate level data. *Research Synthesis Methods* **3**: 177–190.

Saramago P, Sutton AJ, Cooper NJ, Manca A. 2012. Mixed treatment comparisons using aggregate and individual participant level data. *Statistics in Medicine* **31**: 3516–3536.

Jackson C, Best N, Richardson S. 2006. Improving ecological inference using individual-level data. *Statistics in Medicine* **25**: 2136–2159.

Haneuse SJ, Wakefield JC. 2008. The combination of ecological and case–control data. *Journal of the Royal Statistical Society: Series B* **70**: 73–93.

Wakefield JC. 2004. Ecological inference for 2×2 tables. *Journal of the Royal Statistical Society, Series A* **167**: 385–445.

Wakefield JC, Haneuse SJ, Dobra A, Teeple E. 2011. Bayes computation for ecological inference. *Statistics in Medicine* **30**: 1381–1396.

Rubin DB. 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley and Sons, New York.

Wang JG, Staessen JA, Franklin SS, Fagard R, Gueyffier F. 2005. Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome. *Hypertension* **45**: 907–913.

Gueyffier F, Boutitie F, Boissel JP, Coope J, Cutler J, Ekbom T, Fagard R, Friedman L, Perry HM, *et al*. 1995. INDANA: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results. *Thérapie* **50**: 353–362.

Liu L, Wang JG, Gong L, Liu G, Staessen JA. Comparison of active treatment and placebo in older Chinese patients with isolated systolic hypertension. Systolic hypertension in China (Syst-China) collaborative group. *Journal of Hypertension* 1998; **16**: 1823–1829.

Staessen JA, Fagard R, Thijs L, Celis H, Arabidze GG, Birkenhager WH, Bulpitt CJ, de Leeuw PW, Dollery CT, Fletcher AE *et al*. 1997. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. The systolic hypertension in Europe (Syst-Eur) trial investigators. *The Lancet* **350**: 757–764.

Amery A, Birkenhager W, Brixko P, Bulpitt C, Clement D, Deruyttere M, De Schaepdryver A, Dollery C, Fagard R, Forette F *et al*. 1985. Mortality and morbidity results from the European Working Party on high blood pressure in the elderlytrial. *The Lancet* **1**: 1349–1354.

Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. 2001. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* **20**: 2219–2241.

Neuhaus JM, Kalbfleisch JD. 1998. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* **54**: 638–645.

Mancl LA, Leroux BG, DeRouen TA. 2000. Between-subject and within-subject statistical information in dental research. *Journal of Dental Research* **79**: 1778–1781.

Begg MD, Parides MK. 2003. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Statistics in Medicine* **22**: 2591–2602.

Simmonds MC. 2005. Statistical methodology of individual patient data. *Ph.D. Thesis*, University of Cambridge.

Heitjan DF, Rubin DB. 1991. Ignorability and coarse data. *The Annals of Statistics* **19**: 2244–2253.

Little RJA, Rubin DB. 2002. Statistical Analysis with Missing Data, 2nd edition. John Wiley and Sons, New York.

Cheng RCH. 1984. Generation of inverse Gaussian variates with given sample mean and dispersion. *Applied Statistics* **33**: 309–316.

Engen S, Lillegard M. 1997. Stochastic simulations conditioned on sufficient statistics. *Biometrika* **84**: 235–240.

Lindqvist BH, Taraldsen G. 2005. Monte Carlo conditioning on a sufficient statistic. *Biometrika* **92**: 451–464.

Cochran WG. 1954. The combination of estimates from different experiments. *Biometrics* **10**: 101–129.

Higgins JPT, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**: 1539–1558.

Riley RD, Simmonds MC, Look MP. 2007. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology* **60**: 431–439.

Ahmed I, Sutton AJ, Riley RD. 2012. Assessment of publication bias, selection bias, and unavailable data in meta-analysis using individual participant data: a database survey. *The British Medical Journal* **344**, DOI:10.1136/bmj.d7762.

Higgins JPT, Thompson SG, Spiegelhalter DJ. 2009. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, Series A* **172**: 137–159.

Tanner MA, Wong WH. 1987. The calculation of posterior distribution by data augmentation. *Journal of the American Statistical Association* **82**: 528–540.

van Dyk DA, Meng XL. 2001. The art of data augmentation. *Journal of Computational and Graphical Statistics* **10**: 1–50.

Riley RD, Kauser I, Bland M, Thijs L, Staessen JA, Wang J, Gueyffier F, Deeks JJ. 2013. Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. *Statistics in Medicine* **32**: 2747–2766.

Tsiatis AA. 2006. Semiparametric Theory and Missing Data. Springer Series in Statistics, Springer-Verlag, New York.

Pullin DI. 1979. Generation of normal variates with given sample mean and variance. *Journal of Statistical Computation and Simulation* **9**: 303–309.

Gelman A, Carlin JB, Stern HS, Rubin DB. 1995. Bayesian Data Analysis. Chapman and Hall, New York.